Evidence for Serial Coercion: A Time Course Analysis Using the Visual-World Paradigm

Christoph Scheepers
Department of Psychology, University of Glasgow
58 Hillhead Street, Glasgow G12 8QB, UK
phone: +44-141-330-3606, fax: +44-141-330-4606
email: c.scheepers@psy.gla.ac.uk

Frank Keller and Mirella Lapata
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK
phone: +44-131-650-4407, fax: +44-131-650-6626
email: keller@inf.ed.ac.uk

Abstract

Metonymic verbs like *start* or *enjoy* often occur with artifact-denoting complements (e.g., *The artist started the picture*) although semantically they require event-denoting complements (e.g., *The artist started painting the picture*). In case of artifact-denoting objects, the complement is assumed to be type shifted (or *coerced*) into an event representation to conform to the verb's semantic restrictions. Psycholinguistic research has provided evidence for this kind of *enriched composition*: readers experience processing difficulty when faced with metonymic constructions compared to non-metonymic controls. However, slower reading times for metonymic constructions could also be due to *competition* between multiple interpretations that are being entertained in parallel whenever a metonymic verb is encountered. Using the *visual-world paradigm*, we devised an experiment which enabled us to determine the time course of metonymic interpretation in relation to non-metonymic controls. The experiment provided evidence in favor of a serial coercion process.

Keywords: competition, enriched composition, metonomy, coercion, semantic processing, visual-world paradigm, time course analysis.

1. Introduction

The interpretation of sentences such as *The artist started the picture* has attracted much attention in lexical semantics (Bach, 1986; Briscoe, Copestake, & Boguraev, 1990; Copestake, 1995; Pustejovsky, 1995; Vendler, 1968; Jackendoff, 1997) and recently also in psycholinguistics (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002; Lapata,

Keller, & Scheepers, 2003). The primary point of interest here is the verb *start*. Its complement (i.e., *picture*) denotes an entity, but in order to interpret the sentence correctly, the reader has to relate *start* to an event, and assign it an interpretation in which the artist started doing something to the picture, e.g., painting it, drawing it, or framing it.

In general, verbs like *start* (other examples include *finish* or *enjoy*) can select for verbal complements (as in *The artist started painting/to paint the picture*), event-denoting nouns (as in *The artist started the fight*), or entity-denoting nouns (as in *The artist started the picture*). In the latter case the object NP appears to be incongruent with the fact that the verb requires an event-denoting object. Therefore, in order to conform to the semantic restrictions of *start*, the complement must be *type shifted* or *coerced* from an entity to an event (Jackendoff, 1997; Partee, 1992; Pustejovsky, 1995). Pustejovsky (1991) dubs this phenomenon *logical metonymy*. As in the case of conventional metonymy (Nunberg, 1995; Lakoff & Johnson, 1980), one expression (here a noun phrase) is used in place of a related one (here an event associated with the NP). The metonymy is *logical* since it is triggered by type requirements which a verb places onto its arguments. The phenomenon involves interpolating additional meaning that is not present in the sentence containing *start* and its complement. The additional meaning is often an event related to the artifact denoted by the complement (e.g., *painting* or *drawing* for *picture*), but can also be provided by intra-sentential (Lapata et al., 2003) or extra-sentential context (Lascarides & Copestake, 1998). The process of constructing the missing information is sometimes called *enriched composition* (Jackendoff, 1997) since, unlike standard composition, it involves the computation of extra linguistic material. Throughout this article, uses of the term metonymy refer exclusively to logical metonymy.[1] Furthermore, we will refer to *metonymic verbs* as a shorthand for "verbs inducing logical metonymy".

An enriched composition account of metonymy predicts that type shifting incurs a processing cost, as additional structure needs to be constructed. McElree et al. (2001) tested this prediction in a self-paced reading experiment. They contrasted constructions requiring enriched composition (see sentence (1-a)) with constructions involving standard composition (see (1-b) and (1-c)) and found that readers experienced more difficulty with sentences requiring enriched composition. Upon encountering the complement noun, reading times for (1-a) and (1-c) were significantly longer than for (1-b); one word later, reading times for (1-a) were longer than reading times for (1-b) and (1-c) – the latter two conditions were indistinguishable at this point. McElree et al. (2001) interpret these results as evidence for enriched composition: constructions like (1-a) engender longer reading times because the complement noun is coerced into an appropriate event, which requires the costly construction of additional structure.

(1) a. The artist started the picture in his studio in the city.

[1] Although closely related, conventional metonymy (e.g., *Peter read Shakespeare* where *Shakespeare* stands for Shakespeare's works) is typically *not* analyzed in terms of semantic type coercion.

We would like to thank Malte Viebahn for his tireless efforts in designing the materials for Experiment 1 and for conducting this experiment as part of his Erasmus research practical in Dundee. Also, we are grateful to Yuki Kamide for testing some of the participants in Experiment 2 in combination with one of her own experiments. We further acknowledge Martin Pickering, Roger Levy, and the Edinburgh Computational Psycholinguistics Group for valuable comments on this and related topics. A preliminary version of this work was presented as a talk at AMLaP-2005 in Ghent (Belgium), September 5–7, 2005.

b. The artist painted the picture in his studio in the city.
c. The artist analyzed the picture in his studio in the city.

Follow-up experiments by Traxler et al. (2002) and Pickering, McElree, and Traxler (2005) confirmed that sentences requiring enriched composition incur reading difficulties. In eye-tracking, reliable differences emerged at the complement noun (Pickering et al., 2005) or on the two words succeeding it (Traxler et al., 2002). A similar effect was found when type-shifted sentences like (1-a) were matched with control sentences that explicitly verbalized the missing meaning (2).

(2) The artist started painting the picture in his studio in the city.

Further evidence for enriched composition comes from contrasting sentences like (1-a) with controls whose verbal objects are eventive noun phrases (see sentence (3)). Traxler et al. (2002) showed that sentences containing non-eventive complements (e.g., *started the picture*) incurred more processing difficulty than their eventive controls (e.g., *started the fight* in (3)). This result indicates that processing difficulty stems from the combination of metonymic verbs with their complements and it is not solely linked to the complement noun phrase.

(3) The artist started the fight in his studio in the city.

Using the multi-response speed-accuracy tradeoff (SAT) paradigm, McElree, Pylkkänen, Pickering, and Traxler (2005) showed that type-shifted constructions are being processed less accurately (in terms of whether the sentences *made sense* to participants or not) and more slowly than minimally contrasting controls. The fact that type-shifting had a measurable effect on the speed of processing suggests that readers engage additional resources in computing the missing meaning. In this paper, we will pursue a similar approach, but focus more on the number of different interpretations that are computed on-line rather than processing accuracy (we will return to this point in Section 5).

2. Competition vs. Enriched Composition

As we saw in the previous section, existing experimental work on the processing of logical metonymy has almost unequivocally demonstrated that metonymic verbs cause processing difficulty relative to non-metonymic controls (but see de Almeida, 2004). This raises the question of precisely what kinds of cognitive processes are involved in interpreting metonymic constructions, and why these engender additional processing cost. The enriched composition hypothesis offers the following explanation: when speakers interpret a metonymic construction, they have to construct additional structure, over and above the structure that has to be constructed for a non-metonymic construction and this slows down the comprehension process.

An alternative explanation, however, is that comprehending metonymic constructions involves pursuing multiple interpretations at the same time. Sentences like (1-a) allow for several different interpretations, although some may be more dominant than others. For example, *started the picture* could trigger a *painting* interpretation, an *analyzing* interpretation, a *framing* interpretation, and so on. It is possible that these interpretations compete with one another and therefore decelerate the process of establishing a final interpretation. This effect is familiar from the lexical access literature, where it was found that homonymous lexical items (words with multiple, unrelated meanings such as *bark*) are accessed more slowly than unambiguous lexical items (e.g., Rayner & Duffy, 1986; Rodd, Gaskell, & Marslen-Wilson, 2002). Indeed, there are a number of sentence

comprehension models that assume similar competition processes to take place when ambiguity is encountered at the sentence level (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Seidenberg & MacDonald, 1999; Trueswell & Tanenhaus, 1994).

The reading paradigms used in previous studies of metonymic verbs do not rule out such a competition-based explanation. As we will discuss in Section 4, the main reason for this is that in a reading task, it is impossible to establish how strongly speakers commit to a single interpretation, or whether they pursue multiple interpretations while a sentence unfolds over time.

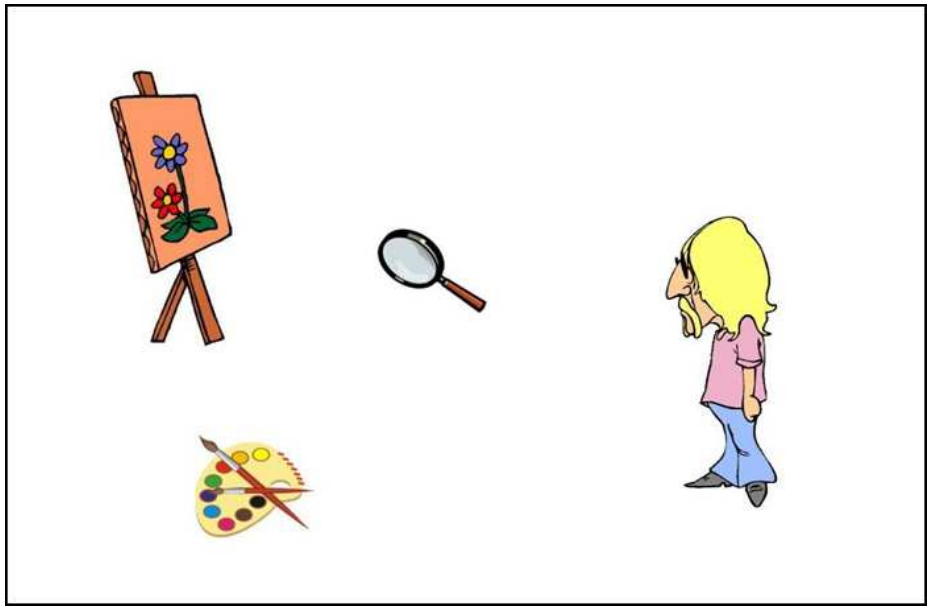

Figure 1. Sample item: picture for the sentence *The artist started/painted/analyzed the flowery picture using the depicted* . . .

In this paper, we will use the *visual-world paradigm* to investigate the processing of metonymic verb constructions. In this paradigm, a visual scene is presented concurrently with a spoken sentence in order to establish how eye-movement patterns on the scene are affected by linguistic variation. Specifically, our experiments will combine pictures such as the one in Figure 1 with sentences of the form *The artist started/painted/analysed the flowery picture using the depicted* . . . . Notice that our pictures will always contain two critical instrument entities; one will be compatible with the dominant interpretation of the metonymic verb (e.g., the paint brushes for *painting*) while the other one will support a subordinate, yet plausible, alternative interpretation (e.g., the magnifying glass for *analyzing*).

Previous visual-world research has shown that participants' eye-movements are closely time-locked with the auditory linguistic input and, more importantly, that *anticipatory* eye-movements occur which indicate the kinds of interpretations participants entertain for ambiguous input (e.g., Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Allopenna, Magnuson, & Tanenhaus, 1998; Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005).

