Journal of Memory and Language 76 (2014) 94–112

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Making sense of Kafka: Structural biases induce early sense commitment for metonyms

Joel Fishbein[a], Jesse A. Harris[a,b,*]

[a] Pomona College, Claremont, CA, USA
[b] University of California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 7 May 2013
revision received 10 June 2014
Available online 12 July 2014

Keywords:
Sentence processing
Metonymy
Polysemy
Eye tracking
Underspecification
Thematic role assignment

ABSTRACT

Prior research suggests that the language processor initially activates an underspecified representation of a metonym consistent with all its senses, potentially selecting a specific sense if supported by contextual and lexical information. We explored whether a structural heuristic, the *Subject as Agent Principle*, which provisionally assigns an agent theta role to canonical subjects, would prompt immediate sense selection. In Experiment 1, we found initial evidence that this principle is active during offline and online processing of metonymic names like *Kafka*. Reading time results from Experiments 2 and 3 demonstrated that previous context biasing towards the metonymic sense of the name reduced, but did not remove, the agent preference, consistent with Frazier's (1999) proposal that the processor may avoid selecting a specific sense, unless grammatically required.

Published by Elsevier Inc.

Introduction

Interpreting polysemes requires that the language processor integrates sentential information with lexical and semantic knowledge to choose one sense of a word from among many, possibly dozens of, related senses (Zipf, 1945). For example, it must be able to entertain the possibility that the word *Vietnam* may refer to a country, a war, a group of people, a government, a United Nations delegation, or a soccer team, and then utilize some salient piece of information to select the correct meaning in context (Cruse, 1986; Nunberg, 1979). Despite the availability of multiple related senses for polysemes, the language processor often seems to avoid simply committing to the most frequent sense, in contrast with homonymous words like *bank* (e.g., Frazier & Rayner, 1990, cf. Duffy, Morris, & Rayner, 1988; Swinney, 1979). In this paper, we aim to demonstrate that syntactic position has a powerful and immediate impact on whether a specific interpretation is selected. In particular, we present evidence from three experiments supporting the view that the language processor utilizes information from a default structural heuristic to make immediate sense selection decisions, but delays when faced with weaker contextual information.

Metonymy

Metonyms are a type of polyseme, a word with two or more senses (i.e., related meanings). For regular metonymy, it is commonly thought that there is a single base sense and one or more related senses derived via various types of metonymic rules (Nunberg, 1995, 2004). From among the several views concerning how related senses are stored and accessed during real time processing (see Frisson & Pickering, 2001, or Frisson, 2009, for review), we concentrate on the *Underspecification Model*, in which the language processor accesses the meaning of a polyseme in two stages (see Frazier & Rayner, 1990; Frisson, 2009; Frisson & Pickering, 1999, 2001, 2007,

* Corresponding author. Address: Department of Linguistics, University of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095-1543, USA.

E-mail addresses: joel.n.fishbein@gmail.com (J. Fishbein), jesse.harris@pomona.edu (J.A. Harris).

http://dx.doi.org/10.1016/j.jml.2014.06.005
0749-596X/Published by Elsevier Inc.

among others). Upon first encountering a polysemous word, the processor activates a semantically underspecified representation, in which certain, specific semantic features of the lexical representation go unexpressed in favor of a more general, vague, or nonspecific representation. This underspecified representation facilitates equal access to all senses of the polysemous word, so that any non-literal senses are no more difficult to access over its literal sense. Second, the processor selects a specific sense consistent with the available contextual and lexical information, if required. This secondary 'homing-in' stage is affected by several factors, including the importance of the word in the sentence, the strength of contextual information, as well as the demands and requirements of the task, although the language processor may elect to forgo this stage if such factors are not sufficiently compelling (Frisson, 2009).

Evidence for semantic underspecification of polysemes was observed by Frazier and Rayner (1990), who found processing costs for homographs like *pitcher*, but not for polysemes like *newspaper*, when subsequent material supported a subordinate sense. Frisson and Pickering (1999) compared contextually relevant place-for-institution interpretations (*convent*) against those with unfamiliar place-for-institution metonymies (*stadium*), finding processing costs only for unfamiliar metonymies. Later work with producer-for-product metonymies suggests that the costs associated with processing unfamiliar metonyms are effectively mitigated by prior supporting context (Frisson & Pickering, 2007). Further evidence for underspecified semantic representations comes from the finding that polysemous words show a processing advantage over homonyms across a variety of paradigms, including lexical decision (e.g., Klepousniotou, 2002; Klepousniotou & Baum, 2007; Wiliams, 1992), eye movements (e.g., Frisson & Frazier, 2004) and magenetoencephalography (Beretta, Fiorentino, & Poeppel, 2005), though there are other interpretations of such effects (Foraker & Murphy, 2012; Klein & Murphy, 2002; see also Pylkkänen, Llinás, & Murphy, 2006 for the nuanced view that individual senses are represented on distinct nodes under a single abstract representation of the polyseme).

One concern that has been raised (see Foraker & Murphy, 2012, for discussion) with the Underspecification Model is that, in studies such as Frazier and Rayner (1990) and Frisson and Pickering (1999), the central evidence presented as support for the model manifests in the form of null results. Even though others (Frisson & Pickering, 2007; McElree, Frisson, & Pickering, 2006) show that readers are sensitive to closely related manipulations like complement coercion (Jackendoff, 1997; Pustejovsky, 1995), it remains a possibility that there is a subtle processing cost for accessing the non-literal sense, and that the disambiguating contexts manipulations used were not strong enough to produce an effect. The studies presented below address this concern indirectly: accessing the non-literal sense of a regular metonym is shown to be costly in one syntactic position, but not another.

Homing in and contextual strength

We assume that selectional information from a verb prompts the processor to select a specific sense of such metonyms, and it does so without penalty. We concentrate here on producer-for-product metonymy, in which the name of an author or artist like *Kafka* refers not to the literal individual Kafka, but to the works associated with the individual: here, Kafka's writings. For instance, the verb *read* in *The students read Kafka* subcategorizes for a readable object, initiating sense selection for the metonymic, literature sense of *Kafka*. We assume an early stage of lexical access in which an underspecified representation, consistent with both literal and metonymic senses, is activated prior to homing in. We posit that at this stage there is no cost for accessing the metonymic interpretation over controls supporting the literal sense, as in *The students met Kafka* (Frisson & Pickering, 1999, 2001). Other sentence contexts may leave the metonym unresolved, as in *The students discussed Kafka*, where either sense is permitted; as such, the processor may *opt out* of a finer sense selection process in certain cases. Of course, the processing system may elect to use more general discourse context, e.g., in a situation that primes or otherwise supports one specific sense, for further sense selection.

Numerous questions regarding sense selection processes remain. Can other types of linguistic information besides subcategorization requirements tempt the processor into making an immediate sense selection? Or is the processor required to select a particular sense only when it is forced to by specific lexical selection requirements? We discuss two possibilities below.

Only lexical constraints mandate sense selection

On one view, the language processor will be forced to make an immediate sense selection only if the metonym is subject to strong and local lexical constraints (as in subcategorization requirements) at the point of interpretation. Under this more restricted account, the processor will be forced to select a specific sense only in such cases, possibly forgoing the sense selection stage in ambiguous cases like *The students discussed Kafka*, mentioned above.

Appealing as such a model is, however, evidence from prior experiments suggests that the processor may be tempted to rule out certain sense interpretations before reaching subcategorization restrictions, as in lexical processing more generally (see, e.g., Morris, 2006, for review). For example, experiments probing the resolution of number ambiguities (Bader & Häussler, 2009), and distributed and collective readings of noun phrases (Frazier, Pacht, & Rayner, 1999) indicate that the processor commits to highly specific interpretations of words before reaching a disambiguating verb. While it remains possible that metonyms are a special case, we believe it is likely that other types of constraining information entice the processor to immediately select for a specific sense, as well. Indeed, the experiments below show that grammatical constraints, in additional to purely lexical constraints, may tempt the processor to select a more specific sense.

Only grammatical constraints mandate sense selection

Another possibility is that the processor must respond to syntactic decisions by immediately committing to whatever interpretation is required by the rules of the grammar. This is not to say that contextual information, or even sense-dominance, does not influence whether and when a specific sense is selected; it merely states that the processor cannot delay when it comes to syntactic information. This is captured by the conjunction of Frazier's (1999) principles of *Immediate Partial Interpretation* and *Minimal Semantic Commitment*:

(1) *Immediate Partial Interpretation*
Perceivers must choose between grammatically incompatible meanings of a word or constituent immediately, by the end of the word or constituent, unless this conflicts with the dictates of the grammar.

(2) *Minimal Semantic Commitment*
Premature or arbitrary semantic commitments (concerning words) are made only when forced by [(1)].

If (1) and (2) hold when processing metonyms, then we would expect the processor to select a specific sense of a metonym whenever required by syntactic constraints imposed by the grammar, avoiding arbitrary semantic decisions whenever possible. This view subsumes the more restricted lexical-only view, as lexical selection information surely counts as grammatical, in that a lexical entry imposes subcategorization requirements on elements with which it forms a syntactic constituent.

It seems to us that both views are possible, in principle. In order to tease the two options apart, we require an environment that will allow us to decouple syntactic demands from stronger, more specific, lexical constraints at the point at which the metonym appears. One such environment in English is the sentential subject position, for which the processor, we argue, assumes a canonical thematic role assignment upon encountering a clause-initial noun. In what follows, we assume a basic vocabulary of thematic (theta) roles (Fillmore, 1968), such as AGENT – *actor*; and THEME – *acted upon*, which are syntactically *assigned* to noun phrases by virtue of standing in a particular structural relationship to a thematic assigner (see, e.g., Grimshaw, 1990; Jackendoff, 1972, among others). We propose that the processor follows the Subject as Agent Principle (SAP), in which the processor provisionally assigns the subject of a clause in English an Agent theta role, all else being equal (see also Ferreira, 2003; Grodinsky, 1986, discussed below). Assuming further that theta role processing places grammatical constraints upon interpretation (e.g., Clifton, 1993; Clifton et al., 2003), we expect that application of SAP on a metonym would lead to immediate sense selection. The justification for SAP is briefly discussed below.

