Dependency resolution difficulty increases with distance in Persian separable complex predicates: Implications for expectation and memory-based accounts

Safavi, Molood Sadat [1], Husain, Samar [2], and Vasishth, Shravan [3]*

[1] *Department of Linguistics, University of Potsdam, Potsdam, (IDEALAB program), Germany* [2] *Department of Humanities and Social Sciences, Indian Institute of Technology, India* [3] *Department of Linguistics, University of Potsdam, Potsdam, Germany*

Correspondence*:
Shravan Vasishth
Department of Linguistics, University of Potsdam Haus 14,
Karl-Liebknecht-Str.24-25, D-14476 Potsdam, Germany ,
vasishth@uni-potsdam.de

ABSTRACT

Delaying the appearance of a verb in a noun-verb dependency tends to increase processing difficulty at the verb; one explanation for this locality effect is decay and/or interference of the noun in working memory. Surprisal, an expectation-based account, predicts that delaying the appearance of a verb either renders it no more predictable or more predictable, leading respectively to a prediction of no effect of distance or a facilitation. Recently, Husain et al (2014) suggested that when the exact identity of the upcoming verb is predictable (strong predictability), increasing argument-verb distance leads to facilitation effects (consistent with surprisal), but when the exact identity of the upcoming verb is not predictable (weak predictability), locality effects are seen. We investigated Husain et al's proposal using Persian complex predicates (CPs), which consist of a non-verbal element ('noun' in the current study) and a verb. In such constructions, once the noun has been read, the exact identity of the verb is highly predictable (strong predictability); this was confirmed using a sentence completion study. In two self-paced reading (SPR) and two eye-tracking (ET) experiments, we delayed the appearance of the verb by interposing a relative clause (Expt. 1 and 3) or a long PP (Expt. 2 and 4). We also included a simple predicate (Noun-Verb) configuration with the same distance manipulation; here, the exact identity of the verb was not predictable (weak predictability). Thus, the design crossed Predictability Strength and Distance. We found that, consistent with surprisal, the verb in the strong predictability conditions was read faster than in the weak predictability conditions. Furthermore, greater verb-argument distance led to slower reading times; strong predictability did not neutralize or attenuate the locality effects. As regards the effect of distance on dependency resolution difficulty, these four experiments present evidence in favor of working memory accounts of argument-verb dependency resolution, and against the surprisal-based expectation account of Levy (2008). However, another expectation-based measure, entropy (which was computed using the sentence completion data) predicts reading times in Experiment 1 (but not the other experiments). Thus, memory overload and entropy are two alternative explanations for the locality effects in Persian.

Keywords: Locality, Expectation, Surprisal, Entropy, Persian, Complex Predicates, self-paced reading, eye-tracking

1 INTRODUCTION

A long-standing claim in sentence processing is that increasing distance in a linguistic dependency, such as a noun-verb dependency, leads to greater processing difficulty (**Chomsky**, 1965; **Just and Carpenter**, 1992; **Gibson**, 2000; **Lewis and Vasishth**, 2005); it is common to refer to this increase in processing difficulty as the locality effect. One explanation for the locality effect is in terms of constraints imposed by working memory. According to one account, the Dependency Locality Theory (DLT; **Gibson** (1998)), the processing difficulty experienced when resolving a long dependency depends on the decay experienced by the noun; a related account by **Lewis and Vasishth** (2005) attributes the locality effect to decay and/or interference. Constraints on working memory may be a plausible explanation given that individuals' working memory capacity seems to affect the processes involved in dependency resolution (**Nicenboim et al.**, 2015; **Caplan and Waters**, 2013). Although there is evidence consistent with the memory-based explanation in English, German, Chinese, Russian, and Hindi, (**Hsiao and Gibson**, 2003; **Grodner and Gibson**, 2005; **Bartek et al.**, 2011; **Vasishth and Drenhaus**, 2011; **Levy et al.**, 2013; **Husain et al.**, 2014, 2015), research on some of these languages has also uncovered evidence that increasing noun-verb distance facilitates processing at the verb(**Konieczny**, 2000; **Vasishth**, 2003; **Vasishth and Lewis**, 2006; **Jaeger et al.**, 2008; **Vasishth and Drenhaus**, 2011; **Levy and Keller**, 2013; **Husain et al.**, 2014; **Jäger et al.**, 2015). One explanation for these anti-locality effects is in terms of surprisal (**Hale**, 2001; **Levy**, 2008). Surprisal extends and formalizes the old idea of predictive sentence processing—which has been extensively investigated in the EEG literature (e.g., **Kutas and Hillyard** 1984)—in terms of probabilistic parse continuations (also see **Jurafsky** 1996). The surprisal account assumes that the comprehender maintains and uses linguistic knowledge probabilistically to parse a sentence incrementally. Surprisal is the claim that rare transitions are difficult: increased processing difficulty is predicted when a parser is required to build a low-probability syntactic structure. Formally, surprisal is defined as the negative log probability of encountering a particular part of speech or word given previous context. We will refer to surprisal as the expectation-based account, following the terminology of **Levy** (2008).[1]