In contrast to standard reading tasks, the visual-world paradigm therefore allows us to es-

tablish the time course of metonymic interpretation in more detail. Anticipatory looks to instrument entities associated with different interpretations of the same metonymic verb provide an index of the different kinds of semantic commitments that are being made on-line in relation to the linguistic input. Moreover, relative proportions of looks to these instrument entities can be taken as an estimate of the relative "strength" of a given interpretation: stronger commitment to a certain interpretation is likely to elicit a stronger visual bias towards one of the critical instrument entities.

This, in turn, will enable us to determine whether enriched composition (i.e., the assumed *coercion* processes) can fully explain the processing difficulty associated with metonymic verbs, or whether at least part of this difficulty stems from competition between alternative interpretations. According to a serial coercion account, anticipatory eye-movements should favor only one of the instrument entities in the display (the entity compatible with the dominant interpretation), since only one interpretation is pursued by speakers at any given time; however, given that metonymic verbs require the computation of additional structure, it should take longer to establish this preferred metonymic interpretation in comparison to a semantically matched non-metonymic control condition. By contrast, a parallel account of coercion, or indeed a competitive account that does not rely upon the notion of coercion, predicts that over a number of trials, anticipatory eye-movements should be more evenly spread across the two possible instrument entities (meaning that competition should manifest itself in a weaker interpretational bias).

Before discussing these hypotheses in more detail (Section 4), we will explain how the experimental stimuli were constructed.

3. Experiment 1: Norming Study

This experiment served as a norming study for the materials used in our subsequent visual world study (Experiment 2). The aim was to establish interpretation preferences for a set of metonymic verbs, and to make sure that the critical instrument entities in the visual stimuli were indeed associated with the events denoted in the non-metonymic control conditions. The experiment was conducted as a spoken sentence completion experiment, in which participants saw pictures combined with one of three types of written sentence fragments: a *metonymic verb* fragment as in (4-a), a *preferred verb* fragment as in (4-b), or a *non-preferred verb* fragment as in (4-c).

(4) a. The artist started the flowery picture using the depicted . . .
b. The artist painted the flowery picture using the depicted . . .
c. The artist analyzed the flowery picture using the depicted . . .

Participants were asked to complete each fragment on the basis of what they saw in the picture and what was given in the fragment. The relevant example picture is shown in Figure 1. Crucially, the pictures always contained two entities that could function as instruments for alternative interpretations of the metonymic verb: the paint brushes, for example are compatible with the *painting* interpretation of the metonymic verb *start*, while the magnifying glass is compatible with the *analyzing* interpretation.

We predicted that the most frequent completion of (4-b) should refer to the paint brushes, while the most frequent completion of (4-c) should refer to the magnifying glass. In the metonymic verb case (4-a), both completions are plausible, but we expected the *painting* interpretation to be preferred and the *analyzing* interpretation to be dispreferred (more references to the paint brushes rather than the magnifying glass).

Such a pattern of results would confirm that the depicted instruments are indeed associated with the events denoted in the non-metonymic control conditions. It would also establish off-line interpretation preferences for the metonymic verbs, in line with the sentence completion preferences reported in McElree et al. (2001) and Traxler et al. (2002).

3.1. Method

3.1.1. Participants

Sixty native English speakers (undergraduates from the University of Dundee) took part in this study, receiving either course credit or £2 subject payment. Participants were tested in individual sessions, each of which took about 15 minutes to complete.

3.1.2. Materials

A set of 34 easily depictable candidate items was selected from McElree et al. (2001) and Traxler et al. (2002). From these, we generated 34 stimulus sets, each of which consisted of a picture and three matching sentence fragments. The sentence fragments were constructed in the same way as example (4), i.e., they only differed in the verb, which was either a metonymic verb (e.g., *start*), or a non-metonymic verb corresponding to the preferred interpretation (e.g, *paint*) or a non-preferred interpretation (e.g., *analyze*) of the metonymic verb. Note that neutral adjectives such as *flowery* were inserted before the object nouns; this was to create longer NPs for the analysis of the visual world data in Experiment 2.

The pictures were generated from clipart libraries such that each of them contained four entities, as illustrated in Figure 1. One entity corresponded to the subject of the target sentence (e.g., *artist*), and one to the object of the target sentence (e.g., *flowery picture*). The other two entities depicted instruments congruent with two alternative interpretations of the metonymic verb (e.g., *paint brushes* and *magnifying glass*). Visual arrangements of the four picture entities were more or less arbitrary and differed across items so as to avoid any systematic viewing patterns. The full set of experimental pictures can be obtained from the first author.

The items were allocated to three stimulus files, each of which contained all of the 34 pictures, but combined with a different sentence fragment across files. Each file contained the same number of items per condition (according to a Latin square) and was presented to 20 participants.

3.1.3. Procedure

The experiment was conducted in a quiet experimental room. Participants were seated approximately 55 cm from a 17” color monitor with 1024×768 pixel resolution. Stimulus presentation and data recording were controlled by an Intel Pentium PC running DMDX (Forster & Forster, 2003).

Each participant was randomly assigned one of the three stimulus files. The order of items per file was determined at random for each participant. In order to initiate a trial, the experimenter pressed a button, triggering the presentation of the picture; the corresponding sentence fragment was displayed in a 24 point font at the bottom of the picture. At the onset of the picture presentation, a 100 ms alert sound was played over speakers (this helped identify trial onsets in the concurrent audio recordings). Participants were asked to use the information both in the picture and in the written sentence fragment so as to generate a complete spoken sentence. They were instructed to produce whole sentences rather than just name the missing nouns. After the participant had finished

	PV	NPV	Other
Metonymic verb	$.69 \pm .05$ $(.07)$	$.19 \pm .03$ $(.06)$	$.13 \pm .03$ $(.05)$
Preferred verb	$.82 \pm .04$ $(.07)$	$.09 \pm .02$ $(.04)$	$.09 \pm .03$ $(.04)$
Non-preferred verb	$.13 \pm .04$ $(.04)$	$.79 \pm .04$ $(.04)$	$.08 \pm .02$ $(.04)$

Table 1: Average completion probabilities per condition for the final set of materials, with 95% confidence limits by participants (items).

speaking, the experimenter pressed a button to proceed to the next trial. The sessions were audio-recorded on minidisk.

3.1.4. Response Annotation

Spoken responses were transcribed and annotated as one of PV, NPV, or Other. A response was scored as PV if the final noun of the completed sentence unambiguously referred to the instrument associated with the preferred verb (*paint brushes* or *palette* in our example). A response was coded as NPV if the final noun unambiguously referred to the instrument associated with the non-preferred verb (e.g., *magnifying glass*). All remaining responses (ambiguous references, references to other entities in the picture, or references to entities that were not displayed in the picture) were coded as Other. Probabilities of responses were taken as the dependent variable for analysis.

3.2. Results and Discussion

The data were analyzed using *k*-means cluster analysis, a procedure that helps identifying homogeneous subsets of items in terms of the degree of similarity between item-specific response patterns. The analysis revealed that ten of the 34 candidate items produced rather idiosyncratic results, which partly disagreed with the desired distribution of completions. Results for the remaining 24 items were as expected (see Table 1): in the metonymic verb condition, about two thirds of the completions referred to the instrument associated with the preferred verb (PV completions, e.g., *The artist started the flowery picture using the depicted paint brushes*), in line with the completion data in McElree et al. (2001) and Traxler et al. (2002).[2] A substantially stronger bias towards PV completions was observed in the preferred verb condition (e.g., *The artist painted the flowery picture using the depicted paint brushes*), as can be seen from the confidence intervals in Table 1. In the non-preferred verb condition, there was a strong preference in favor of the instrument associated with the non-preferred verb (NPV completions, e.g., *The artist analyzed the flowery picture using the depicted magnifying glass*). Importantly, the bias towards *appropriate* instruments (PV in the preferred verb condition, NPV in the non-preferred verb condition) was roughly the same for the two control verb conditions, accounting for about four out of five responses in each case.

4. Experiment 2: Visual-World Study

Experiment 1 established the interpretation preferences for a set of 24 picture-sentence combinations involving metonymic and non-metonymic verbs. The present experiment used these ma-

[2]For 20 of the selected 24 items, there was at least one reference to the non-preferred instrument in the metonymic verb condition. The remaining four items showed a high proportion of Other responses in this condition (19% on average). The off-line data therefore confirm the existence of potentially competing alternative interpretations for metonymic verb constructions.

terials to investigate the time course of the interpretation of metonymic constructions. Participants saw pictures such as the ones in Figure 1 and at the same time listened to spoken sentences such as the ones in (5), while their eye-movements were recorded. Note that half of the time metonymic sentences (5-a) ended in an instrument noun compatible with the preferred interpretation, while in the other half of trials, they ended in the non-preferred instrument noun.

(5) a. The artist started the flowery picture using the depicted paint brushes/magnifying glass.
b. The artist painted the flowery picture using the depicted paint brushes.
c. The artist analyzed the flowery picture using the depicted magnifying glass.

This experiment was based on the assumption that participants' eye-movements should reflect their on-line interpretation preferences. We expected that for the sentences in (5), anticipatory eye-movements (i.e., eye-movements to scene entities in advance of their referring expressions in the auditory material) should be launched to the depicted instruments as soon as the verb and its object have been processed. Given that perceivers are likely to anticipate the forthcoming direct object when they encounter the verb (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003, see also Arai, Gompel, & Scheepers, 2006), a plausible triggering point for instrument anticipation would be the point in time where the object noun is available. This however does not exclude the possibility that, in some instances at least, instrument anticipation may already take place before the object noun, while in other instances it may occur well after the object noun. Our analyses will take the presumed probabilistic nature of instrument anticipation into account by focusing on *distributions* of looks to different instrument entities over time, measured from the verb (the earliest point in which conditions differ) until a point in time where the critical instrument noun has been processed. These temporal distributions will then allow us to determine the degree to which different interpretations compete with one another in metonymic verb constructions, and also, whether interpretation of logical metonymy is truly associated with a slowdown in processing.

In the following, we will distinguish between three accounts of metonymic verb processing: Serial Coercion, Immediate Competition, and Parallel Coercion. Each of these predict different outcomes for comparisons between condition (5-a) and condition (5-b).[3]

According to the Serial Coercion account, a single interpretation is pursued in metonymic verb constructions such as *The artist started the picture*. That is, the processor only considers the dominant interpretation (as established in the previous norming study) while other interpretations are ignored unless information supporting them is encountered. However, constructing the one (dominant) interpretation requires a time-consuming type shifting operation, meaning that the processor needs to build additional semantic structure to obtain an interpretation that complies with the verb's selectional restrictions (e.g., *The artist started painting the picture*). The Serial Coercion account therefore predicts a difference in *dynamics* between (5-a) and its non-metonymic counterpart in (5-b): anticipatory eye-movements should favor the preferred-verb instrument (*paint brushes*) about equally strongly in (5-a) and (5-b) (only the dominant *painting* interpretation is considered in each case), but processing should be decelerated in (5-a) relative to (5-b) because (5-a) requires a time-consuming type shifting operation to take place.