Subject as agent principle

There is evidence from a wide range of sources that the language processor engages in default assignment of thematic roles with respect to the position of the noun in sentential structure (e.g., Bever, 1970; Ferreira, 2003; Garnham & Oakhill, 1987; Grodinsky, 1986; MacWhinney, Bates, & Kliegl, 1984, among others). For example, Grodinsky (1986) presented an account of differential processing for active and passive sentence processing for agrammatic patients (Schwartz, Saffran, & Marin, 1980). Although agrammatic aphasics successfully interpret non-reversible passives (*The ball is kicked by the boy*) on par with active counterparts, they struggle to reach the correct interpretation of reversible passives containing multiple agentive nouns (*The boy is pushed by the girl*), which lack semantic cues guiding thematic role assignment. On Grodzinsky's analysis, if the processor encounters a noun not assigned a theta role in a position typically associated with a specific theta role, then it assigns that noun the default theta role for that position. The subject NP position (canonically the first NP in the sentence) is typically the Agent (or proto-Agent, as in Dowty, 1991) in English, and so the language processor provisionally assigns the subject NP an Agent theta role. Of course, this decision may turn out to be incorrect. While normal readers may easily correct the misanalysis, agrammatic patients fail to consistently construct the appropriate structure that would lead to the appropriate analysis, in which the thematic default must be overturned.

A related, but distinct, account for thematic structure preferences was offered in Ferreira (2003), who found that readers misinterpreted non-canonical sentences in English (specifically, passives and object-clefts) more often than canonical counterparts (actives and subject-clefts), regardless of the surface frequency of the construction. Ferreira (2003) proposed a processing architecture in which simple processing heuristics may sometimes dominate more deliberate syntactic algorithms, assigning the structure a shallow parse which is "good enough" for the purpose at hand (see also Ferreira, Bailey, & Ferraro, 2002; Oakhill & Garnham, 1996; Oakhill, Garnham, & Vonk, 1989; Sanford, 2002). In this account, the NVN heuristic (Townsend & Bever, 2001) offers a rough and ready approximation of the thematic structure, in which the first noun is expected to provide the (proto-)Agent, the verb an action, and the final noun a (proto-)Patient.

The NVN account diverges from Grodzinsky's in that the NVN heuristic applies to an entire clause, and operates independently, or in-tandem with, syntactic algorithms, whereas Grodzinksy's approach is consistent with a two stage model of the parser, which attempts to resolve grammatical dependencies immediately. In other words, the problem of which theta role to assign to a clause-initial noun is resolved by a simple, cost-effective strategy, defaulting to the Agent role when no other role has been assigned to the noun. Another way of putting the difference between Grodzinky's and Ferreira's approach is in terms of a *reflex* vs. an *expectation*, respectively.

As we are primarily concerned with how such defaults might plausibly affect the resolution of metonymic senses, we first formulate a general principle, the Subject as Agent Principle (SAP), in terms compatible with either view, and then discuss two variants differing in terms of their relation to more general context.[1]

(3) *Subject As Agent Principle*
Provisionally assign an (animate) sentential subject an Agent thematic role.

Online processing studies of normal reading have yielded results generally consistent with the default thematic role assignment proposed in SAP. For example, Clifton et al. (2003) found that inanimate nouns in sentence-initial position elicited slower reading times in relative clause structures, as in *The ransom paid by the parents was unreasonable*, than animate counterparts, an effect which was not modulated by disambiguating the relative clause structure (also Ferreira & Clifton, 1986; McRae, Spivey-Knowlton, & Tanenhaus, 1998). Thus, while the processor may in principle be influenced by task-specific demands or other special circumstances, we propose that in normal reading, the initial thematic processing of possible subjects follows SAP, which allows the processor to continue interpreting the sentence without having to delay thematic role assignment until it encounters additional information. A processing system that assigns an initial theta role by SAP is compatible with multiple models of sentence processing, including a serial depth-first model, as in the classic garden-path model (Frazier & Rayner, 1982), as well as a variety of parallel models which rank preferred analyses above others (e.g., MacDonald, Perlmutter, & Seidenberg, 1994), and good enough processing (Ferreira et al., 2002). For concreteness, we speak of reanalysis in which the processor may revise an earlier processing decision, but remain uncommitted to a general model of sentence processing, for the simple reason that we believe SAP is compatible with multiple instantiations of the sentence processor. Whatever general model one favors, however, principles like SAP allow us to account for the highly incremental and economical nature of sentence processing. When paired with the possibility of underspecified representations, the processor is allowed to get on with the business of parsing sentences, making rough and ready grammatical decisions when required, while leaving finer semantic distinctions vague if possible, cf. Just and Carpenter's (1980) *Immediacy Assumption*.

However, SAP could nevertheless be formulated according to different strengths regarding the situations in which it applies. A weaker version of SAP would assign an Agent theta role to an initial noun just in case it had no other context to inform its decision. In this case, SAP could be considered a *last resort* mechanism, which applies only when the discourse context is impoverished (4).

(4) *Subject As Agent Principle (weak version)*
Provisionally assign an (animate) sentential subject an Agent thematic role only in the absence of other contextual indicators to guide interpretation.

A much stronger version of SAP operates regardless of biasing context. This variant of SAP could be considered an automatic, perhaps modular, syntactic response generated whenever structurally possible, in line with Grodinsky (1986).

(5) *Subject As Agent Principle (strong version)*
Provisionally assign an (animate) sentential subject an Agent thematic role regardless of contextual support.

If correct, the strong variant would provide an economical method of processing highly regular structures, conceivably leaving more attentional resources available for other, potentially more demanding tasks.

To preview the findings below, Experiment 1 below was designed to test whether SAP (3) in general is a viable principle. According to the more constrained version of the homing-in process in which only lexical constraints require sense selection, we would expect that the processor could opt to leave a noun with multiple related senses in subject position underspecified. However, we find evidence in multiple experiments suggesting that such subjects are preferentially interpreted in their literal sense, consistent with the general variant of SAP, thus supporting the view that grammatical constraints, in addition to lexical ones, may necessitate sense selection.

Experiments 2 and 3 address the strength of SAP. If general context is sufficient to override processing costs associated with the provisional Agent role assignment, then it would seem that the weaker variant of SAP (4) is warranted: SAP would then operate, as it were, only as a last resort mechanism. On the other hand, if SAP operates independently of context, assigning an Agent role whenever semantically possible, we then have evidence for a stronger variant of SAP (5). The results from both experiments support the stronger variant over the weaker one.

Finally, we stress that we expect that SAP under any guise is not a universal processing principle. Rather, it may reflect strategic processing for languages with impoverished morphosyntactic information on its nominal arguments, for which structural position provides a strong indication of thematic function. SAP might not hold for languages like Russian, which have retained relatively rich agreement systems that more transparently indicate the structural role of the noun (see MacWhinney et al., 1984, for support).

The present experiments

The following experiments probe whether the application of a structural default like SAP initiates immediate sense commitment to producer-for-product metonyms.

[1] Alternatively, the general preference for agentive subjects might be driven by a general discourse constraint that associates the first noun as the sentence topic by default (see, e.g., Erteschik-Shir, 2007, for review of sentence topics). As sentence topics are typically animates and animates make good agents, SAP could be considered a by-product of an unrelated preference. While our experiments do not eliminate the possibility of a discourse-oriented preference, we believe that SAP as formulated above provides a simpler alternative that should be pursued as a first hypothesis. In addition, the results of Experiment 3 cast doubt on the reduction of SAP to general topic preferences, although we do not doubt that agentivity and topicality are intertwined. We thank Dave Kush for raising this suggestion.

We expect that the language processor will home in immediately on one specific sense of a metonym when it is required to make a grammatical decision. The default Agent-role assignment commits the processor to the sense of a metonym like *Kafka* that is consistent with an Agent role assignment – i.e., its literal, person sense, only when in subject position. In contrast, information biasing towards, but not grammatically constraining, sense selection, should not necessarily force immediate commitment to a specific sense, but instead facilitate potentially costly reinterpretation processes.

These expectations were tested in the three experiments. Experiment 1 tests the claim that processing of potential metonyms is guided by the general version of SAP (3), in offline and online measures. Experiments 2 and 3 provide support for the stronger version of SAP, in that the processor appears to obey SAP even in the face of extra-sentential context (Experiment 2), as well as strongly biasing topic-setting clauses immediately preceding the subject (Experiment 3). However, readers are not wholly insensitive to contextual bias: when context supports the metonymic sense, the penalty for violating SAP is reduced, but not entirely eliminated.

Sense norming

The metonymic names used in all three experiments below were normed in a post hoc Internet experiment consisting of two parts presented on Ibex Farm (Drummond, 2012). In the first part, 20 native speakers of English recruited through Amazon Mechanical Turk provided three words or phrases that they immediately associated with the name, interspersed with 32 unrelated fillers, and presented in random order. After unfamiliar names and a few nonsensical responses were removed, responses were scored as Ambiguous (22%), Literal (35%), or Metonymic (43%),[2] showing no offline preference for the literal sense of the expression. In addition, we analyzed the responses by order. The first response showed an overall increase in Literal responses, Ambiguous (25%), Literal (42%), Metonymic (33%). The second showed a prevalence of Metonymic responses, Ambiguous (19%), Literal (33%), Metonymic (48%), as did the third, Ambiguous (21%), Literal (28%), Metonymic (50%).

In the second part of the norming task, the same subjects chose which sense immediately occurred to them upon viewing a metonymic name like *Kafka* from four alternatives: *The actual person Kafka* (13%), *The writings of Kafka* (47%), *Both equally* (19%), and *Neither/Not sure* (21%). Removing the last two response types, only 22% of trials elicited the actual person sense unambiguously. Even when the *Both equally* responses were added to *The actual person* response, the 41% combined total still failed to surpass the metonymic option (59%). It is therefore highly unlikely that a literal sense of the potentially metonymic name dominated during sentence processing independently of sentence context.

Although anonymous reviewers proposed an additional a corpus study or cloze measure to determine predictability of the verbs in our manipulation, we unfortunately could not find a way to do so that did not crucially interact with our theoretical hypothesis. Corpus examples were too often ambiguous or in incomparable syntactic positions, and placing a potentially metonymic name like *Kafka* in subject position of a sentence frame for completion would, by hypothesis, bias its interpretation towards the literal interpretation via SAP. Finding a preference in cloze measures towards verbs that support the literal sense in sentence frames like *Kafka was* ________ could be taken as support of our hypothesis, rather than as a genuine norming measure; indeed, Experiment 1A finds such a bias in a distinct, but related, fill-in-the-blank completion task.

Experiment 1

Experiment 1 consisted of three studies. The first presents results from an offline fill-in-the-blank study (Experiment 1A), which provides a proof of concept that SAP guides the interpretation of producer-for-product metonyms. The remaining two experiments test how sentence position affects the processing of potentially metonymic names during online processing in self-paced reading (Experiment 1B) and eye tracking (Experiment 1C) studies. The items for Experiment 1 can be found in Appendix A.