In many of these studies, evidence has been found for both the memory-based account and the expectation-based account. One conclusion that has emerged is that both memory and expectation play a role. For example, in his eye-tracking study investigating processing difference in English object vs subject relative clauses, **Staub** (2010) finds evidence for both expectation-based processing and locality constraints, although these occur in different regions of the target sentence. An example of Staub's design is provided below. In this study, processing difficulty was found on the noun phrase *the fireman* in the ORC (object relative clause) 1b, compared to the SRC (subject relative clause) 1a; this is consistent with the expectation account because the reader would be forced to build a rare object relative in the ORC condition when he/she encounters the noun phrase. However, this study also found greater processing difficulty at the relative clause verb in ORCs than SRCs, which is predicted by memory accounts.

(1) a. The employees that noticed the fireman hurried across the open field.

b. The employees that the fireman noticed hurried across the open field.

As further examples, both **Vasishth and Drenhaus** (2011) and **Levy and Keller** (2013) have argued that locality effects may appear when high working memory load is experienced; anti-locality effects may be present when the load is low.

[1] Another expectation-based account in the literature is the entropy reduction hypothesis or ERH (**Hale**, 2006); we do not investigate ERH in this paper, but we will briefly discuss a related idea, entropy, in the General Discussion.

In a recent development, **Husain et al.** (2014) argue that the strong predictability for a head (predicting an exact lexical item) can neutralize the locality effect; locality may manifest itself only when predictability strength is weak, that is, when only a verb phrase is predicted, and not the exact identity of the verb. In their self-paced reading study, **Husain et al.** (2014) used a 2×2 design, crossing Predictability and Dependency Distance to investigate locality and anti-locality effects. In the strong predictability conditions, Hindi complex predicates were used. In these noun-verb sequences, the noun strongly predicted the upcoming light verb, e.g. the noun *khayaal*, 'care', strongly predicts the verb *rakhnaa*, 'put', in *khayaal rakhnaa*, literally, 'care put' ('to take care of'). The weak predictability condition, on the other hand, used the same verb used in the complex predicate, but the noun did not predict the verb. An example is *gitaar rakhnaa*, 'guitar put'; 'to put (down) a guitar'; here, the verb retains its literal meaning. Thus, when the reader see *gitaar*, they cannot predict the exact identity of the verb, because many other verbs are possible here (e.g., bought). To summarize, in the strong predictability condition, the noun predicted the exact identity of the verb, while in the weak predictability condition the exact identity of the verb was not predicted with high certainty—although *a* verb was predicted. The second factor, dependency distance, was manipulated by placing one to two adverbials between the nominal predicate/object and the verb in the short condition. The long condition had two to three intervening adverbials. Reading time was measured at the verb. The results showed that CP light verbs were read faster in long vs short distance conditions, but for non-CP verb there was a tendency towards a slowdown in long vs short conditions. Finally, there was weak evidence for an interaction (estimate on the log ms scale: 0.03, Bayesian 95% credible interval [-0.02, 0.07], posterior probability of the effect being greater than 0 was 0.77). That is, there was some evidence that with increased distance there was a speedup at the light verb in the CP conditions and a slowdown in the non-CP conditions. These results were interpreted by Husain and colleagues as strong predictability of the head canceling the locality effect, and the locality effect manifesting itself only when predictability strength was weak.

In the present study, we build on the work by **Husain et al.** (2014) described above. Husain and colleagues' work suggested that the strength of the predictability may modulate whether locality effects occur or not; we investigate the cross-linguistic generality of this claim using Persian, which, like Hindi, also has a complex predicate construction that allows us to manipulate strong and weak predictability. We turn next to a short discussion of the complex predicate construction in Persian as it relates to our experiments.

2 COMPLEX PREDICATES IN PERSIAN

Complex Predicates (CPs) (**Samvelian**, 2001) consist of a sequence containing a non-verbal element (often a noun) and a verb, where the meaning of the sequence is non-compositional. An example is shown in (2).

(2) Maryam be man latme zad
Maryam to me damage hit
'Maryam caused damage to me (Maryam harmed me).'

The verb, often called a light verb, lacks sufficient semantic force to function as an independent predicate (**Vahedi-Langrudi**, 1996; **Karimi-Doostan**, 1997; **Karimi-Doostan**, 2005) and can be combined with different types of non-verbal items such as nominal, adjectivals or prepositional phrases (**Dabir-Moghaddam**, 1997).

In our study, we used separable complex predicates as defined by **Karimi-Doostan** (2011). According to Karimi-Doostan, a complex predicate can be separated if it satisfies both of the following two conditions: (1) if the nominal part is a noun to which adjectives, demonstratives, and wh-words, etc. can be attributed, and (2) if this noun has an internal argument structure (referring to an action or event). From this perspective, Persian complex predicates are categorized in three groups: (1) predicative verbal nouns (e.g.

anja:m da:dan, perform+to give), (2) predicative nouns (e.g. *latme zadan*, damage+to hit), and (3) non-predicative nouns (e.g. *gush da:dan*, ear+to do). Among these three types, only the second one satisfies both of the conditions.