[3]For the non-preferred control condition in (5-c), we predict an anticipatory bias towards the non-preferred instrument (magnifying glass). Given previous reading data from McElree et al. (2001) and Traxler et al. (2002), we also expect evidence for a processing slowdown in (5-c) relative to (5-b). Note, however, that the non-preferred control condition (5-c) is not as vital for distinguishing between different accounts of metonymic verb processing as the other two conditions.

The basic assumption behind Immediate Competition is that several competing interpretations are generated as the sentence unfolds, based on a multi-level, probabilistic constraint satisfaction process (McRae et al., 1998; Tanenhaus, Spivey-Knowlton, & Hanna, 2000). Under this assumption, no additional semantic structure needs to be computed for metonymic verbs – all relevant interpretations (in our example, *painting the picture*, *analyzing the picture*, etc.) are immediately available in the competitor set.[4] In our materials, different degrees of competition might be observed as early as during the verb itself (e.g., *The artist started* . . . in (5-a) allows a wider range of plausible continuations than *The artist painted* . . . in (5-b)), or, in the context of instrument anticipation, upon processing the object noun. Since no additional structure is generated, Immediate Competition does not predict any differences in processing dynamics between (5-a) and (5-b), but instead a difference in *interpretation strength*, as measured in the proportions of looks to the instruments in the picture. If two competing interpretations are generated for the metonymic verb in (5-a), then anticipatory eye-movements should be more likely to alternate, both within and across trials, between the two possible instruments in the display shortly after the verb or the following object has been encountered. On average, this implies that the preferred instrument (paint brushes) should receive fewer anticipatory looks in (5-a) than in (5-b) where competition should be considerably weaker, or even absent.

The Parallel Coercion account combines features of Serial Coercion and Immediate Competition. Like Serial Coercion, it assumes that the interpretation of metonymic verb constructions requires the computation of additional semantic structure, which should decelerate the processing of (5-a) relative to (5-b). However, in contrast to Serial Coercion, not only the dominant interpretation, but also alternative (dispreferred) interpretations are being computed during this enriched composition process. In this respect, Parallel Coercion is similar to Immediate Competition. The interpretations are pursued in parallel and compete with one another to a degree that is proportional to the meaning dominance established off-line (see Experiment 1). Parallel Coercion therefore predicts a combined effect of reduced interpretation strength (fewer anticipatory eye-movements to the preferred instrument overall) and decelerated processing in (5-a) relative to (5-b). The former follows from the assumption that multiple competing interpretations are being entertained in parallel in (5-a), the latter from assuming that additional semantic structure needs to be computed in (5-a).

Figure 2 provides a schematic illustration of the predictions made by the three accounts. The black line in each subfigure represents the preferred non-metonymic condition (*The artist painted the picture*), the gray line the metonymic condition (*The artist started the picture*). The time from having encountered the verb until the onset of the critical instrument noun (or some earlier point in time) is plotted on the X-axis. The Y-axis represents the strength of commitment to the preferred-verb interpretation (*painting*), which, in the context of this visual world experiment, should be measurable in terms of numbers of looks to the critical instruments in the visual display.

We assume that commitment to the preferred *painting* interpretation gradually increases as a non-linear function of time until a maximum is reached. This maximum, in turn, can be taken as an indicator of the overall strength of commitment to the *painting* interpretation. The dashed lines in the plots mark points in time where a given percentage (here, 50%) of the relevant interpretational maximum is achieved and help to demonstrate cross-condition differences in dynamics vs. strength of interpretation.

[4] As we shall see below, increased reading difficulty for metonymic verb constructions (5-a) would follow directly from competition between alternative interpretations in such an account.

Figure 2a illustrates the predictions of the Serial Coercion account: the two conditions achieve the same maximum strength of commitment to the preferred *painting* interpretation (no difference in asymptote – recall that Serial Coercion predicts only the dominant interpretation to be pursued in each case, while possible competing interpretations are ignored); however, in case of a metonymic verb, accretion of this interpretational bias takes more time because additional semantic structure needs to be computed. As can be seen from the dashed lines in Figure 2a, the point in time where interpretation strength reaches 50% of the maximum differs considerably between the two conditions.

A different state of affairs is predicted by Immediate Competition, as illustrated in Figure 2b. Because of competing interpretations, the metonymic verb condition reaches a lower maximum than the non-metonymic verb condition, which corresponds to a difference in overall interpretation strength. In terms of processing dynamics, however, the two conditions are identical, as shown by the dashed lines in Figure 2b: both conditions accumulate a given percentage of the relevant interpretational maximum at exactly the same point in time (Immediate Competition does not assume computation of extra semantic structure in metonymic verb constructions).

The Parallel Coercion account (see Figure 2c) predicts a combined effect of decelerated processing and reduced interpretation strength in metonymic-verb constructions, where competition is assumed to be mediated via a costly coercion process. Accordingly, the two curves in Figure 2c achieve different maxima, and the 50% point is located at different points in time.

The plots also illustrate why it is difficult to distinguish between these accounts in a reading experiment. The open circles in each plot are taken to indicate hypothetical points in time where readers would decide to move their eyes to the next region in the sentence (e.g., after having read the object noun *picture*), or to press a button for the next presentation segment, respectively. Assuming that this decision often takes place before maximum commitment to a given interpretation is achieved, a difference in reading time between the metonymic verb condition and the non-metonymic control is compatible with a difference in dynamics (Figure 2a), a difference in interpretation strength (Figure 2b), or a combined effect (Figure 2c). Since we usually do not know how strongly (in relation to the potential maximum) readers commit themselves to a given interpretation, e.g., after having processed *The artist started/painted the picture . . .*, cross-condition differences in reading time cannot fully decide between the three hypothetical accounts of metonymic verb processing. (Corresponding off-line data are not necessarily a valid measure of the kinds of semantic commitments that are established on-line.) Our solution to this problem is to use the visual world paradigm to map out how interpretational preferences (as measured in proportions of looks to indicative instrument entities in the visual display) change over time. A precise functional description of these interpretational changes (analogous to the curve fitting approach in speed-accuracy tradeoff paradigms, see, e.g., McElree & Dosher, 1993; McElree & Griffith, 1995; McElree et al., 2005; Reed, 1976; Wickelgren, Corbett, & A.Dosher, 1980) will then allow us to determine cross-condition differences in processing dynamics and interpretation strength, respectively.

4.1. Method

4.1.1. Participants

Eighty-eight native speakers of English (undergraduates from the Universities of Dundee and Edinburgh) took part in this study, receiving either course credit or £5 subject payment. Participants were tested in individual sessions, each of which took about 30 minutes to complete. Thirty-two

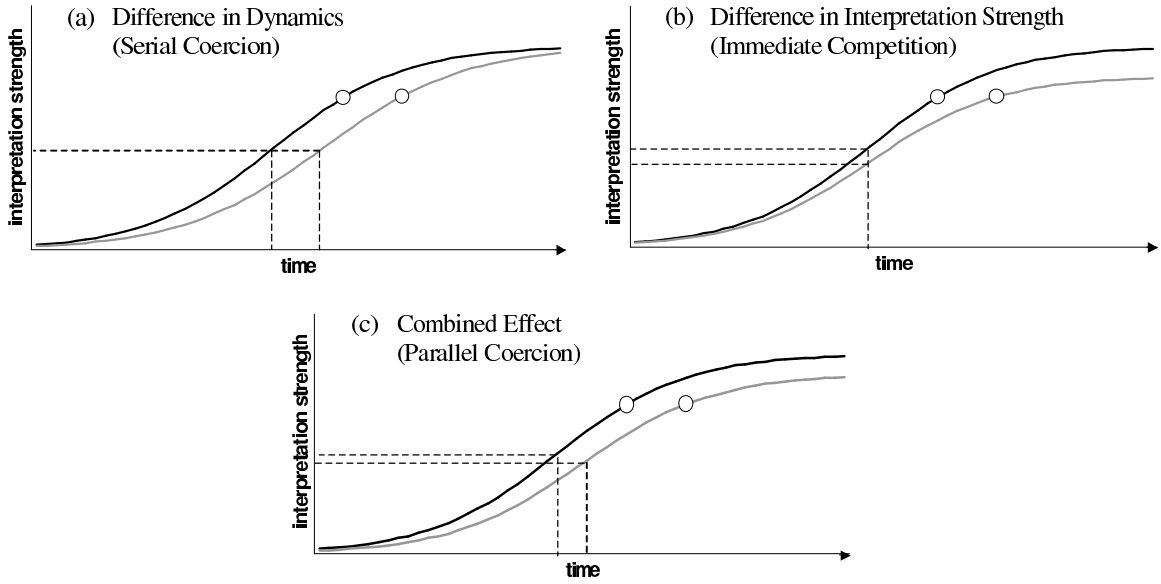

Figure 2. Hypothetical time course predictions derived from (a) the Serial Coercion account, (b) the Immediate Competition account, and (c) the Parallel Coercion account. Black line: preferred non-metonymic condition (*The artist painted . . .*), gray line: metonymic condition (*The artist started . . .*). Time is plotted on the X-axis and strength of commitment to the preferred interpretation (*painting*) on the Y-axis. The dashed lines indicate points in time where interpretation strength has reached 50% of the maximum, while the open circles represent hypothetical threshold points at which readers decide to move on to the next region in the sentence (see text).

participants were tested in Edinburgh, the remaining participants in Dundee. Both labs possessed the same experimental apparatus and software.

4.1.2. Materials

The materials used in this experiment were the 24 items that were identified as showing the appropriate biases in Experiment 1 (see Section 3.2). The pictures were identical to those in Experiment 1, but instead of written sentence fragments, we now used complete spoken sentences for our experimental manipulations. The sentences were constructed according to the completion preferences in Experiment 1. Each picture was combined with three different versions of spoken sentences, resulting in three experimental conditions: the metonymic verb condition, the preferred verb condition and the non-preferred verb condition. The metonymic verb condition ended either with the preferred instrument noun or with the non-preferred instrument noun (see (5-a)) in an equal number of trials. Hence, there was a 50% chance of metonymic verb sentences to end in either of the two instrument nouns. The preferred and the non-preferred verb conditions always ended in the corresponding preferred vs. non-preferred instrument nouns (see (5-b) and (5-c)). Appendix A lists the full set of sentences used.

The sentences were read by a male native speaker of Scottish English, recorded on minidisk in a sound-proof booth. The speaker was instructed to use a neutral intonation. To normalize the auditory stimuli, cross-splicing was used, ensuring that the recordings were identical across conditions between the offset of the verb and the onset of the instrument noun.

Furthermore, a set of 30 fillers was constructed. Each of the fillers consisted of a visual scene containing four entities, similar to the experimental pictures. Auditory filler sentences employed a variety of different structures unrelated to the critical target sentences. They were read and recorded in the same way as the experimental items.

The materials were allocated to four master files, each of which contained all of the 24 target pictures, but combined with different versions of spoken stimuli across files. Hence, only the linguistic input varied across conditions while the pictures stayed the same. Each file contained the same number of items per condition, according to a Latin square. The 30 fillers were added to the four files, and two fixed randomizations were generated for each file, making sure that the first three items per file were fillers. This yielded eight stimulus files, each of which was seen by a total of 11 participants.