Experiment 1A: offline fill-in-the-blank

An offline fill-in-the-blanks task conducted over the Internet tested the most basic prediction of the model: animate nouns, including potentially metonymic names, should be preferentially interpreted in their literal sense when in subject position. Therefore, readers should be more likely to supply verbs supporting the literal (person) meaning of the name in passive frames, which place the potential metonymy in subject position, than in active frames, which place the potential metonym in object position.

Method

Participants. Forty-six participants were recruited using Amazon's Mechanical Turk who had amassed an approval rate of 98% or greater for 50 or more previous tasks and whose IP address was located within the United States. One participant self-reported as a non-native English speaker and was removed from the analysis. A unique anonymous identifier formed in part from the subject's IP address was created in order to identify participants who attempted to repeat the experiment. No such individuals were found. Subjects were paid $4 for completing the task through Amazon.

[2] We annotated responses as 'Literal' when they referred only to the person, 'Metonymic' when they referred only to the person's works, and 'Ambiguous' otherwise. The responses largely clustered into patterns of relationships to the name, which were analyzed consistently throughout. Responses referring to a person's occupation (e.g. 'poet' in response to 'Maya Angelou') were annotated as 'Literal.' In contrast, we coded references to the category of works the person produced (e.g. 'book' in response to 'Proust') as 'Metonymic.' Responses that were the first name of the person in the manipulation (e.g. 'Mark' in response to 'Twain') were coded as Ambiguous, since in those cases the person's full name has the same metonymic relationships as the last name alone. Responses that were nonsensical or which referred to a person or concept other than the one in our manipulation (e.g. 'Hollywood' in response to 'Marx') were removed.

Materials. Sentence items consisted of 24 pairs of sentences, which manipulated the Voice (Active: 6a vs. Passive: 6b) of the sentence frame with a blank in verbal position.

(6) a. *Active*
As planned, the publisher __________ Kafka shortly after the revisions were in.
b. *Passive*
As planned, Kafka was _________ by the publisher shortly after the revisions were in.

Procedure. Items were counterbalanced and randomly interspersed with 36 unrelated experimental fill-in-the-blank items and 16 non-experimental fillers in a Latin Square design using Ibex Farm. We included 5 high probability fill-in-the-blanks items to remove participants who were unfit for participation in the study. One participant was removed on this basis. In addition, 8 participants were excluded prior to testing for counterbalancing purposes, leaving 32 participants in total. Through a short guided practice, participants were instructed to fill in the blank for each sentence with the first sensical single word that came to mind. After each item, subjects were asked whether they knew who the famous figure was before the experiment. We removed individual responses in case the subject indicated lack of familiarity (96 responses from a total of 528).

Results

We annotated the verbs by what sense interpretation of the metonym was plausible given the supplied verb and the sentential context, according to whether they allowed for only a literal interpretation of the metonym in its sentence (*Literal-only*), only a metonymic interpretation (*Metonymic-only*), or if it allowed for either interpretation (*Ambiguous*).[3] A high percentage of the verbs were annotated as Ambiguous. We report the categorization treatment which most greatly biases against our hypothesis, namely treating ambiguous cases as Literal-only responses. Responses were subjected to a linear mixed effects logistic regression model (Jaeger, 2008), treating Voice as the sole fixed effect variable, and random effects specified as by-items and by-subjects random slopes and intercepts using the lme4 package (Bates & Maechler, 2009) with ANOVA-style sum contrast coding in R (R Development Core Team, 2008).

As predicted by SAP, fewer Metonymic verbs were supplied for Passives ($M = 20\%$, $SE = 3$) than for Actives ($M = 40\%$, $SE = 2$); (Intercept): Coefficient = 1.18, $SE = 0.27$, Wald-$Z = 4.36$, $p < .001$; Voice: Coefficient = -0.8148, $SE = 0.25$, Wald-$Z = -3.30$, $p < .001$. We obtained similar results when we removed the Ambiguous responses altogether.[4]

[3] In order to reduce the possibility of the Verb condition biasing the annotators, we also conducted a much stricter analysis that categorized responses by verb only, without consideration of the sentential context, and observed qualitatively similar results. For example, the verb "contacted" would be annotated according to its lexical semantics as *Literal-only* without regard to the sentence in which it appeared.

[4] The same basic result obtained for all annotation schemes considered, which we have omitted for reasons of space.

Discussion

The experiment lends initial support for the basic prediction that in the absence of other contextually salient cues, the processor preferentially assigns an agent theta role to a noun in canonical subject position – i.e., the first noun in a clause. In the case of potentially metonymic names, an agent assignment eliminates a non-literal interpretation, resulting in fewer Metonymic verbs for passives, which place the metonymic name in subject position, than actives, in which the metonymic name appears in object position. This finding is consistent with the general formulation of SAP in (3). The next two experiments report on how the preference to treat clause initial nouns as agents manifests during online sentence processing.

Experiment 1B: self-paced reading

Materials in Experiment 1B–C manipulated the Voice of the sentence: Active sentences (name in object position; 7a–b) vs. Passive (name in subject position; 7c–d). Additionally, sentences varied on Verb type: crossing verbs subcategorizing most strongly for a literal interpretation of the name (Literal; 7a and 7c), with those subcategorizing for a metonymic interpretation of the name (Metonymic; 7b and 7d). The quartet in (7) shows the regions of analysis in both experiments. In order to distract from the manipulation, and to keep readers engaged, comprehension questions like (8) appeared after each trial with the order of responses individually randomized. No feedback on correctness was provided. Crucially, the questions did not probe the subject's interpretation of the name, as there is independent evidence that so doing might encourage them to abandon a vague representation in favor of a more specific one (Frisson, 2009).

(7) a. *Active Literal verb*
$|_1$ As planned, $|_2$ the publisher $|_3$ contacted $|_4$ Kafka $|_5$ shortly after $|_6$ the revisions were in.
b. *Active Metonymic verb*
$|_1$ As planned, $|_2$ the publisher $|_3$ printed $|_4$ Kafka $|_5$ shortly after $|_6$ the revisions were in.
c. *Passive Literal verb*
$|_1$ As planned, $|_2$ Kafka $|_3$ was contacted $|_4$ by the publisher $|_5$ shortly after $|_6$ the revisions were in.
d. *Passive Metonymic verb*
$|_1$ As planned, $|_2$ Kafka $|_3$ was printed $|_4$ by the publisher $|_5$ shortly after $|_6$ the revisions were in.

(8) What was submitted?
a. Revisions
b. Reviews

Method

Participants. Thirty-three individuals at the Claremont Colleges participated in this experiment. One participant self-identified as a non-native English speaker, and was removed from the analysis. All participants were compensated $10, and did not take part in any of the other experiments presented here.

Materials. Twenty-four sentences were created using the frames from Experiment 1A, with each sentence appearing in four versions varying over the Voice and Verb conditions, as in (7). Prior to testing, the main verbs of the sentences were matched for length (Literal: $M = 8.17$, $SD = 1.37$ vs. Metonymic: $M = 8.04$, $SD = 1.26$) and log frequency measured in terms of Balota et al.'s (2007) Hyperspace Analogue to Language (Literal: $M = 7.63$, $SD = 1.56$ vs. Metonymic: $M = 7.75$, $SD = 1.54$) in paired t-tests, t's < 1.

Procedure. Sentences were presented visually on a 17-in. Dell LCD monitor via a 32-bit Dell Optiplex tower running Windows 7 using Linger software (Rohde, 2003). In order to minimize interference during recording (Plant & Turner, 2009), peripheral programs such as anti-virus software were closed, and the computer was disconnected from the Internet. Participants were instructed to keep their hands resting on a Logitech PS/2 keyboard and to read the text silently for comprehension. Experimental items were interspersed with 52 items from unrelated manipulations and 38 non-experimental fillers, for a total of 114 items per individual experimental session. Participants were encouraged to take breaks in between items.

After completing the reading portion, participants completed an exit survey probing for familiarity with the metonymic names. For example, one such question was, *before this experiment, were you aware that Kafka was an author?* These questions were designed to determine whether the participant's existing knowledge facilitated the producer-for-product metonymy associated with each name during the self-paced reading experiment. Participants answered each question with *Yes*, *No*, or *Not sure*. Subjects were told to answer honestly and that their responses would not be revealed or affect compensation. As noted by an anonymous reviewer, the form of the questions in the exit survey may have yielded confirmation bias. However, analysis computed on the entire data set, disregarding familiarity scores, yields qualitatively similar, though slightly less robust, results.

Results

Prior to analysis, we first examined familiarity responses from the exit questionnaire, which showed a high general familiarity of the metonyms tested across subjects ($M = 90\%$), with no subject indicating less than 75% familiarity with the critical names. We then removed any trial from any condition presented containing a name with which the subject was not familiar, eliminating approximately 11% of the total trials. Although we intended to remove subjects whose performance on the comprehension questions was below 80%, no such subjects were found. We also removed reaction times greater than three standard deviations from the mean for each region, resulting in the elimination of less than 2% of the data. The means and standard errors for all conditions are provided in Table 1.

Responses to the comprehension questions were then subjected to a linear mixed-effects logistic regression model in R. All data presented below, including the experiments that follow, were fit with maximal random effect structures (Barr, Levy, Scheepers, & Tily, 2013) – with by-subjects and by-items random slopes for condition using the lme4 package (Bates & Maechler, 2009) with ANOVA-style sum contrast coding. (Given the lack of consensus on how to best compute p-values in such models, we report t-values above 2 as significant; however, other statistical models, including logistic regressions, are provided with p-values as well.) On average, participants answered the comprehension question correctly 98% of the time. Analysis revealed no effects of Voice, Verb, or interaction between the two, z's < 1.

Next, we analyzed response times for all regions of the target sentence (Region 1 through Region 5). As Active and Passive conditions contained very different types of constituents in the region of interest, we analyzed Passives and Actives separately, each with linear mixed effects regression models region by region, with the single predictor variable of Verb type, and random effects as described above. All significant effects were observed on the post verbal region (Region 4) in Passive sentences: items with Metonymic verbs elicited longer reading times than those with Literal verbs, $t = 2.88$, shown in Table 2. In contrast, we found no difference between items with Metonymic and Literal verbs in any region of the Active sentences. Comparable effects were found with a fully crossed model, with region length added as a predictor. No other effects were observed in any region.

Discussion

The findings discussed above are clearly consistent with the prediction that SAP entices immediate sense selection of a subject-position metonym. Furthermore, we replicated previous findings showing no cost for accessing the metonymic sense of object-position producer-for-product metonyms (Frisson & Pickering, 2001; McElree et al., 2006) with different items and subjects, and with more comprehension questions.