4.1.3. Procedure

Participants were seated approximately 65 cm from a 21" color monitor with 1024×768 pixel resolution; twenty-four pixels equaled about one degree of visual angle. Participants wore an SR Research Eyelink II head-mounted eye-tracker running at 500 Hz sampling rate. Viewing was binocular, but only the participant's dominant eye was tracked (the right eye for about 68% of the participants, as determined by a simple parallax test prior to the experiment). Participants were instructed to avoid strong head movements throughout the experiment. The auditory stimuli were presented via a pair of speakers situated to the left and right of the screen. The recordings were played from the hard disk as 16 kHz mono sound clips. A USB gamepad was used to record button responses. Stimulus presentation and data recording were controlled by two PCs running experimental software developed by the Psycholinguistics Group at Saarland University on the basis of the Eyelink API.

Each participant was randomly assigned one of the eight master files. At the start of the experiment, the experimenter performed the standard Eyelink calibration routine, which involves participants looking at a grid of nine fixation targets in random succession. Then a validation phase followed to test the accuracy of the calibration against the same targets. Calibration and validation was repeated at least twice throughout the experiment, or if the experimenter noticed that measurement accuracy was poor (e.g., after strong head movements or a change in the participant's posture).

Each trial was structured as follows: first a fixation point was displayed in the middle of the screen, accompanied by a brief alert sound. Once the participants had fixated this point, the experimenter performed drift correction and started the trial. The picture was displayed, and after a fixed 1000 ms preview period, the spoken sentence was played over speakers. Each picture remained on the screen for 7000 ms before the next trial was initiated. The auditory sentence typically ended 1000–2000 ms before the end of the picture presentation. Participants were instructed to view the pictures and listen to the sentence attentively, so that they were able to answer subsequent questions. In 25% of the cases (determined at random), the trial was followed by a written question on the screen, replacing the picture. The question could refer either to the picture (e.g., *Did the artist have a beard?*) or to the sentence (e.g., *Did the artist sell the picture?*) of the immediately preceding trial. Whenever such a question appeared, subjects had to answer it by pressing either the "yes" button or the "no" button on the gamepad.

4.1.4. Primary Data Processing

For each picture, a template was generated consisting of a 1024×768 pixel bitmap in which the entities in the visual scene and the background were color-coded. For example in Figure 1, the painting, the artist, the paint brushes, and the magnifying glass all formed separate regions with distinct colors. Each region was defined in terms of a 12-pixel halo around the relevant entity's contour. The output of the eye-tracker included the X- and Y-coordinates of participants' fixations, which were converted into region codes using the templates. The region codes were then mapped onto three scoring regions: preferred instrument (the paint brushes in our example), non-preferred instrument (magnifying glass), or other (artist, painting or background). Fixations shorter than 80 ms (approximately 2–3% of all fixations) were pooled with preceding or following fixations if these fixations were within 0.5 degrees of visual angle, otherwise they were deleted (short fixations often result from false saccade planning rather than meaningful information processing, e.g., Rayner & Pollatsek, 1989). Times for blinks were added to the immediately preceding fixations (assuming that processing does not pause during a blink) and fixations outside the screen area (less than 1% of all fixations) were deleted. Finally, all consecutive fixations within one region (i.e., before a saccade to another region occurred) were added together and counted as one *gaze*.

The eye-movement data per trial were then analyzed as follows. The time period between 1000 ms from picture onset (start of sentence) and 7000 ms from picture onset (end of picture presentation) was divided into 50 ms timeslots, accounting for the fact that saccades may require up to 50 ms execution time to cover the relatively large angular distances between different scoring regions in the display (see, e.g., Abrams, Meyer, & Kornblum, 1989). For each time slot, we counted the number of gazes that were observed for each of the three scoring regions. For instance, if a gaze on a region started at 1000 ms and lasted until 1130 ms, then one gaze on the region would be scored for the timeslots 1000–1050 ms, 1050–1100 ms, and 1100–1150 ms. Preceding and subsequent timeslots would score zero gazes for that region, unless the region was inspected several times within the same trial. The resulting data were then used to compute gaze probabilities (across trials) per region, defined as number of gazes on a given region in a given timeslot divided by the total number of gazes in the relevant time slot.

4.2. Results and Discussion

4.2.1. Anticipation

Figure 3 plots gaze probability distributions over time (50 ms resolution) for the two regions of interest, namely, the preferred instrument (e.g., paint brushes) and the non-preferred instrument (e.g., magnifying glass), separately for the metonymic verb condition (Figure 3a), the preferred verb condition (Figure 3b), and non-preferred verb condition (Figure 3c). The data in Figure 3a are collapsed across the two versions of the metonymic verb condition (ending either in the preferred or non-preferred instrument noun, see (5-a)).

Each plot in the figure spans a time period of 1000–7000 ms from picture onset, i.e., the preview phase is not included. Also shown are the average onsets of the verb and the critical instrument noun in each condition (solid vertical lines); the dotted vertical lines indicate 99% confidence limits (across items) for these onsets. For each time slot, we performed a binomial test on raw gaze counts to determine whether there is a significant difference in numbers of gazes between the two regions of interest. The results of these tests are highlighted by the gray boxes in the plots, indicating

time periods with a significant difference at $p < .01$.[5] Complementary chi-square tests confirmed that these differences did not reliably interact with participants or items ($ps > .05$), i.e., that they can be generalized across individuals and materials. For time-slots outside the boxes, either no reliable differences were established, or there were significant interactions with participants and items, respectively.

As can be seen, participants launched anticipatory eye-movements to the depicted instruments well in advance of the onset of the instrument noun, but considerably later than the onset of the verb.[6] In the preferred verb condition (Figure 3b), there were significantly more gazes on the preferred instrument, while in the non-preferred verb condition (Figure 3c), there were more gazes on the non-preferred instrument. For the metonymic verb condition, Figure 3a indicates significantly more looks to the preferred rather than non-preferred instrument region, suggesting that participants interpret metonymic verbs in a way that is comparable to the preferred verb interpretation (i.e., they take *started the flowery picture* to mean *started painting the flowery picture* in our example). Importantly, in each condition, the relevant gaze probability differences between the two instrument regions reached significance even before the lower 99% confidence limit of the temporal starting point of the instrument noun (i.e., the onset of *paint brushes* or *magnifying glass*, respectively). This can be taken as evidence for anticipation. Another observation concerns the durations of the visual preferences across conditions: while the non-metonymic conditions elicited rather long-lasting preferences for the appropriate instrument regions (see Figures 3b and 3c), the bias towards the preferred instrument region in the metonymic verb condition was comparatively short-lived (Figure 3a). This is most likely due to the fact that 50% of the metonymic verb sentences ended in the non-preferred instrument noun, and that participants would shift their attention accordingly in those trials after recognizing the noun.

A comparison between Figures 3a and 3b also appears to indicate that gaze probabilities for the two instruments diverge later in the metonymic verb condition than in preferred verb condition. This may reflect a dynamics difference due to decelerated processing in the metonymic verb condition, as predicted by both Serial and Parallel Coercion. To find out whether this is truly the case, we conducted a more rigorous time course analysis comparable to the curve-fitting approach in SAT and related paradigms.

4.2.2. Time Course Analysis

The purpose of this time course analysis was to provide a precise functional description of how strength of interpretation develops over time in each condition, and to determine whether cross-condition contrasts are better characterized in terms of differences in processing dynamics, differences in interpretation strength, or a combination of both.

The data in Figure 3 were first converted into *probability differences* (ΔP) to quantify the strength of the bias towards the preferred instrument region in each condition (see Figure 4): for both the metonymic verb and the preferred verb condition, gaze probabilities on the non-preferred instrument region (magnifying glass) were subtracted from the corresponding gaze probabilities on the

[5] Given the large number of tests, we employed a stricter significance criterion than the commonly assumed 5% rule. Post-hoc adjustments like the Bonferroni correction, on the other hand, would result in unacceptably high type II error probabilities.

[6] In fact, it appears that the curves in Figure 3 start to diverge just around the onset of the auxiliary verb *using*, whose average onset occurred 1068 ms before the onset of the critical instrument noun. This could mean that anticipation of the instrument noun was triggered by *using*, or alternatively, that the processes enabling anticipation were completed after the object noun *picture* had been integrated.

preferred instrument region (paint brushes), such that more positive values indicate a stronger visual bias towards the preferred instrument region; for the non-preferred verb condition, gaze probabilities on the preferred instrument region (paint brushes) were subtracted from the gaze probabilities on the non-preferred instrument region (magnifying glass), such that more positive values indicate a stronger visual bias towards the non-preferred instrument region (note that in the present analyses, we were more interested in the strength of the bias rather than its direction). The difference curves in Figure 4 span a 3500 ms time period from the onset of the verb (determined individually for each item) until about 1000 ms after the onset of the critical instrument noun, whose cross-condition average[7] is indicated by the solid vertical line in the plot, together with 99% confidence limits (dashed vertical lines). Probability differences in the given time period formed the basis for the present time course analysis.

Figure 4 indicates that there were time periods where ΔP (our measure of interpretation strength) was roughly at zero (no visual preference for either instrument region), followed by time periods where ΔP was rising, reaching a peak (maximum visual preference for the preferred or non-preferred instrument region), and then declining again. We fitted a range of differently shaped peak distribution functions to our data, and identified the *Logistic Power Peak* (*LPP*) function as the best description of the variance both within and between conditions. A mathematical definition of this function is given in equation (1) below.

$$(1) \quad \Delta P(t) = \frac{\lambda}{\gamma}\left(1+\exp\left(\frac{t+\beta\ln(\gamma)-\delta}{\beta}\right)\right)^{\frac{-\gamma-1}{\gamma}} \exp\left(\frac{t+\beta\ln(\gamma)-\delta}{\beta}\right)(\gamma+1)^{\frac{\gamma+1}{\gamma}}; \quad \text{for } \gamma \geq 1,\ \beta \neq 0$$

The function comprises four independently adjustable parameters which describe different characteristics of the observed ΔP distributions over time. Figure 5 provides an illustration of how variation in each of these parameters affects the shape of the function: each plot in the figure shows three curves associated with three different settings of an individual parameter while keeping the remaining parameters constant. As can be seen, there is one parameter (the peak amplitude λ) which captures variation in overall interpretation strength (the maximum ΔP value achieved), while the remaining parameters all characterize differences in dynamics. The β parameter, for instance, describes the width of the distribution over time. Its interpretation is comparable to that of the *rate* parameter known from bounded exponential models in SAT-analysis: a higher β value results in a wider peak distribution over time, i.e., a slower rise to the peak value in the left tail of the distribution and a slower decline from the peak value in the right tail (but see below). The location parameter δ has an interpretation similar to (though not quite the same as) the *intercept* parameter in bounded exponential SAT models. It provides a millisecond index of the peak location in time: lower values of δ imply an early peak (fast processing), higher values a late peak (slow processing). The symmetry parameter γ does not possess an analog in bounded exponential models, where Y-values rise to an asymptote rather than a peak. In the present *LPP* model, γ determines whether interpretation strength is distributed symmetrically around the peak ($\gamma = 1$) or not ($\gamma > 1$); in combination with a positive width-parameter ($\beta > 0$), increasingly higher γ-values imply an increasingly slower decline from the peak in the right tail of the distribution (as shown in the figure); in combination with a negative width-parameter ($\beta < 0$), increasingly higher γ-values imply an increasingly slower rise to the

[7] Instrument noun onsets were virtually the same across conditions (see Figure 3).

peak in the left tail of the distribution (which would produce a mirror-inverted image of the curves in the bottom right panel of the figure). Thus, by varying the sign of β, the function is capable of modeling both positively and negatively skewed distributions. Importantly, in case of an asymmetry ($\gamma > 1$), the rate of processing in the left tail of the distribution (i.e., before reaching the peak value) is appropriately described by β, provided β is positive; if β is negative, rate of processing in the left tail of the distribution is better captured by γ (this complication in the parameter relations will be dealt with in Section 4.2.3).

To determine the amount of parameter variation necessary to describe the cross-condition differences in Figure 4, we employed a *competitive nested model fitting* approach, exploring a range of possible models starting with a simple 1λ-1β-1δ-1γ model (adjusting a single amplitude, width, location, and symmetry parameter to all three conditions) and ending with a full 3λ-3β-3δ-3γ model (fitting a unique set of parameters to each of the three conditions). This was done not only for the grand average data (Figure 4), but also for subsets of data (corresponding to the eight sub-groups of participants that shared the same stimulus files)[8] so as to explore the consistency of the model fits. The models were compared in terms of the *adjusted* R^2 statistic (explained variance adjusted by degrees of freedom – increasing numbers of parameters lead to a decrease in *adjusted* R^2 unless a substantially improved fit is achieved). Importantly, we also determined whether a given model produced any systematic residuals that could be explained by additional parameters. The analyses were performed in TableCurve-2D, using the Levenberg-Marquardt fitting algorithm.

It turned out that an 11-parameter model (3λ-3β-3δ-2γ) yielded the best description both of the grand average data and of the eight data subsets. The model comprised a separate λ, β and δ estimate for each condition, while two of the three conditions shared the same γ estimate (see below). The model achieved an *adjusted* R^2 of .951 on the grand average data, ranging from .947 to .965 across conditions. Across data subsets, *adjusted* R^2s ranged from .867 to .918. Models with less parameter variation produced systematic misfits and comparatively low *adjusted* R^2 values; the full 3λ-3β-3δ-3γ model, on the other hand, obtained slightly lower *adjusted* R^2s due to unnecessary parameter variation. The amount of parameter variation in the best model suggests systematic cross-condition differences in the overall strength of interpretation (variation in λ) as well as in processing dynamics (variation in β, δ, and γ). These can be explored more fully in terms of the model's actual parameter estimates in Table 2.

As can be seen in the table, there were consistent, but fairly minor differences in amplitude (λ). In particular, the metonymic verb condition still achieved around 95% of the maximum bias that was estimated for the preferred verb condition. This does not provide a lot of support for competition – be it immediate or mediated via coercion – which should have manifested itself in a considerably smaller amplitude (λ) for the metonymic verb condition than for the preferred verb condition (reduced interpretation strength due to competing interpretations for metonymic verbs).

In stark contrast to this, the remaining parameter estimates in Table 2 indicate rather pronounced cross-condition differences in processing dynamics, consistent with the assumption that enriched composition incurs extra processing costs: in comparison to the preferred verb condition, the metonymic verb condition engendered substantially decelerated processing, as indicated by consistently larger width (β) and location (δ) estimates; the β estimates were all positive, indicating a slower rate of processing in the left tail of the distribution (i.e., before reaching the peak value) for the metonymic rather than preferred verb condition; also, maximum interpretation strength was ac-

[8] Unfortunately, the range of possible models to be explored, as well as the number of data points required to achieve reasonably stable parameter estimates, rendered analyses by participants or items unfeasible.

	Parameter			
Condition	Amplitude (λ)	Width (β)	Location (δ)	Symmetry (γ)
Metonymic verb	.189	296	2754	1.06
	$(.192 \pm .007)$	(290 ± 47)	(2752 ± 61)	(1.12 ± 0.11)
Preferred verb	.198	159	2403	25.71
	$(.201 \pm .006)$	(153 ± 39)	(2391 ± 59)	(25.71 ± 7.98)
Non-preferred verb	.207	459	2893	1.06
	$(.208 \pm .006)$	(447 ± 58)	(2881 ± 68)	(1.12 ± 0.11)

Table 2: Parameter estimates per condition derived from the best *LPP* fit of the grand average data. Figures in parentheses refer to mean parameter estimates (with 95% confidence limits) across the eight data subsets, based on the same 3λ-3β-3δ-2γ model.

cumulated about 350 ms later in the metonymic verb condition than in the preferred verb condition (difference in δ). The slowest condition, both in terms of processing rate and peak location, was the non-preferred verb condition, presumably because the event denoted in this condition (*artist analyzing picture*) is rather atypical compared to the (*artist painting picture*) interpretation that is generated both in the preferred verb condition and – at least temporarily – in the metonymic verb condition.

Finally, there were marked differences in symmetry (γ): while processing rate was more or less symmetrically distributed around the peak in both the metonymic verb and the non-preferred verb condition ($\gamma \approx 1$), there was a clear positive skew in the preferred verb condition (positive β and $\gamma > 1$), indicating a very slow decline from the peak in the right tail of the distribution. This corresponds with the observation in Section 4.2.1, that the visual preference for the preferred instrument region lasted much longer in the preferred verb condition than in the metonymic verb condition. Indeed, Figures 3 and 4 indicate that in the preferred verb condition, the visual bias towards the preferred instrument region extended well into time periods where the instrument noun was available, whereas in the metonymic verb condition, the corresponding bias declined quite rapidly after the onset of the instrument noun (apparently because half of the time, this instrument noun was incompatible with the preferred interpretation). The lack of such a symmetry contrast between the metonymic verb and the non-preferred verb condition (in spite of the latter showing a longer-lasting visual bias in Figures 3 and 4) is most likely due to the fact that the relevant difference was captured in width (β) rather than symmetry (γ).

4.2.3. Parameter Validation

One objection against the previous time course analyses might be that the *LPP* function in equation (1) is rather complex and prone to parameter tradeoff. Recall that in this model, processing rate in the left tail of the data distributions (which is of particular interest in that it mostly reflects how quickly visual preferences accumulate before the critical instrument noun is available) is conjointly determined by two parameters, β and γ. It could be that for some conditions or data sets, processing rate before the peak was better captured by β, while in other conditions or data sets, it was more appropriately described by γ. This might question the validity of our previous claims about cross-condition differences in processing rate, particularly for time periods before the instrument noun has been processed.

We therefore tried to replicate our findings by fitting a simpler model to the left tail of the

Condition	Parameter		
	Asymptote (λ)	Rate (β)	Location (δ)
Metonymic verb	.249	293	2372
	$(.249 \pm .041)$	(272 ± 45)	(2356 ± 112)
Preferred verb	.217	160	1888
	$(.225 \pm .008)$	(159 ± 23)	(1898 ± 52)
Non-preferred verb	.271	415	2270
	$(.272 \pm .013)$	(412 ± 48)	(2282 ± 78)

Table 3: Parameter estimates per condition derived from the 9-parameter sigmoid fit of the truncated grand average data. Figures in parentheses refer to mean parameter estimates (with 95% confidence limits) across the eight data subsets, based on the same 3λ-3β-3δ sigmoid model.

data distributions in Figure 4. For this purpose, only data points between verb onset (zero) and the peak location, as estimated by the best *LPP* fit of the data (respectively, the time slot closest to the relevant peak location), were considered in each condition, as shown in Figure 6.

The model used for fitting these truncated data distributions was a standard sigmoid function, as defined in equation (2) below.

$$\Delta P(t) = \frac{\lambda}{1 + \exp\left(-\frac{t-\delta}{\beta}\right)} \text{ ; for } \beta \neq 0 \quad (2)$$

The sigmoid function describes a symmetrically S-shaped curve in terms of three parameters: an asymptote λ which determines an upper limit on ΔP at maximum time, a rate parameter β indexing the speed of transition from zero to asymptote, and a location parameter δ which determines the point in time where $\Delta P = 0.5\lambda$ (the central inflexion point of the S-curve).

Informed by the findings in Section 4.2.2, we fitted a 3λ-3β-3δ sigmoid model to the truncated data in Figure 6 (the model fits are indicated by solid curves in the figure). The model achieved a mean *adjusted* R^2 of .939 on the grand average data, ranging from .932 to .946 across conditions. Across data subsets, *adjusted* R^2s ranged from .848 to .893. Table 3 shows the relevant parameter estimates. As can be seen, the previous results, especially for the processing rate parameter β, were closely replicated: the metonymic verb condition was associated with a higher β (meaning slower processing) than the preferred verb condition, and the non-preferred verb condition produced the highest β (in fact, these processing rate estimates were numerically very close to those derived from the *LPP* model, see Table 2). Cross-condition differences in the other two parameters (λ and δ) were also comparable to those observed in the relevant *LPP* counterparts. However, it should be noted that the sigmoid model tended to produce rather excessive asymptote estimates for our data distributions, presumably due to underfitting. The crucial point is that both models converge on the fact that there was substantial processing slowdown in the metonymic verb condition (as well as the non-preferred verb condition) relative to the preferred verb condition. There was, however, little or no evidence for competition in the metonymic verb condition relative to the preferred verb condition, as would have become manifest in substantially lower λ estimates for the former.

4.2.4. Looks to Competitor Instruments

The previous time course analyses all relied on probability differences (looks to preferred instruments minus looks to dispreferred instruments, ΔP) as a measure of interpretation strength. The measure essentially reflects how strongly perceivers discriminate between preferred and non-preferred instruments in the display, enabling us to model the *interaction* between condition (metonymic vs. preferred vs. non-preferred verb) and type of instrument (preferred vs. non-preferred) as a function of time during picture viewing.

From a statistical point of view, the use of probability differences may not be without problems. Higher proportions of looks to one instrument entity are likely to entail lower proportions of looks to the other instrument entity, suggesting that part of the information contained in ΔP is redundant. However, note that this does not necessarily compromise our previous conclusions. First, proportions of looks to either of the two instrument entities were not fully complementary: in a considerable number of cases, looks to a given instrument entity were launched at the expense of looks to a non-instrument entity (e.g., the artist, the painting or the background) rather than looks to the alternative instrument entity. Second, and more importantly, redundancies in ΔP are bound to *amplify*, not attenuate, any cross-condition differences in the visual preferences of interest. In our view, this renders the lack of conclusive evidence for competition even more striking.

A more theoretical concern might be that some competitive accounts emphasize differences in looks to dispreferred entities (so-called *competitor objects*) more than differences in looks to preferred, *target* entities (see, e.g., Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001) – our previous analyses, by contrast, treated both as equally important. In the following analyses, we therefore focused only on probabilities of looks to competitor instruments and compared them between the two most critical experimental conditions (metonymic verb vs. preferred verb). Figure 7a shows the relevant data, spanning a time period from 1000 ms (sentence onset) until 7000 ms (end of picture presentation) in 50 ms resolution. Note that the data for the metonymic verb condition are again collapsed across the two spoken versions, ending in either the preferred or the non-preferred instrument noun.