Experiment 1C: eye tracking

The following experiment replicates the findings of the previous experiment with eye movements. We did so for three reasons. First, one could be concerned that the self-paced reading method is too coarse to capture subtle differences between Metonymic and Literal verbs in Active conditions (though see Mitchell, 2004, for a general defense of the self-paced reading method). A second, related reason was that showing a similar penalty for potentially metonymic subjects and Metonymic verbs in eye tracking would give us additional confidence in the self-paced reading results in Experiments 2 and 3. Third, it might be argued that the effects observed in self-paced reading reflect an artificial strategy in which the reader, deprived of parafoveal preview, is forced to interpret each clause more fully upon encountering it (see Clifton et al., 2003, for discussion). That is, the limited window might have prevented subjects from retaining an underspecified representation just when subcategorization information could not determine which sense was required – e.g., in subject but not object position, forcing subjects to default to the Agent reading. If so, the online preference to interpret subject metonyms in their literal sense could reflect

Table 1
Experiment 1B: Means by region and condition in milliseconds (standard errors in parentheses). Outliers, incorrect answers, and unfamiliar metonyms were removed. Significant contrasts marked in italics.

	Region 1 Adverbial	Region 2 Subject	Region 3 Verb	Region 4 Post-verb	Region 5 Spillover	Region 6 Final
Active Literal verb	661 (24)	586 (17)	534 (16)	558 (19)	664 (21)	690 (22)
Active Metonymic verb	670 (26)	574 (17)	530 (15)	582 (20)	646 (20)	700 (23)
Passive Literal verb	692 (29)	553 (16)	540 (15)	*615 (21)*	684 (22)	726 (23)
Passive Metonymic verb	653 (24)	555 (13)	554 (16)	*702 (24)*	716 (24)	723 (23)

Table 2
Experiment 1B: Self-paced reading. Linear mixed effects regression models for Active and Passive conditions on all regions.

		Active				Passive		
		Coef	*SE*	*t*		Coef	*SE*	*t*
Region 1	Intercept	667.417	34.591	19.295*	Intercept	671.815	38.471	17.463*
	Verb	4.592	17.104	0.268	Verb	−18.398	15.815	−1.163
Region 2	Intercept	584.441	28.399	20.58*	Intercept	555.414	21.532	25.795*
	Verb	−4.996	8.979	−0.556	Verb	3.761	9.691	0.388
Region 3	Intercept	539.471	24.642	21.893*	Intercept	547.798	25.097	21.827*
	Verb	−0.193	8.609	−0.022	Verb	9.974	8.986	1.11
Region 4	Intercept	566.538	28.429	19.928*	Intercept	647.675	28.777	22.506*
	Verb	12.989	11.283	1.151	Verb	42.98	14.596	2.945*
Region 5	Intercept	650.855	30.121	21.608*	Intercept	699.388	30.28	23.097*
	Verb	−5.926	13.448	−0.441	Verb	17.858	16.302	1.095
Region 6	Intercept	701.221	29.08	24.113*	Intercept	722.304	28.903	24.991*
	Verb	10.069	14.487	0.695	Verb	−0.756	15.797	−0.048

a task-dependent strategy, rather than an independent grammatical heuristic. Although we think this possibility is unlikely, we address these methodological concerns in Experiment 1C.

In keeping with the findings above, we predicted a processing penalty for Metonymic verbs only in passive constructions, where the processor must interpret a name like *Kafka* without the benefit of prior lexical constraints. Although such a penalty could in principle take many forms in the eye movement record, we expected that it would most likely manifest in 'later' processing measures after encountering the critical verb, such as the time spent in a region before moving past the verb or post verbal region, and the proportion of regressions out of those regions, as are often observed for manipulations of semantic composition (e.g., Pickering, Frisson, McElree, & Traxler, 2004), though, of course, 'early' processing measures have also been implicated in semantic processing, such as the detection of semantic anomalies (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004; see, in addition, Clifton, Staub, & Rayner, 2007, for discussion of the possibly misleading categorization of eye movement measures into early and late measures).

Participants

Thirty-six native English speakers at the Claremont Colleges originally participated in this experiment, but nine were excluded due to excessive eye-blinks or inaccurate calibration, as described below, and were subsequently replaced under the same counterbalancing condition. The final data set distributed subjects evenly across counterbalancing conditions. All participants self-reported English as their native and dominant language.

Materials

The sentence items used during the eye tracking experiment were identical to the items from Experiment 1B, except that a few non-critical typos were corrected. Items were analyzed according to the same 6 regions as before.

Procedure

The experimental items were presented using the UMass Amherst presentation software EyeTrack (http://www.psych.umass.edu/eyelab/) and recorded on an SR Research Eyelink 1000 eye tracker, mounted on the table approximately 50 cm away from a 19" Mitsubishi Diamond Pro 900u flat-screen CRT monitor running at 170 Hz. Materials were presented in a sound isolated room on a 32-bit Dell Optiplex tower, running Windows 7, with peripheral programs and the Internet connection turned off. Text was presented in black 11pt monospaced font against a white background; approximately 3 characters subtended 1 degree of the visual angle. Sampling rate was set to 1000 Hz.

Except for recording method, presentation procedure was identical to Experiment 1B. The experimenter performed a drift correct between each trial. The 24 experimental items were interspersed with 62 experimental items from a number of other studies, as well as 7 practice items and 34 filler items for a total of 127 items for each experimental session. Sentences were presented in individually randomized order, counterbalanced across conditions in Latin Square design. Participants completed an exit survey as in Experiment 1B. Each experimental session lasted approximately 40 min, for which participants were compensated $10.

Results

Individual trials were removed if the participant blinked at least once during the first pass on the verb or the post verbal region (which was the by-phrase for Passive items, and the metonym for Active items), but not if blinks occurred during re-reading. Trials were also removed if excessive blinking led to significant track loss, or if track loss occurred for some other reason during the experiment (approximately 6% of total trials). We removed participants when three or more of any one of the four configurations of an item (e.g. three or more sentences of the Passive Literal verb condition) was removed prior to analysis. These subjects were replaced with other subjects in the same counterbalancing condition.

Additionally, short (under 80 ms) and long (over 800 ms) fixation times were removed from the data (Rayner, 1998), as were blinks on the critical region and track losses using the program EyeDoctor (http://www.psych.umass.edu/eyelab). Several standard eye tracking measures were used in the analysis, computed with the DOS version of EyeDry analysis software: *first pass durations* (also known as *gaze duration*), the sum of all fixation durations within a region before leaving that region in any direction, *go past time*, the time spent after first entering a region to first moving past the region to the right, *percentage of regressions out* of and *percentage of regressions into* a region, *second pass time*, the time spent re-reading a region once the region has been exited to the right including zero times indicating failure to re-read, and *total time*, the sum of all fixation times in a region during any point in reading (see, e.g., Staub & Rayner, 2007, for a concise review of these measures). Means and standard errors for these measures are presented in Table 3.

As before, we removed items with metonymic names with which subjects were unfamiliar on the basis of the exit questionnaire ($M = 87\%$ familiarity). We did not analyze or remove incorrect answers from the data set, given the high accuracy rate of Experiment 1B. However, observations over three standard deviations from the mean for each region were removed from typically normally distributed measures (here, first pass times). As before, passive and active constructions were first analyzed separately, region-by-region, as before, due to incomparable differences between the regions of interest. For each measure and each region, we first fit a linear mixed-effects regression model with Verb type as the fixed effect predictor to Actives and Passives separately. Although the vast syntactic differences between the regions in actives and passives, which cannot simply be captured by adding region length into the model, make splitting the data by Voice the most accurate analysis, an additional model with Voice, Verb, and their interaction as predictors, along with region length was computed for each measure in each region. As we found essentially the same patterns, this analysis is omitted.

As we predicted that Metonymic verbs would elicit processing costs in passive, but not active, constructions, we first present regions at which the processor had encountered the main verb, i.e., regions 3–6, concentrating initially on go past and regression measures. We then briefly report any other effects observed at any other point in processing. As no significant results were observed on first pass durations for any region, they are omitted here.

A differential effect for Metonymic verbs on passive conditions was observed in the expected go past measure and proportions of regressions out of the post verbal region of Passive conditions, *by the publisher*. For go past times, a 137 ms penalty for Metonymic verbs was observed, $t = 3.38$. The expected processing penalty for Passives was again observed in a higher percentage of regressions out following any type of fixation of the post-verbal region following a Metonymic verb for Passives ($d = 16\%$) than for Actives ($d = 7\%$), $z = 3.36$, $p < .001$. In addition, readers made more regressions into the verb region (Region 3) for Metonymic over Literal verbs when in Passive ($d = 18\%$) as compared to Active constructions ($d = 8\%$), $z = 2.88$, $p < .01$. In contrast, Metonymic verbs ($M = 32\%$, $SE = 4$) elicited marginally *fewer* regressions into the region containing the subject (Region 2) than Literal verbs ($M = 44\%$, $SE = 4$), $z = 1.77$, $p = .08$. The general finding of a processing difficulty for sentences with Metonymic verbs and metonyms in subject, but not object, position supports the central prediction of SAP: a producer-for-product metonym is interpreted literally at first when in subject position.

Additional support for this interpretation comes from measures of second pass re-reading times. In the verb region, Metonymic verbs elicited longer second pass times (in which 0 values were retained) than Literal verbs in the Passive ($d = 99$ ms), $t = 3.44$. While there was a 53 ms penalty on the verb for Metonymic verbs in the Active conditions, the effect did not reach significance, $t = 1.75$. In the following region, a similar pattern for second pass times emerged: the 74 ms penalty for Metonymic verbs in Passive conditions was significant, $t = 2.55$, while the 46 ms cost in Active conditions was marginally significant, $t = 1.93$.

Finally, the cost for processing Metonymic verbs again manifested in total reading times for Passive, but not Active, conditions in verbal and post-verbal regions. In Region 3, Metonymic verbs elicited longer reading times than did Literal verbs, $t = 3.97$. The pattern continues for Region 4, in which Metonymic verbs elicited longer total reading times than Literal verbs did, $t = 3.48$. As there were no effects observed in first pass duration, the differences reported in total reading times are likely due to second pass differences discussed above. No other significant effects were observed (see Tables 4 and 5).

Discussion

All told, we find a good deal of evidence to support the central prediction that SAP guides early stages of metonym processing. The patterns observed in both the eye movements and self-paced reading studies were consistent with our prediction that SAP tempts the processor to immediately commit to the literal sense of a possibly metonymic name like *Kafka* in subject position, in keeping with the preference for Literal verbs observed in the offline fill-in-the-blanks task. Reading costs were observed when readers encountered a verb selecting for the metonymic sense of a subject metonym, suggesting that they had already

Table 3
Experiment 1C: Eye tracking study. Means by region and condition in milliseconds (or percentages for regression data) for all eye movement measures collected. (Standard errors in parentheses.) Outliers and unfamiliar metonyms were removed. Significant contrasts marked in italics.

	Region 1 Adverbial	Region 2 Subject	Region 3 Verb	Region 4 Post-verb	Region 5 Spillover	Region 6 Final
Go past times						
Active Literal verb	303 (21)	436 (24)	431 (20)	348 (23)	665 (41)	1085 (78)
Active Metonymic verb	320 (20)	408 (19)	417 (20)	404 (27)	817 (61)	1017 (78)
Passive Literal verb	282 (19)	433 (18)	433 (19)	*525 (22)*	589 (29)	1240 (87)
Passive Metonymic verb	307 (20)	461 (19)	461 (19)	*662 (28)*	674 (37)	1290 (93)
Regressions out						
Active Literal verb	0 (0)	12 (3)	14 (3)	22 (3)	15 (3)	61 (4)
Active Metonymic verb	0 (0)	10 (2)	12 (3)	29 (4)	26 (3)	63 (4)
Passive Literal verb	0 (0)	15 (3)	14 (3)	*16 (3)*	13 (3)	58 (4)
Passive Metonymic verb	0 (0)	14 (3)	14 (3)	*32 (4)*	19 (3)	57 (4)
Regressions in						
Active Literal verb	69 (4)	48 (4)	34 (4)	16 (3)	34 (4)	0 (0)
Active Metonymic verb	55 (5)	39 (4)	42 (4)	24 (3)	31 (3)	0 (0)
Passive Literal verb	59 (5)	30 (4)	*35 (4)*	20 (3)	43 (4)	0 (0)
Passive Metonymic verb	50 (5)	34 (4)	*53 (4)*	28 (3)	34 (4)	0 (0)
Second pass times						
Active Literal verb	160 (17)	196 (18)	176 (18)	83 (10)	190 (23)	0 (0)
Active Metonymic verb	170 (19)	176 (16)	229 (19)	129 (15)	199 (20)	0 (0)
Passive Literal verb	160 (18)	101 (12)	*175 (18)*	*150 (18)*	228 (23)	0 (0)
Passive Metonymic verb	148 (19)	119 (16)	*274 (23)*	*224 (23)*	202 (23)	0 (0)
Total times						
Active Literal verb	505 (31)	593 (28)	584 (24)	414 (20)	759 (36)	636 (31)
Active Metonymic verb	528 (33)	583 (27)	634 (30)	517 (31)	792 (30)	586 (26)
Passive Literal verb	486 (27)	364 (17)	*561 (25)*	*637 (28)*	794 (37)	614 (32)
Passive Metonymic verb	487 (28)	398 (22)	*738 (34)*	*774 (30)*	805 (34)	616 (28)

committed to its literal sense by the time they had reached the disambiguating verb in the next region.

Nevertheless, the penalty manifested at different places in the sentence in the two experiments: solely on the post verbal region in self-paced reading, and on both verb and post verbal regions in eye tracking, raising the interesting issue of what prompts the putative 'reanalysis' of the Agent role assignment. We discuss a few possibilities in the General Discussion.

Turning to object position metonyms, we found no processing penalties for the metonymic sense vs. the literal sense, except for marginal costs on the verb and post-verbal regions in re-reading times. The differential processing pattern is entirely expected under an account which can delay committing to a specific sense, except when a forced by a grammatical decision (Frazier, 1999). However, it is possible that the ease with which the processor selects metonymic interpretations is ultimately due to 'digging-in' effects (e.g., Ferreira & Henderson, 1991a, 1991b; Tabor & Hutchins, 2004), so that the more frequent or familiar sense is selected first (Foraker & Murphy, 2012; Klein & Murphy, 2002).[5] The online penalty observed in the above experiments may reflect the processor increasing its commitment to a more familiar literal sense over the course of reading the sentence. That is, without lexical or contextual information, the processor need not revise its commitment to the literal sense of a metonym in subject position until reaching contradictory information in the verb. Metonyms in object position, in contrast, receive information from the verb that allows the processor to forgo such a commitment, leaving it relatively unattached to the literal sense.

Although we believe that this explanation is unlikely given that the sense norming study failed to show a consistent preference for the literal sense, as well as previous findings in the literature (e.g., Frisson & Frazier, 2004; see the General Discussion section), a digging-in account is perhaps worth considering in light of the marginal re-reading costs for metonymic interpretations in actives. Note that a general cost for Metonymic verbs is not inconsistent with our basic account: the processor might encounter greater difficulty when incorporating metonymic or other non-literal representations into a semantic interpretation or situation model of the sentence. Yet, the precise effects observed with grammatical position are not, to our knowledge, predicted by any other account (though see Lowder & Gordon, 2013, for evidence that metonyms in adjunct positions are processed differently than those in argument positions).

Further, the role of grammatical voice is not necessarily anticipated in a model of metonymic processing which treats different metonymic senses as distinct nodes within the lexicon (e.g., Klein & Murphy, 2002). However, the results might still be consistent with a single stage model, given that we have dealt exclusively with *regular* metonymy. For it might well be that the literal sense for regular polysemes, including the producer-for-product metony-

[5] Thanks to Chuck Clifton for raising the possibility of a digging-in alternative.

Table 4
Experiment 1C: Eye tracking study. Linear mixed effect regression model results for go past, second pass, and total times. Values for the coefficient, standard error, and *t*-statistic for the main effect of Verb type are presented for Active and Passive voice by region. An asterisk * indicates that the effect is significant at the α = 0.05 level on the criterion that |*t*| > 2. A plus + indicates that the effect is marginally significant on the criterion that |*t*| > 1.9. Region 6 is omitted from the second pass times, since the second pass measure does not apply to the final region.

		Active				Passive		
		Coef	*SE*	*t*		Coef	*SE*	*t*
Go past								
Region 1	Intercept	311.118	34.117	9.119*	Intercept	302.667	33.643	8.996*
	Verb	0.329	14.301	0.023	Verb	15.562	11.067	1.406
Region 2	Intercept	422.847	29.228	14.467*	Intercept	281.225	22.504	12.497*
	Verb	−6.664	18.014	−0.37	Verb	7.017	14.706	0.477
Region 3	Intercept	420.371	26.207	16.04*	Intercept	451.459	20.932	21.568*
	Verb	−10.703	20.736	−0.516	Verb	13.367	15.324	0.872
Region 4	Intercept	376.324	37.841	9.945*	Intercept	598.285	32.759	18.263*
	Verb	10.71	23.195	0.462	Verb	67.297	19.924	3.378*
Region 5	Intercept	732.785	62.439	11.736*	Intercept	631.261	39.185	16.11*
	Verb	59.078	45.123	1.309	Verb	36.993	24.241	1.526
Region 6	Intercept	1081.9	121.911	8.874*	Intercept	1290.896	125.117	10.318*
	Verb	−2.775	77.431	−0.036	Verb	22.375	56.365	0.397
Second pass								
Region 1	Intercept	169.542	27.783	6.102*	Intercept	156.056	25.116	6.214*
	Verb	9.405	13.12	0.717	Verb	−5.137	16.057	−0.32
Region 2	Intercept	187.823	21.879	8.584*	Intercept	113.8	21.131	5.386*
	Verb	−6.579	12.163	−0.541	Verb	10.038	12.655	0.793
Region 3	Intercept	203.297	20.055	10.137*	Intercept	229.129	26.505	8.645*
	Verb	26.849	15.39	1.745	Verb	51.47	14.976	3.437*
Region 4	Intercept	106.676	17.121	6.231*	Intercept	189.459	25.14	7.536*
	Verb	23.7	12.27	1.932+	Verb	36.193	14.198	2.549*
Region 5	Intercept	198.431	29.351	6.761*	Intercept	221.947	30.232	7.342*
	Verb	11.317	15.135	0.748	Verb	−9.845	17.481	−0.563
Total time								
Region 1	Intercept	508.593	50.764	10.019*	Intercept	480.425	45.279	10.61*
	Verb	11.247	26.138	0.43	Verb	9.772	18.549	0.527
Region 2	Intercept	591.414	44.521	13.284*	Intercept	386.882	30.695	12.604*
	Verb	11.981	25.432	0.471	Verb	16.572	15.308	1.083
Region 3	Intercept	603.116	36.614	16.472*	Intercept	662.223	40.983	16.159*
	Verb	19.443	27.157	0.716	Verb	91.336	22.99	3.973*
Region 4	Intercept	458.326	40.267	11.382*	Intercept	716.74	43.086	16.635*
	Verb	33.43	29.794	1.122	Verb	63.275	18.197	3.477*
Region 5	Intercept	777.14	53.466	14.535*	Intercept	802.903	57.607	13.937*
	Verb	22.388	30.143	0.743	Verb	2.868	25.338	0.113
Region 6	Intercept	610.981	36.837	16.586*	Intercept	618.65	47.312	13.076*
	Verb	−21.003	23.7	−0.886	Verb	−5.53	22.617	−0.245

mies examined here, are sufficiently, but perhaps not singularly, activated by sentential structure. That is, assigning an Agent role to the metonym might activate the literal sense by virtue of sharing an animate feature.[6] Many more types of metonymic relationships would need to be tested to determine whether all metonyms are equally subject to grammatical heuristics like SAP, but the possibility remains open that the language processor utilizes just one general strategy for processing all types of metonyms. We leave this important task open for future study. Importantly, the above results effectively rule out the more constrained version of the Underspecification Model in which only prior lexical information like subcategorization requirements imposed by a verb can lead to immediate sense selection. We now turn to assessing the weak and strong variants of SAP discussed in the Introduction.

[6] Many thanks to Andreas Brocher for bringing the issue of regular metonymy to our attention.

Experiment 2

Insofar as the studies in Experiment 1 indicated that SAP promotes the literal sense of a subject position metonym, the results cannot adjudicate whether SAP operates independently of more general discourse context – that is, whether SAP is affected by general context (4) or not (5). In order to address this question, we investigated how biasing information from context affects subject-position metonyms by considering two types of context: weakly constraining information from prior text (Experiment 2) and more strongly constraining information from topic setting clauses immediately preceding the subject (Experiment 3). In both cases, we observe a similar processing penalty as in Experiment 1 above, except that the effects appeared to be somewhat weaker, consistent with a stronger, context-insensitive conception of SAP (5).

Our items in Experiment 2 (see Appendix B) consisted of context sentences which either biased towards the metonymic sense of the metonym (9a) or were neutral (9b),

Table 5
Experiment 1C: Eye tracking study. Linear mixed effect logistic regression model results for percentage of regressions in and out of a region. Values for the coefficient, standard error, and Wald-*Z* for the main effect of Verb type are presented for Active and Passive voice by region, along with the *p*-value.