As the figure indicates, there were indeed slightly higher proportions of gazes on the competitor instrument (magnifying glass) in the metonymic verb condition compared to the preferred verb condition, most markedly so in a time period of approximately 200–950 ms before the onset of the instrument noun, and in a time period of approximately 350–1850 ms after the onset of the instrument noun. This would suggest a certain degree of competition in the metonymic verb condition. However, what can also be seen from the corresponding 95% confidence intervals in Panels (b) and (c) of the figure (these are equivalent to a series of paired samples t-tests, each assuming significance at $p \leq .05$.) is that the *early* difference was significant (by participants and items) only within a single 50 ms time slot which comprised a potential negative outlier in the preferred verb condition (marked by an arrow in Figure 7a). Thus, evidence for competition before the onset of the instrument noun appears to be rather faint, if not spurious. The *late*, post-instrument difference, by contrast, extended over longer time periods and was comparatively robust. However, this post-instrument difference is hardly surprising given that half of the metonymic verb trials ended in the non-preferred instrument noun, whereas preferred verb trials always ended in the preferred instrument noun.

We also fitted *LPP* functions to each of the data series in Figure 7a. However, these obtained comparatively poor fits (R^2s $< .8$) or no solution at all due to insufficient systematic variation over

time in some data subsets. The previously reported probability differences did not suffer from this problem.

5. General Discussion

Previous experimental studies have shown that participants experience processing difficulty when reading metonymic constructions such as *the artist started the picture*. This slowdown in processing has often been attributed to enriched composition, which claims that additional structure has to be constructed when a noun referring to an artifact (such as *picture*) is coerced into the event representation required by the verb *start* (McElree et al., 2001; Traxler et al., 2002; Lapata et al., 2003; McElree et al., 2005). However, as discussed in the introduction of Section 4, such differences in reading time could also be due to the fact that readers compute multiple interpretations for metonymic verbs (e.g., *the artist started painting/analysing/framing the picture*) which compete with one another and thus decelerate the process of establishing an interpretation that is unambiguous enough for the reader to decide to move on in the text.

In this paper, we employed the visual-world paradigm to distinguish between three possible accounts of metonymic verb interpretation: Serial Coercion (where difficulty associated with metonymic verbs is solely due to enriched composition of a single interpretation), Parallel Coercion (where enriched composition triggers a competition between alternative interpretations), and Immediate Competition (where competition alone is sufficient to explain the difficulty associated with metonymic verbs).

Participants listened to metonymic sentences such as *the artist started the flowery picture using the depicted . . .* while at the same time looking at a visual array comprising instruments for two potential interpretations of *start*, e.g., a palette with paint brushes (for the dominant *started painting* interpretation) and a magnifying glass (for the subordinate *started analyzing* interpretation). This setup allowed us to investigate the time course of metonymic verb interpretation: proportions of looks to the depicted instruments were taken as an indicator of the kinds of interpretations that listeners pursue at any given point in time, as well as of the strength of commitment to one of these interpretations against potential alternative interpretations.

Detailed analyses of how visual preferences for the relevant instrument entities developed over time revealed that processing was substantially slowed down in metonymic verb constructions relative to non-metonymic constructions with comparable interpretations. This became evident in substantially slower processing rates and delayed peak locations for the metonymic verb condition after fitting a *Logistic Power Peak* model to gaze probability differences over time. Enriched composition predicts such a slowdown for metonymic verbs via type-shifting of an artifact-denoting object noun into an event representation required by this type of verb. Hence, models that assume this costly type-shifting operation to take place in metonymic verb constructions (i.e., Serial or Parallel Coercion) can easily explain the obtained differences in processing dynamics. Immediate Competition, on the other hand, fails to explain those differences because no mechanism other than competition is provided in this framework.[9]

Most importantly, our experiment also revealed that cross-condition differences in amplitude (the estimated maximum visual preference for the preferred instrument in each condition, taken to

[9]This is not to say that competition-based frameworks are generally incapable of modeling differences in processing dynamics. In some implementations, the temporal behavior of the system can be modulated, for example, by varying the timings with which different constraints enter the competition (see McRae et al., 1998). However, any such solution would require additional theoretical as well as empirical justification in our view.

indicate strength of commitment to the relevant interpretation) were virtually negligible. In other words, our experiment failed to provide evidence for the simultaneous activation of multiple interpretations in metonymic verb constructions, a conclusion that is further corroborated by the lack of a convincing effect in looks to competitor instruments, see Section 4.2.4. These findings are difficult to reconcile not only with the Immediate Competition account but also with the Parallel Coercion account – both assume a stronger degree of competition in metonymic verb constructions as opposed to non-metonymic controls. Taken together, this leaves us with the Serial Coercion account as the best explanation of our data.

It seems likely that further experimental work is required to conclusively establish that there are no competition effects in the interpretation of metonymic verbs. An additional, very informative test would be an experiment in which *irrelevant* instruments are shown alongside preferred and non-preferred instruments in the visual display (thus following more closely the design in Allopenna et al., 1998, for example).[10] That is, the example picture would not only include the paintbrushes (preferred instrument) and the magnifying glass (non-preferred instrument), but also, e.g., a jackhammer (an irrelevant instrument in the sense that it is not compatible with any of the interpretations triggered by *The artist started the picture* ...). Given such a design, a parallel/competitive account predicts that in the metonymic verb condition, the non-preferred competitor instrument (magnifying glass) would attract more anticipatory looks than the irrelevant instrument (jackhammer). This is because, of these two instruments, only the competitor instrument would relate to one of the competing interpretations taken into consideration by the processor. The Serial Coercion account, on the other hand, predicts that there should be no difference in proportions of looks to competitor vs. irrelevant instruments – both would be largely ignored since only the instrument compatible with the dominant interpretation (i.e., the preferred instrument) would be taken into consideration. We are currently conducting a follow-up experiment to test these predictions.

Most authors (starting with McElree et al., 2001) hypothesize that the processing of coerced verbs requires the comprehender to construct additional structure, which leads to increased processing effort, and therefore to increased reading times (or to a deceleration effect such as the one observed in the present study). However, there have been no attempts to explain the cognitive mechanisms that underlie the hypothesized construction of additional structure. In fact, the only explicit model of metonymic verb interpretation that we are aware of is the one of Lapata et al. (2003). They propose a Bayesian account where the comprehension of a metonymic expression is modeled as the computation of i, the interpretation which maximizes $P(i,v,o,s)$, the joint probability of i, the metonymic verb v, its subject s, and object o. This probability can be broken down as:

$$(3) \qquad \arg\max_i P(i,v,o,s) = \arg\max_i P(i)P(o|i)P(v|i,o)P(s|i,o)$$

This equation implicitly assumes a serial coercion mechanism, as it returns a single interpretation i (the one that maximizes the joint probability); a parallel coercion model would return a list of interpretations (possibly ordered by probability). It is important to note, however, that Lapata et al.'s (2003) model only captures the process of computing the preferred interpretation of metonymic constructions (such as *the artist started the picture*). It does not deal with the interpretation of non-metonymic constructions (such as *the artist painted the picture*). This means it offers only a partial account of the data from Experiment 2, as it does not allow a comparison between the metonymic

[10] We owe this suggestion to an anonymous reviewer.

and the non-metonymic conditions. Extending the model to handle non-metonymic constructions is a topic for future research.[11]

How do our data compare to the results from a recent SAT study investigating the time course of metonymic verb interpretation (McElree et al., 2005)? In this experiment, which was based on sensicality judgments, metonymic verb constructions were found to be processed less accurately and, most importantly, more slowly than non-metonymic controls. The difference in accuracy could be taken as evidence for readers being less likely to compute a sensible eventive interpretation in metonymic constructions. The difference in processing speed is, again, consistent with a time-consuming enriched composition process. Using a different paradigm, our experiment clearly confirmed the latter finding, while the difference in accuracy does not seem to have a direct equivalent in our data. However, it has to be noted that the two experiments measured two qualitatively rather different aspects of comprehension – sensicality judgments on the one hand and looks to instrument entities associated with different verb interpretations on the other. While McElree et al.'s (2005) study was primarily concerned with the problem of whether readers can easily come up with a sensible interpretation for metonymic verb constructions, our study was tailored around the question of *how many* different interpretations are computed for such verbs, and whether there is evidence for on-line competition between different metonymic verb interpretations. The different interpretations were always visually supported in our paradigm, making it more likely that listeners came up with a sensible interpretation. However, the lack of a convincing competition effect in our data (in spite of the fact that our experimental setup should *encourage* the generation of multiple interpretations, given that instruments associated with different interpretations are always present in the visual display) strongly suggests that only a single, most dominant metonymic-verb interpretation is computed at a given time. In this respect, our findings go beyond the conclusions from McElree et al. (2005).

Importantly, the relevant interpretational preferences (as evidenced by visual biases towards associated instrument entities) were already established before sufficient information about the indicative instrument noun became available in the linguistic input (see Section 4.2.1). Our visual world experiment therefore adds to a growing body of research which shows that listeners are able to anticipate forthcoming reference to entities in the visual display on the basis of incremental interpretation of concurrent spoken material (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Knoeferle et al., 2005; Arai et al., 2006).

Future research will show whether the present anticipatory effects were dependent on recognizing the word *using*, or on integration of the preceding complement noun *picture* (see Footnote 3). The fact that instruments *were* anticipated in the present experimental setting has important implications, however. It suggests that the observed cross-condition differences in processing dynamics were triggered by comprehension processes that started before the instrument noun was available, even though the slower experimental conditions reached their maximum interpretational biases during time periods beyond the onset of the instrument noun (see Figure 4).

A final point concerns the processing of non-preferred control constructions such as *the artist analyzed the picture*. In line with previous findings from reading (McElree et al., 2001; Traxler et al., 2002), we found evidence for a processing slowdown in this type of constructions relative to preferred non-metonymic controls. This effect can be attributed to the fact that non-preferred con-

[11] Another problem is the fact the model is not incremental as it stands; to account for time course data, the model would have to be extended to work with partial information (e.g., if only the verb is available to compute the preferred interpretation, but not the object).

structions denote rather atypical situations which generally make less sense to language comprehenders. In our experiment, we found that non-preferred sentences caused even higher processing costs than coerced metonymic sentences, which is somewhat puzzling given that reading studies suggested the latter to be slightly harder to process than the former. Our data might indicate that prediction of an instrument (which forms the basis of the present findings, in contrast to previous reading data) is particularly difficult for less stereotypical events such as *the artist analyzed the picture*; in stereotypical events such as *the artist painted the picture* or the preferred interpretation of *the artist started the picture*, prediction of the instrument may be relatively easy. This could be an interesting question for further investigation. Future work might also include other types of logical metonymy (e.g. adjective-noun combinations), and enriched composition processes outside the domain of logical metonymy (Todorova, Straub, Badecker, & Frank, 2000; Pullman, 1997).