		Active					Passive			
		Coef	*SE*	Wald-*Z*	*p*-Value		Coef	*SE*	Wald-*Z*	*p*-Value
Regressions out										
Region 2	Intercept	−2.204	0.199	−11.099	<.001***	Intercept	−2.252	0.26	−8.67	<.001***
	Verb	−0.071	0.184	−0.384	.701	Verb	−0.08	0.268	−0.298	.765
Region 3	Intercept	−2.069	0.222	−9.314	<.001***	Intercept	−1.901	0.175	−10.852	<.001***
	Verb	−0.104	0.174	−0.597	.551	Verb	0.035	0.169	0.208	.835
Region 4	Intercept	−1.281	0.203	−6.313	<.001***	Intercept	−1.366	0.182	−7.518	<.001***
	Verb	0.169	0.169	1.001	.317	Verb	0.555	0.165	3.355	<.01**
Region 5	Intercept	−1.644	0.243	−6.759	<.001***	Intercept	−1.878	0.195	−9.61	<.001***
	Verb	0.266	0.15	1.775	.076	Verb	0.238	0.217	1.093	.274
Region 6	Intercept	0.569	0.189	3.007	<.01**	Intercept	0.426	0.207	2.054	<.05*
	Verb	0.017	0.133	0.126	.9	Verb	0.009	0.138	0.062	.951
Regressions in										
Region 1	Intercept	0.527	0.34	1.549	.121	Intercept	0.234	0.475	0.492	.623
	Verb	−0.275	0.171	−1.607	.108	Verb	−0.355	0.258	−1.375	.169
Region 2	Intercept	−0.296	0.188	−1.576	.115	Intercept	−0.824	0.183	−4.494	<.001***
	Verb	−0.205	0.116	−1.771	.077	Verb	0.085	0.137	0.62	.536
Region 3	Intercept	−0.614	0.231	−2.663	<.01*	Intercept	−0.292	0.217	−1.345	.179
	Verb	0.183	0.179	1.02	.308	Verb	0.451	0.157	2.881	<.01**
Region 4	Intercept	−1.567	0.2	−7.826	<.001***	Intercept	−1.229	0.151	−8.13	<.001***
	Verb	0.203	0.154	1.319	.187	Verb	0.21	0.143	1.467	.142
Region 5	Intercept	−0.898	0.254	−3.529	<.001***	Intercept	−0.516	0.191	−2.697	<.01**
	Verb	−0.011	0.138	−0.08	.936	Verb	−0.247	0.165	−1.5	.134

+ Significance level: $p < .1$.
*** Significance level: $p < .001$.
** Significance level: $p < .01$.
* Significance level: $p < .05$.

followed by sentences which were closely related to our items from Experiment 1. In this experiment, the initial adverbial phrase was replaced by a sentence establishing discourse context. The regions of interest (Regions 3–6) are similar to the Passive conditions in Experiment 1. Questions like (10) followed each item, and were designed to relate the context sentence to the target sentence, without probing the interpretation of the potentially metonymic name.

(9) a. *Metonymic context*
$|_1$ In 1781, Vienna hosted a music festival where many popular orchestral pieces were played. |
b. *Neutral context*
$|_1$ In the spring of 1781, a wonderful gala was held at the palace in Vienna. |
Target sentence with {Literal / Metonymic} verb
$|_2$ Mozart $|_3$ was {invited / conducted} $|_4$ by Vienna's orchestra conductor, $|_5$ which was $|_6$ a great honor.

(10) Where was the event held?
i. Vienna
ii. Paris

On Frazier's (1999) account, the processor may delay commitment to a specific sense if no grammatical (i.e., lexical or syntactic) decision requires it. As supporting prior context is highly informative, but not grammatical, we predicted that Metonymic contexts like (9a) would not force immediate sense selection of the metonym – especially if it would contradict SAP. Nevertheless, we expected that the effects of that biasing information would ease processing when the language processor was forced to revise its initial sense selection.

Participants

Fifty-two participants were recruited from the Claremont Colleges for participation in this experiment. Two subjects self identified as non-native speakers of English and were removed from the analysis. Participants were paid $10 as compensation.

Materials

Items each consisted of a context sentence and a target. The first sentence always provided a general context for the second sentence, and the second sentence was nearly identical to the sentence form of the passive sentence items in Experiment 1. The experimental design crossed Contextual bias (Metonymic or Neutral) by Verb type (Literal or Metonymic), as in Experiment 1). Prior to testing, 16 subjects rated 26 context–target pairs of items for naturalness on a 7-point scale (7 = completely natural). We found no differences with respect to naturalness, t's < 1.5, suggesting that any effects in reading times cannot be attributed to independent differences in goodness of fit between context sentences and targets (Metonymic context–Metonymic verb: $M = 5.20$, $SE = 0.16$; Metonymic context–Literal verb: $M = 5.09$, $SE = 0.16$; Neutral context–Metronymic verb: $M = 4.84$, $SE = 0.17$; Neutral context–Literal verb: $M = 5.12$, $SE = 0.15$). Twenty-four items balanced for naturalness were selected for presentation in the reading experiment. Differences in length and HAL frequency for critical verbs were controlled for prior to testing: in paired t-tests, Literal verbs (Length: $M = 8.17$, $SD = 1.55$; Frequency: $M = 7.31$, $SD = 1.24$) did not differ from Metonymic verbs (Length: $M = 8.25$, $SD = 1.3$; Frequency: $M = 7.77$, $SD = 1.53$), t's < 1.5.

Procedure

The self-paced reading procedure in Experiment 2 was virtually identical to the procedure in Experiment 1B (and was tested on the same computer under the same testing conditions described above), except that there was no exit survey to determine subject's familiarity with the names mentioned in the experimental items. As the majority of the names overlap with names in Experiment 1, and subjects from the same population were largely familiar with those names, we suspect that the names were mostly familiar to our subjects.

Results

Prior to analysis, we removed the data from two participants whose accuracy on comprehension questions in the experiment was less than 80%. These subjects were then replaced by two more in the same counterbalancing condition, for a completely counterbalanced data set. We then examined participants' general comprehension of the sentence items by analyzing question response data. Participants answered comprehension questions correctly on approximately 94% of trials. Data from item 12 were excluded due to presentation error. Finally, we removed all reaction times above three standard deviations from the mean for each region, resulting in less than 2% data loss.

Across all the items, we analyzed response times to all the regions of the target sentence (Region 1 through Region 5) by subjecting the data to linear mixed effects regression modeling, treating sum coded Context, Verb type, and their interaction as fixed effects, and by-subject and by-item random slopes and intercepts in R. All significant effects from all analyses that were run are reported here. Means and standard deviations are presented in Table 6; results of the linear mixed effects regression models are presented in Table 7.

Prior context had no effect on the processing times on either the subject or verb region. However, on the post verbal region (Region 4), there are several effects to report. First, biasing context elicited shorter reading times than did neutral contexts, $d = 83$ ms, $t = -3.73$, indicating that subjects were attuned to the contexts provided. Second, there was a 115 ms penalty for Metonymic verbs over Literal ones, $t = 4.14$, as predicted by SAP. Although there was a greater penalty for Metonymic verbs in Neutral contexts ($d = 142$ ms) than in Metonymic contexts ($d = 88$ ms), the effect did not approach significance, $t = 1.12$. There were no other effects to report.

Discussion

The results obtained in this experiment support a stronger version of SAP (5) over a weaker alternative (4). That is, biasing contextual information did not remove the temptation to interpret the clause initial noun as a subject. Rather, it seems that SAP likely places powerful, if defeasible, constraints on initial interpretation, plausibly operating as a predictive mechanism independent of general contextual concerns. We now turn to a different contextual manipulation to determine whether the biasing contexts above were somehow too weak. Despite making every effort to provide sufficiently biasing contexts that would interfere with initial sense-selection, we found that such contexts fail to eliminate effects consistent with an early commitment to the literal sense of clause initial metonyms, as predicted by the strong variant of SAP.

Experiment 3

In the following experiment, subject metonyms were preceded by topic-setting clauses that provided an explicit literal or metonymic interpretation. For example, in (11) the topic-setting clause *As for* provided either a metonymic works-of-individual (*plays*; 11a) or a literal-individual (*playwrights*; 11b) context explicitly, again manipulating Verb type between Literal (*honored*) and Metonymic (*staged*), followed by comprehension questions as in the previous studies.[7]

(11) a. *Metonymic topic with {Literal / Metonymic} verb*
$|_1$ As for plays, $|_2$ Shakespeare $|_3$ was {honored / staged} $|_4$ by the actors $|_5$ at the celebration $|_6$ of the arts.
b. *Literal topic with {Literal / Metonymic} verb*
$|_1$ As for playwrights, $|_2$ Shakespeare $|_3$ was {honored / staged} $|_4$ by the actors $|_5$ at the celebration $|_6$ of the arts.

If SAP is indeed influenced by general context at early processing stages and those contexts in Experiment 2 were simply too weak to induce an effect, then we would expect that the cost of Metonymic verbs observed in the previous experiments would be eliminated, or at least greatly reduced, in the presence of a metonymic context (11a) compared to (11b). However, if SAP is a grammatical principle operating as a powerful default independent of discourse context, we expect that such clauses will fail to eliminate the relevant processing costs. Further, the manipulation allows us to replicate Experiment 2 with different subjects and modified items.

Methods

Participants

Forty-three participants from the Claremont Colleges were recruited for the study, which lasted approximately

[7] We thank Lyn Frazier for suggesting this manipulation.

Table 6
Experiment 2: Means by region and condition in milliseconds. (Standard errors in parentheses) Outliers and incorrect answers removed. Significant contrasts marked in italics.

	Region 1 Context	Region 2 Subject	Region 3 Verb	Region 4 Post-verb	Region 5 Spillover	Region 6 Final
Context–verb type						
Metonymic context–Literal verb	4169 (120)	580 (14)	502 (13)	*597 (18)*	598 (14)	789 (25)
Metonymic context–Metonymic verb	4105 (130)	574 (14)	490 (13)	*685 (25)*	611 (14)	752 (24)
Neutral context–Literal verb	3885 (137)	607 (16)	502 (13)	*653 (22)*	606 (15)	797 (27)
Neutral context–Metonymic verb	3733 (119)	583 (15)	514 (14)	*795 (28)*	637 (16)	751 (21)

Table 7
Experiment 2: Self-paced reading. Values for the coefficient, standard error, and t statistic for the linear mixed effects regression models computed with Context, Verb, and their interaction Context:Verb as fixed effects. An asterisk * indicates that the effect is significant at the $\alpha = 0.05$ level on the criterion that $|t| > 2$. Region 1, which contained the context, is omitted.

		Coef	*SE*	*t*
Region 2	Intercept	590.717	22.048	26.792*
	Context	−8.456	6.882	−1.229
	Verb	−6.865	6.071	−1.131
	Context:Verb	3.581	6.29	0.569
Region 3	Intercept	506.171	21.573	23.464*
	Context	−6.023	5.148	−1.17
	Verb	2.615	5.575	0.469
	Context:Verb	−4.562	5.597	−0.815
Region 4	Intercept	690.331	42.257	16.336*
	Context	−38.048	10.215	−3.725*
	Verb	58.87	14.23	4.137*
	Context:Verb	−11.67	10.425	−1.119
Region 5	Intercept	616.137	22.484	27.403*
	Context	−8.331	7.203	−1.157
	Verb	12.169	7.981	1.525
	Context:Verb	−4.756	8.424	−0.565
Region 6	Intercept	783.455	39.053	20.061*
	Context	−3.951	10.346	−0.382
	Verb	−20.619	13.916	−1.482
	Context:Verb	−1.129	11.713	−0.096

30 min. Participants were paid $10 as compensation. Five participants self identified as non-native speakers and were removed from analysis. Two more were removed for not properly understanding the experimental task. The remaining subjects were evenly distributed across counterbalancing conditions.

Materials

We tested 24 sentence quartets like (11), presented in 6 moving windows, as above. Each sentence was followed by a comprehension question as in previous experiments, which did not probe the interpretation of the famous name. Differences in length and log HAL frequency of critical verbs were controlled for prior to testing: in paired t-tests, Literal verbs (Length: $M = 8.38$, $SD = 1.35$; Frequency: $M = 7.68$, $SD = 1.61$) did not differ from Metonymic verbs (Length: $M = 8.08$, $SD = 1.35$; Frequency: $M = 7.72$, $SD = 1.40$), t's < 1.

Prior to testing, we presented the items with verbs removed to 28 subjects recruited on Amazon's Mechanical Turk in a fill-in-the-blanks task, with a procedure nearly identical to Experiment 1A. As before, many responses were potentially ambiguous. In an analysis that categorized Ambiguous responses as Literal, thus biasing against our hypothesis, Metonymic topics (16a; $M = 60\%$, $SE = 3$) elicited more Metonymic verb responses than did Literal topics (16b; $M = 51\%$, $SE = 3$); (Intercept): Coefficient = 0.4615, $SE = 0.3265$, Wald-$Z = 1.414$, $p = .15744$; Topic: Coefficient = −0.5640, $SE = 0.2159$; Wald-$Z = -2.613$, $p < .001$. Though significant, the size of the effect is small (10–15%, depending on scoring). Additionally, items were normed in a separate acceptability rating study. Twenty subjects rated the items with relatively high acceptability ($M = 5.56$; 7 = completely acceptable), with no significant differences between conditions (condition means were within ±0.20 points of the grand mean).

Procedure

The experimental procedure was the same as Experiment 1B, and was conducted on the same computer using an identical recording procedure. Items were interspersed with 62 items from other experiments and 20 non-experimental fillers, for a total of 106 items per experimental session, presented in individually randomized order. After the self-paced reading section, participants completed an exit questionnaire establishing their familiarity with names presented in the previous section.

Results

In the exit questionnaire, familiarity with the metonymic names was generally high at 88%, although 3 participants were familiar with fewer than 75% of the names. We removed trials presenting unfamiliar names to participants on an individual basis, as in Experiment 1B–C, resulting in less than 12% data loss. In addition, participants performed well on the comprehension questions following the items in self-paced reading, with 95% accuracy overall, and no differences between conditions were observed. Each subject answered the comprehension questions with greater than 80% accuracy. We removed outliers above three standard deviations away from the mean of the region. Less than 3% of the data was eliminated for each region. Means and standard errors are presented in Table 8.

Reading time data for each region were then subjected to a linear mixed effects regression model treating sum-coded Topic, Verb type, and their interaction as fixed effects, and random effects as before. No differences between conditions or interactions were observed prior to the verb region (Region 3), where Metonymic verbs elicited marginally longer reading times than Literal verbs, $t = 1.95$, regardless of the topic-setting context. In the post verbal

Table 8
Experiment 3: Self-paced reading. Means by region and condition in milliseconds. (Standard errors in parentheses.) Outliers and incorrect answers removed. Significant contrasts marked in italics.

	Region 1 Topic clause	Region 2 Subject	Region 3 Verb	Region 4 Post-verb	Region 5 Spillover	Region 6 Final
Metonymic topic, Literal verb	1083 (41)	572 (15)	527 (16)	*679 (23)*	785 (28)	783 (28)
Metonymic topic, Metonymic verb	1085 (39)	557 (15)	582 (18)	*725 (25)*	750 (24)	780 (26)
Literal topic, Literal verb	1005 (35)	556 (15)	577 (19)	*682 (22)*	822 (29)	786 (26)
Literal topic, Metonymic verb	1110 (43)	563 (13)	583 (20)	*747 (26)*	818 (28)	772 (27)

region (Region 4), there was a fully significant processing cost for Metonymic verbs over Literal counterparts, $t = 2.57$. However, there was no discernable effect of topic or an interaction until Region 5, in which Literal contexts elicited marginally longer reading times ($d = 53$ ms) than Metonymic contexts, $t = 1.94$, although no other effects obtained in that region. We fit the data again with half of the items that showed the strongest bias against the literal sense from the fill-in-the-blanks norming task above – that is, items for which the manipulation resulted in the greatest increase for Metonymic verbs, again finding the simple effect of increased reading times for Metonymic ($M = 754$ ms, $SE = 23$) over Literal ($M = 679$ ms, $SE = 21$) verbs regardless of topic; (Intercept): Coefficient = 731.53, $SE = 35.70$, $t = 20.49$; Topic: Coefficient = −6.90, $SE = 16.62$, $t = -0.42$; Verb: Coefficient = 38.24, $SE = 16.14$, $t = 2.37$; Topic:Verb interaction: Coefficient = 21.43, $SE = 16.40$, $t = 1.30$. No other significant effects were observed (see Table 9).

Discussion

The experiment replicated results from the previous experiments, in which Metonymic verbs elicited a processing cost for metonyms in subject position (Experiment 1) in spite of contextual bias (Experiment 2). A plausible explanation for the small effect of contextual bias in Experiment 2 was that our contexts were weak and indirect – readers may simply not have bothered to integrate the target sentences into the text. However, this argument is not supported by the present findings: even biasing contexts, occurring *within* the sentence did not overturn the preference for a literal interpretation of the subject position metonym. The result is compatible with a strong version of SAP (5), which, though ultimately defeasible, persists even in the presence of strong contextual bias, shown in an offline study to reduce the bias towards the literal interpretation of subject metonyms compared to Experiment 1A.

Table 9
Experiment 3: Self-paced reading. Values for the coefficient, standard error, and *t* statistic for the linear mixed effects regression models computed with Topic, Verb, and their interaction Topic:Verb as fixed effects. An asterisk * indicates that the effect is significant at the $\alpha = 0.05$ level on the criterion that $|t| > 2$. A plus + indicates that the effect is marginally significant on the criterion that $|t| > 1.9$.

		Coef	*SE*	*t*
Region 1	Intercept	1080.602	58.954	18.33*
	Topic	−15.787	25.703	−0.614
	Verb	20.899	19.635	1.064
	Topic:Verb	24.164	17.537	1.378
Region 2	Intercept	575.036	24.162	23.799*
	Topic	−3.862	5.845	−0.661
	Verb	−5.223	6.275	−0.832
	Topic:Verb	9.265	6.14	1.509
Region 3	Intercept	574.694	30.594	18.785*
	Topic	12.31	7.714	1.596
	Verb	13.645	7	1.949+
	Topic:Verb	−8.379	8.261	−1.014
Region 4	Intercept	719.825	35.859	20.074*
	Topic	4.137	11.854	0.349
	Verb	27.735	10.799	2.568*
	Topic:Verb	8.566	10.874	0.788
Region 5	Intercept	797.007	43.061	18.509*
	Topic	21.33	10.984	1.942+
	Verb	−14.027	14.181	−0.989
	Topic:Verb	9.884	12.327	0.802
Region 6	Intercept	789.501	41.041	19.237*
	Topic	−5.195	10.88	−0.477
	Verb	−6.197	13.02	−0.476
	Topic:Verb	4.381	12.222	0.358

General discussion

According to the Underspecification Model, the language processor may elect to forgo selecting a specific sense upon encountering lexical items with multiple related senses, instead opting to leave a vague or underspecified representation, consistent with some or even all of its senses. As the factors prompting sense selection are not yet completely understood, the model might be developed in multiple ways. One possibility is that the processor might delay, or even ignore, sense selection unless a specific sense is required by lexical selection. Another is that the decision to select a more specific sense is sensitive to grammatical heuristics, assigning a default interpretation on the basis of a word's structural position.

Results from three experiments support a version of the Underspecification Model in which structural defaults like SAP, in which the processor provisionally assigns an Agent theta role to a noun likely to be in subject position, prompt the language processor to home in on a specific sense of a metonym regardless of previous context. In Experiment 1, we found that the processor encountered difficulty when forced to interpret a subject-position metonym using its metonymic sense, but had no such difficulty when that same metonym was in object position. This finding is clearly consistent with SAP. Since thematic role assignment places grammatical constraints on interpretation, the processor would be forced to make a grammatical decision when assigning an Agent role to the metonym, and so would need to home in on the literal sense of the metonym.

Experiments 2 and 3 investigated the influence of context supporting the metonymic sense. In both experiments, there was still a penalty for processing the metonymic sense of a subject-position metonym when presented after contextual information biasing towards a metonymic interpretation; nevertheless, the penalty was reduced compared to neutral preceding contexts (Experiment 2) or topic phrases biasing towards a literal interpretation (Experiment 3). The results together suggest a stronger variant of SAP, in which the Agent theta role is assigned to the subject noun even in the presence of biasing context. This is not to say, however, that other factors such as lexical bias or even discourse context as a whole are wholly irrelevant in the sense selection process. Far from it, the results of compatible with multiple models in which such factors might interact. Nevertheless, we have presented initially evidence that a structural default like SAP is a particularly strong factor in the process.