In conclusion, the present research provided evidence for the cost of enriched composition, confirming earlier results from reading and SAT experiments. However, previous findings were ambivalent as to whether processing difficulty associated with metonymic verbs is solely due to enriched composition, or whether competition between alternative interpretations licensed by such verbs could explain at least part of the difficulty. The present data suggest that the answer to the latter is negative.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye-movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 529–543.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Arai, M., Gompel, R. P. G. van, & Scheepers, C. (2006). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *forthcoming*.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, *9*, 5–16.

Briscoe, T., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. In *Proceedings of 13th International Conference on Computational Linguistics* (pp. 42–47). Helsinki: Association for Computational Linguistics.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 84–107.

Copestake, A. (1995). Representing lexical polysemy. In J. Klavans (Ed.), *Proceedings of the aaai spring symposium on representation and acquisition of lexical knowledge: Polysemy, ambiguity and generativity* (pp. 21–26). Stanford, CA: American Association for Artificial Intelligence.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Journal of Memory and Language*, *38*, 419–439.

de Almeida, R. G. (2004). The effect of context on the processing of type-shifting verbs. *Brain and Language*, *90*, 249-261.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 116-124.

Jackendoff, R.(1997). *The architecture of the language faculty*. Cambridge MA: MIT Press.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M.(2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*, 37–55.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Lakoff, G., & Johnson, M.(1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lapata, M., Keller, F., & Scheepers, C.(2003). Intra-sentential context effects on the interpretation of logical metonymy. *Cognitive Science*, *27*(4), 649–668.

Lascarides, A., & Copestake, A.(1998). Pragmatics and word meaning. *Journal of Linguistics*, *34*(2), 387–414.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

McElree, B., & Dosher, B. A.(1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*(122), 291–315.

McElree, B., & Griffith, T.(1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 134–157.

McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. J. (2005). A timecourse analysis of enriched composition. *Psychonomic Bulletin and Review, forthcoming*.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R.(2001). Reading time evidence for enriched composition. *Cognition*, *78*, B17–B25.

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*(3), 283–312.

Nunberg, G.(1995). Transfers of meaning. *Journal of Semantics*, *12*(1), 109-132.

Partee, B.(1992). Syntactic categories and semantic type. In M. Rosner & R. Johnson (Eds.), *Computational linguistics and formal semantics* (pp. 97–126). Cambridge: Cambridge University Press.

Pickering, M., McElree, B., & Traxler, M. J.(2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, *93*, 1–9.

Pullman, S.(1997). Aspectual shift as type coercion. *Tansaction of the Philological Society*(95), 279–317.

Pustejovsky, J.(1991). The generative lexicon. *Computational Linguistics*, *17*(4), 409–441.

Pustejovsky, J.(1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191–201.

Rayner, K., & Pollatsek, A.(1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Reed, A. V.(1976). The time course of recognition in human memory. *Memory and Cognition*, *4*, 16–30.

Rodd, J., Gaskell, G., & Marslen-Wilson, W.(2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266.

Seidenberg, M. S., & MacDonald, M. C.(1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (2000). Modelling discourse context effects: A multiple constraints approach. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

Todorova, M., Straub, K., Badecker, W., & Frank, R.(2000). Aspectual coercion and the on-line computation of sentential aspect. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Philadelphia, PA: Cognitive Science Society.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*, 530–547.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Prespectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vendler, Z.(1968). *Adjectives and nominalizations*. The Hague: Mouton.

Wickelgren, W. A., Corbett, A. T., & A.Dosher, B.(1980). Priming and retrieval from short-term memory: A speed-accuracy tradeoff analysis. *Journal of Verbal Learning and Verbal Behavior*, *19*, 387–404.

Appendix A. Materials

Linguistic materials selected for Experiment 2 (the corresponding pictures can be obtained from the first author). For each item, the relevant verb triplet is provided, consisting of metonymic verb / preferred verb / non-preferred verb, along with the corresponding preferred / non-preferred instrument noun pair.

(6) The artist {started / painted / analysed} the flowery picture using the depicted {paintbrushes / magnifying glass}.

(7) The engineer will {start / write / read} the urgent memo using his new {ballpoint pen / pair of glasses}.

(8) The editor will {finish / write / read} the leading article using his old {ballpoint pen / pair of glasses}.

(9) The interior designer will {begin / design / decorate} the new kitchen using the depicted {drawing-board / wallpaper}.

(10) The editor will {finish / edit / read} the drafted newspaper using his good old {pencil / desk lamp}.

(11) The publisher will {begin / publish / read} the exciting novel using the practical {computer / desk lamp}.

(12) The expert {started / evaluated / painted} the valuable picture using his new {magnifying glass / paintbrushes}.

(13) The director will {start / write / read} the new script using a conventional {typewriter / desk lamp}.

(14) The banker will {start / make / drink} the morning coffee using his own {coffee maker / cup}.

(15) The boy will {finish / write / send} the letter for Santa Claus using a blue {fountain pen / stamp}.

(16) The mechanic will {finish / repair / wax} the articulated lorry using his special {toolkit / car polish}.

(17) The teenager will {begin / read / write} a new novel using his stylish {glasses / ballpoint pen}.

(18) The student {finished / read / wrote} the heavy book using her new {pair of glasses / pen}.

(19) The chef will {start / cook / eat} the rich dinner using the depicted {frying pan / cutlery}.

(20) The composer will {begin / write / direct} a grandiose symphony using his beloved {quill / baton}.

(21) The builder will {start / build / demolish} the small house using his approved {bricks / wrecking ball}.

(22) The guitarist will {attempt / play / sing} a spirited solo using his new {electric guitar / microphone}.

(23) The diner will {start / eat / cook} the luscious meal using his high-quality {cutlery / frying pan}.

(24) The cook will {try / taste / buy} the hot spices using his old {spoon / purse}.

(25) The mechanic will {finish / repair / switch off} the old TV using his all-purpose {tools / remote control}.

(26) The builder will {master / drive / construct} the winding road using the impressive {sports car / steamroller}.

(27) The diva will {enjoy / sing / watch} the ambitious aria using her new {microphone / binoculars}.

(28) The carpenter will {finish / clean / plane} the wooden commode using his special {polish / planer}

(29) The distiller will {begin / make / taste} the superb whisky using a traditional {pot still / glass}.

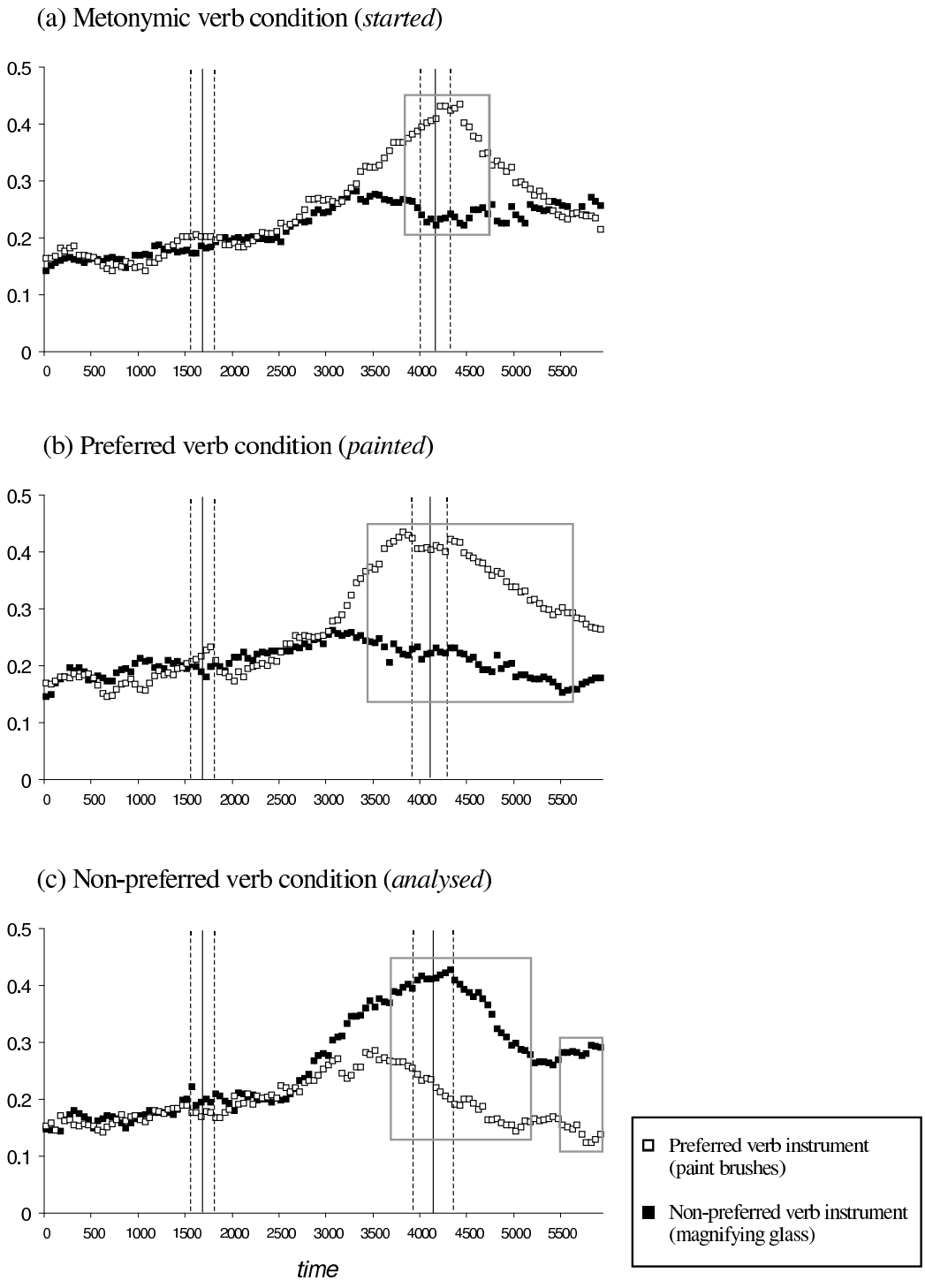

Figure 3. Gaze probabilities (by time steps of 50 ms) for the two critical target regions (preferred vs. non-preferred instrument) in each experimental condition. Solid vertical lines represent the average onsets of the verb and the instrument noun in each condition, dotted lines indicate 99% confidence limits (by items) for these onsets. The gray boxes highlight time periods where numbers of gazes differed significantly between the two target regions (*binomial* $p < .01$).