On the merit of these results, we propose that there is a general structure to what types of information can lead to immediate sense selection during metonym interpretation: when sufficiently constraining information makes separate interpretations of a word incompatible, the processor must make an immediate decision (Frazier, 1999). In the face of more weakly constraining information, on the other hand, the processor may delay arbitrary semantic decisions until provided with subsequent information or decide not to resolve them at all. We suspect that, broadly speaking, the more reliable the information is, the more likely it is to be used at earlier stages of sentence processing. To speculate: if, as in the case of Experiments 2 and 3, the processor elects to follow a grammatical heuristic over more general, contextual information, it may respond to the need to make an immediate grammatical decision by employing the most reliable information it has at its disposal at the time. This view at least has the advantage of plausibly uniting the constraints imposed by individual lexical items and more general structural heuristics. At any rate, this general structure for ordering constraining information with respect to the metonyms studied here is compatible with the Underspecification Model.

Another alternative model exists, however, which would accurately predict our findings in the present experiment. The above results are consistent with a model in which the processor immediately defaults to the literal sense of a metonym if there are not lexical constraints which would otherwise guide interpretation of that metonym. To an extent, this default-to-literal model is an adaptation of the literal-first model presented in the theoretical literature (e.g. Grice, 1975; Searle, 1979), with the qualification that the processor only defaults to the literal sense in cases when lexical constraints do not guide interpretation.

While our results do not, strictly speaking, rule out such an alternative, other experimental findings cast considerable doubt on its general plausibility. For example, Frisson and Frazier (2004) investigated processing polysemous words like *book* in the absence of prior information biasing towards a specific sense interpretation. There was no processing cost after a verb requiring the subordinate sense (*the book looked tattered*) as compared to one supporting the dominant sense (*the book looked enjoyable*), suggesting that when the language processor reached the polysemous noun, it had maintained an underspecified representation of the word (see Frisson, 2009, for description). Crucially, the default-to-literal model incorrectly predicts that when the processor encounters a possible metonym, without the benefit of prior lexical restrictions towards one sense interpretation, it would commit to one of the senses immediately.

Furthermore, the results do not straightforwardly support an alternative model of polysemy that stores related senses on separate lexical nodes, which are accessed according to sense frequency (Foraker & Murphy, 2012; Klein & Murphy, 2002), as is standard in models of homophony (e.g., Simpson, 1981). Given that potentially metonymic names showed a slight bias towards the metonymic sense in offline norming when presented in isolation, the processor could not have relied on the dominant sense of a producer-for-product metonym for sense selection in subject position, as there was a penalty for such senses, nor in object position, as there was no cost for either sense (confirming Frisson & Pickering, 1999, 2007, and McElree et al., 2006).

A remaining point of interest regards the *timing* of the processing cost. In all three self-paced experiments above, a processing penalty was observed for passives on the region immediately following the verb, rather than on the verb itself. In the eye tracking study, there was an immediate penalty on the verb region itself, which could, at least in principle, be attributed to accessing text in the post verbal region while on the verb through parafoveal preview. Thus, the penalty for combining the literal sense of a metonymic name with an incompatible verb might seem delayed, at least in comparison to lexical ambiguity, in which senses are thought to be listed separately so that the processor must select one immediately (e.g., Frazier & Rayner, 1990), as suggested by an anonymous reviewer. We consider three possible explanations for the timing of the penalty, though others may exist. The first is that the timing reflects the use of *regular* metonymy in our manipulation: the clash between senses could be resolved by accessing a regular semantic relation between authors and their works (as in Pustejovsky, 1995). The additional processing resources required might simply not become available until after the source of the semantic clash has been identified and incorporated into the linguistic structure. Further, the penalty might be characterized in terms of a penalty for recruiting additional information to find the appropriate metonymic relation. Such an explanation would predict that less accessible or less regular relations would require more processing resources, thus eliciting larger penalties.

A second type of explanation would attribute the timing of the delay to thematic role processing: if the thematic processor values assigning the agent of the sentence above all else, it might well wait until the post verbal region, which contains a *by*-phrase associated with Agent theta role assignment, before revisiting theta role assignment of the sentential subject. As such, the processor might immediately identify that the initial noun was assigned an incorrect thematic role, but delay in its attempt to rectify it in order to determine which thematic role is appropriate. Accordingly, this approach would not predict any processing differences between *types* of thematic roles utilized in the revision process; for instance, an Agent following the verb (*Kafka was printed by the publisher*) would be no more favored than another type of thematic role, e.g., a Beneficiary (*Kafka was printed for the publisher*).

In a third and related proposal, the processor would assign an Agent role to both the sentential subject, by virtue of its structural position, and the noun in the *by*-phrase, by virtue of being a sister to the agent theta assigner *by* (see Grodinsky, 1986, for such a proposal). The clash of assigning the same thematic role to multiple arguments in distinct positions would itself generate the reanalysis process. A strong version of this account would predict that *only* multiply assigned Agent theta roles would elicit processing penalties for the cases studied above, at least on the post verbal region (other processes may enact a reinterpretation of the sentence). At present, these comments must remain speculative.

Many facets of SAP must be explored in greater depth to understand how the processor treats semantic knowledge as compared to structural information during interpretation. For example, it is not at present clear whether the processor applies SAP only to animate nouns, which are typically eligible for Agent assignment, or whether it also applies SAP to inanimate nouns, which typically are not eligible for Agent assignment, hence "animate" is placed in parentheses in (3) above. At this point, it would seem plausible that SAP applies only to potential or perhaps prototypical agents – i.e., animate nouns. Thus, open questions remain regarding under what conditions the processor follows the dictates of SAP and to what extent.

To conclude, we presented novel evidence in favor of the Underspecification Model, and, more specifically, for a general structure to the effects that different types of constraining information have on the homing in stage of metonym interpretation. More strongly constraining contexts, such as grammatical contexts, create mutually incompatible interpretations that force the processor to immediately commit to one sense over others. The processor is not necessarily forced to make immediate decisions between incompatible representations by weakly constraining information, though it may look to such information when attempting to recover from a mistaken commitment. In general, we hope that these findings enhance our understanding of how the language processor comes to adopt one specific interpretation of a metonym, and moreover, provide clues to how the processor manages to rapidly comprehend sentences despite continuously navigating sense ambiguities.

Acknowledgments

Financial support for this research was provided by Pomona College. The contents of this paper are the sole responsibility of the authors and do not necessarily represent the views of Pomona College. We thank Chuck Clifton, Lyn Frazier, and Adrian Staub, as well as the editor and two anonymous reviewers, for generous feedback on a previous draft. We benefited greatly from discussion with Jay Atlas, Lewis Bott, Andreas Brocher, Deborah Burke, Michael Diercks, Steven Frisson, Heeju Hwang, Elsi Kaiser, Meredith Landman, Mary Paster, and Sarah VanWagenen. Alex Drummond generously provided technical assistance with Ibex Farm. We also thank audiences at the second Experimental Psycholinguistics Conference in 2012, the 87th Annual Linguistics Society of America in 2013, and the 26th Annual CUNY Conference on Human Sentence Processing, where versions of this work have been presented, for their comments, criticisms, and suggestions.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2014.06.005.

References

Bader, M., & Häussler, J. (2009). Resolving number ambiguities during language comprehension. *Journal of Memory and Language, 61*, 352–373.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., & Maechler, M. (2009). *lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31*.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24*, 57–65.
Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 1–61). New York: Wiley.
Clifton, C. Jr., (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology, 47*, 222–246.
Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. Van Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: A window on mind and brain* (pp. 341–372). Oxford: Elsevier Ltd..
Clifton, C. C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317–334.
Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*, 547–619.
Drummond, A. (2012). *Ibex Farm*. <http://www.spellout.net/latest_ibex_manual.pdf> (computer program).
Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446.
Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.
Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164–203.
Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.
Ferreira, F., & Clifton, C. Jr., (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.
Ferreira, F., & Henderson, J. M. (1991b). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725–745.
Ferreira, F., & Henderson, J. M. (1991a). How is verb information used during syntactic parsing. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Amsterdam: North-Holland.
Fillmore, Charles J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart and Winston.
Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language, 67*, 407–425.
Frazier, L. (1999). *On sentence interpretation*. Dordrecht, The Netherlands: Kluwer.
Frazier, L., Pacht, J., & Rayner, K. (1999). Taking on semantic commitments, II: Collective versus distributive readings. *Cognition, 70*, 87–104.
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.
Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29*, 181–200.
Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3*, 111–127.
Frisson, S., & Frazier, L. (2004). Processing polysemy: Making sense of sense. *Poster presented at the annual meeting of the CUNY on human sentence processing*, College Park, MD.
Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology, 25*, 1366–1383.
Frisson, S., & Pickering, M. J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol, 16*, 149–171.
Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word; why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes, 22*, 595–613.
Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *The Quarterly Journal of Experimental Psychology, 39*, 611–627.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics. Speech Acts* (Vol. 3, pp. 41–58). New York: Academic Press.
Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
Grodinsky, Y. (1986). Language and the theory of syntax. *Brain and Language, 27*, 135–159.
Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 447–456.
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations between polysemous senses. *Journal of Memory and Language, 47*, 548–570.
Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language, 81*, 205–223.
Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*, 1–24.
Lowder, M. W., & Gordon, P. C. (2013). It's hard to offend the college: Effects of sentence structure on figurative-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 993–1011.
MacDonald, M. C., Perlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*, 127–150.
McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science, 30*, 115–124.
McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283–312.
Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond* (pp. 15–32). Hove, UK: Psychology Press.
Morris, R. K. (2006). Lexical processing and sentence context effects. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 377–402). London: Academic Press.
Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*, 143–184.
Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics, 12*, 109–132.
Nunberg, G. (2004). The pragmatics of deferred interpretation. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 344–364). Malden, MA: Blackwell Publishing Ltd..
Oakhill, J., & Garnham, A. (1996). The mental models theory of language comprehension. In B. Britton & A. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Hillsdale, NJ: Erlbaum.
Oakhill, J., Garnham, A., & Vonk, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes, 4*, 263–286.
Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond* (pp. 33–50). Hove, UK: Psychology Press.
Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems? *Behavioral Research Methods, 41*, 598–614.
Pustejovsky, J. (1995). *The generative lexicon*. Cambridge MA: MIT Press.
Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience, 18*, 97–109.
R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1290–1301.

Rohde, D. (2003). *Linger: A flexible program for language processing experiments*. <http://www.tedlab.mit.edu/~dr/Linger/> (computer software).

Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind and Language, 17*, 188–206.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. (1980). The word order problem in agrammatism: I. Comprehension. *Brain and Language, 10*, 249–262.

Searle, J. (1979). *Expression and meaning*. Cambridge, UK: Cambridge University Press.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Learning and Verbal Behavior, 20*, 120–136.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford, UK: Oxford University Press.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 431–449.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Wiliams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research, 21*, 193–218.

Zipf, G. K. (1945). The meaning-frequency relationship of words. *Psychology, 33*, 251–256.