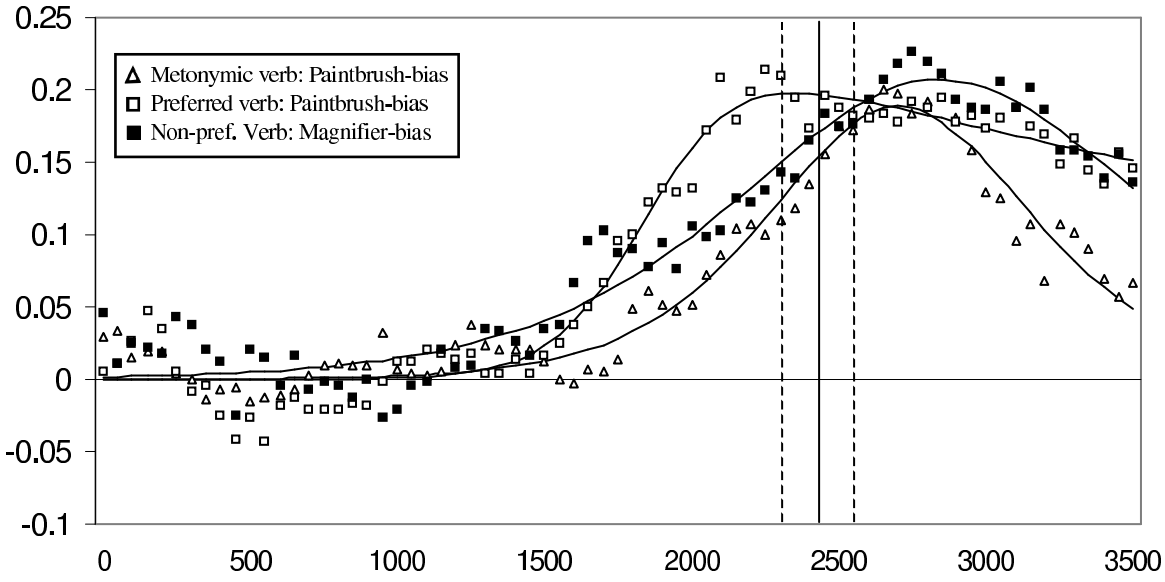

Figure 4. Gaze probability differences (Δ*P*) from the onset of the verb until about 1000 ms after the onset of the instrument noun. Solid lines indicate the best grand average fit of the data, based on a 11-parameter *Logistic Power Peak* model (see text).

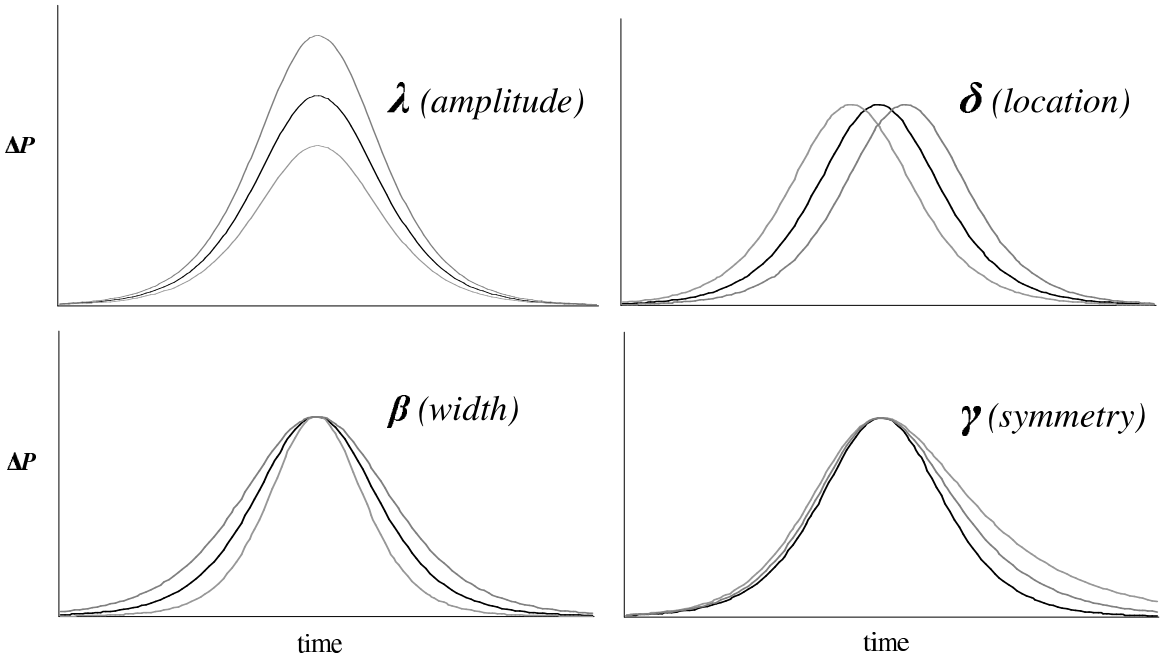

Figure 5. Illustration of the *logistic power peak* (*LPP*) parameters used to fit the probability difference distributions in Figure 4.

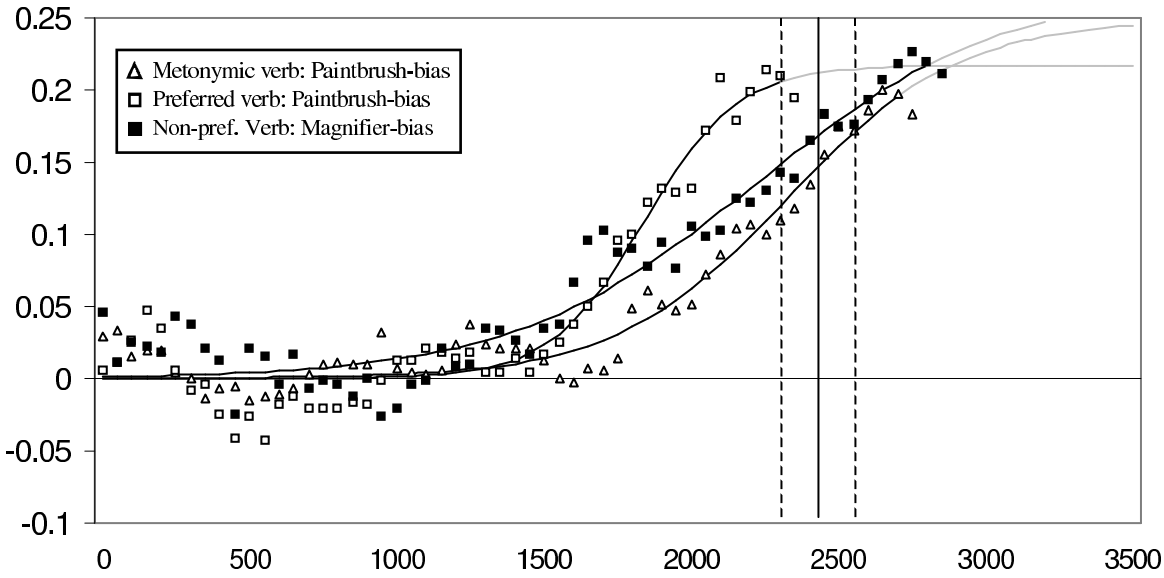

Figure 6. Gaze probability differences (ΔP) from the onset of the verb until the peak location derived from the best *LPP* model (see previous section). Solid lines indicate the best grand average fit of the data, based on a 9-parameter *sigmoid* model (see text).

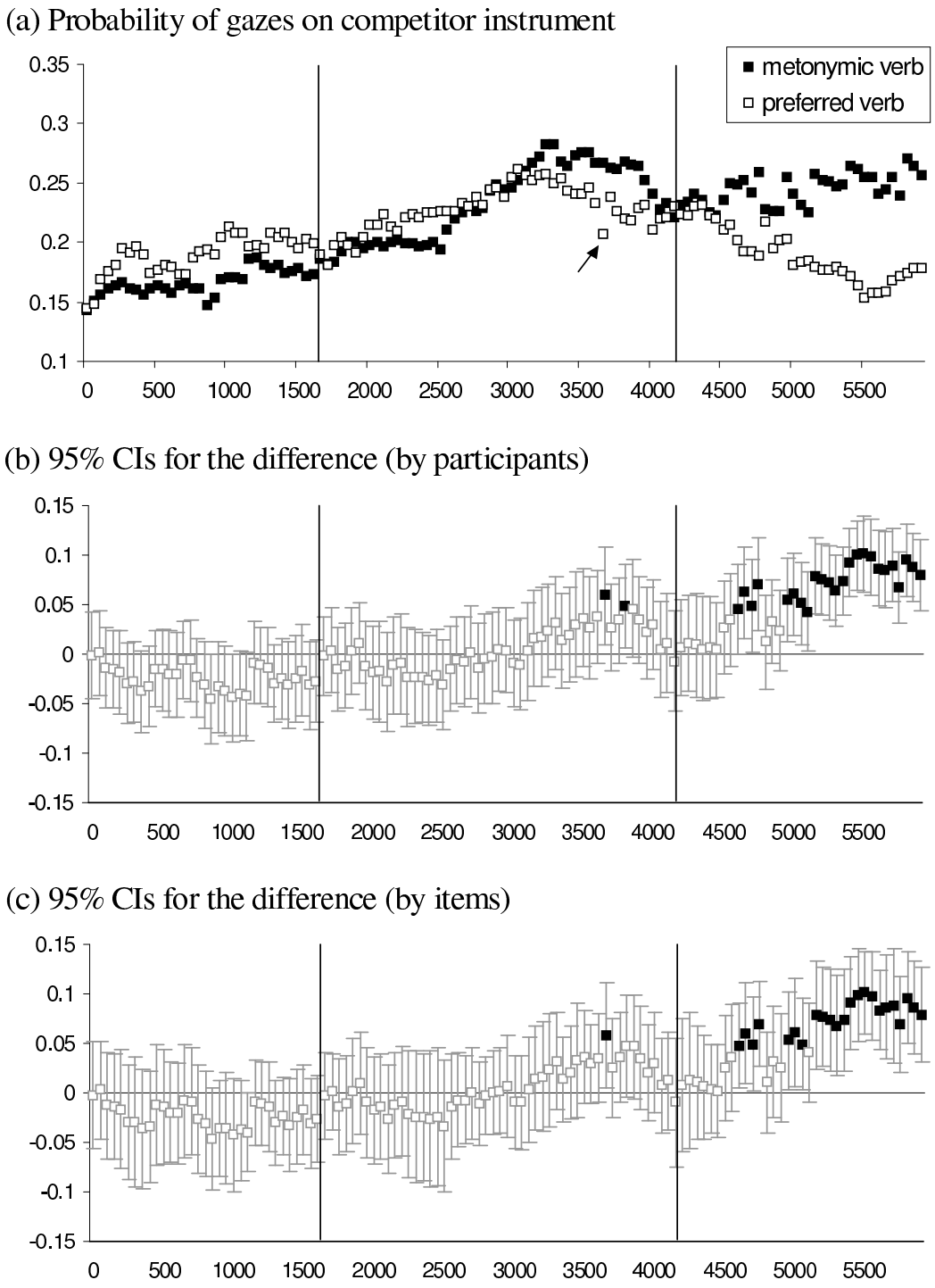

Figure 7. Top panel: probabilities of gazes on the competitor instrument (magnifying glass) for the metonymic verb and the preferred verb condition. Solid vertical lines represent the average onsets of the verb and the instrument noun. Panels (b) and (c) show 95% confidence intervals (by participants and items, respectively) for the difference between the two conditions in each 50 ms time slot (positive values imply higher proportions of gazes on the competitor instrument in the metonymic verb condition). Significant differences are highlighted by filled symbols in Panels (b) and (c).