We began by independently validating our assumption that the CPs we used in our experiments are predictable and separable. We first conducted a norming study (a sentence completion task), to establish that the light verbs (of the separated CPs) are highly predictable when the nominal is provided, as compared to non-CP verbs in simple predicate conditions. We then conducted an acceptability rating study to determine how acceptable Persian CPs are when they get separated.

3 NORMING STUDIES

In order to prepare appropriate stimuli, two norming studies were run. The first study involved sentence completion and served to validate (i) whether the identity of the verb in the complex predicate is highly predictable, and (ii) whether the identity of the verb in the control condition is not predictable.

The second study involved acceptability rating; the goal was to choose complex predicates for our experiments which are separable. That is, we wanted to identify complex predicates which native speakers would find acceptable even if an intervener occurs between the noun-verb sequence.

The sentence completion study was carried out to derive the predictions of the expectation account. Previous work on expectation effects suggests that sentence completion data may be useful for this purpose. For example, **Levy and Keller** (2013) used sentence completion data to complement their corpus analyses for deriving their predictions. In their study, the key issue was whether the intervening material (e.g., a dative marked NP) leads to a prediction of a dative verb. Their Table 4 shows that the intervening material sharpened the expectation for the type of verb predicted. This shows that sentence completion data can be used to determine empirically whether the prediction for a specific verb or a verb type is sharpened by intervening material; in the Levy and Keller case, it makes sense that the intervener sharpens the expectation, but clearly the nature and content of the intervening phrases will be crucial in determining whether expectations are sharpened (**Konieczny**, 2000; **Grodner and Gibson**, 2005).[2] Similarly, **Husain et al.** (2014) used sentence completion to establish that the identity of the verb in a complex predicate is highly predictable given the preceding context, but the identity of the verb in a simple predicate is not (see their Table 4). A third example is **Jäger et al.** (2015); they used both corpus data and sentence completion to establish that a sentence starting with a determiner, classifier, and an adverb leads to the prediction of a relative clause continuation in Chinese, and that the conditional probability of a subject relative continuation is higher than that of an object relative continuation (see their Table 2). Given these previous results, we assume that sentence completion data is informative about the predictions of the expectation-based account.

3.1 SENTENCE COMPLETION STUDIES

Two groups (32 participants each) of Persian native speakers, who did not take part in any of the other experiments, participated in two sentence completion pre-tests in which they were asked to complete the sentences after they were presented the sentence fragment until the pre-critical word. For example, as shown in 3, subjects were shown incomplete sentences which they had to complete; in this example, the missing verb is shown in parentheses. The participants were allowed to complete the sentence with as many words as they wanted, but our interest was only in the first word that they would write, which would most likely be a verb. This allowed us to calculate the proportion of continuations in which the exact verb was produced.

[2] We return to this point in the General Discussion, where we discuss the effect of entropy on reading times.

(3) a. Ali a:rezouyee bara:ye man (kard) . . .
Ali wish-INDEF for 1.S (do-PST . . .
'Ali (made) a wish for me . . . '

b. Ali a:rezouyee ke besya:r doost-da:sht-am bara:ye man (kard) . . .
Ali wish-INDEF that a lot like-1.S-PST for 1.S (do-PST) . . .
'Ali (made) a wish that I liked a lot for me . . . '

The materials were exactly the same as the ones used in the experiments presented below. For experiment 1 items, the average prediction accuracy for the exact verb in the strong predictability conditions was 76% for the short condition and 74% for the long condition; for experiment 2 items it was 77% and 76% for the short and long conditions respectively. By contrast, the average prediction accuracy for the exact verb in the weak predictability conditions in experiment 1 was 22% and 20% for the short and long conditions; and in experiment 2 it was 19% and 22% for the short and long conditions. An analysis using generalized linear mixed models shows a main effect of predictability in both the first experiment (coef = -1.32 , SE = 0.07, z = -16.93) as well as the second experiment (coef = -1.46, SE = 0.08, z = -17.25). As is clear from the mean percentages for each condition, the light verbs used in the complex predicate conditions were highly predictable, and the heavy verbs used in the simple predicate conditions were highly unpredictable. It is also clear from this study that, in our materials, increasing the amount of intervening material does not render the upcoming verb more predictable. The additional information provided by the intervening material for predicting the upcoming verb has been suggested by **Konieczny** (2000) as one possible explanation for shorter reading times at the verb in long- vs short-distance conditions. Although this proposal is likely to be correct for some constructions (see discussion in **Grodner and Gibson** 2005), in our materials, the sentence completion data do not provide any evidence that the intervening words we used in our design sharpen the expectation for the verb.[3]

3.2 ACCEPTABILITY RATING OF SEPARABLE VS INSEPARABLE CPS

Because the noun-verb sequences must be separable for our design to work, we also carried out an acceptability rating pre-test to make sure that the separability of the complex predicates used in our study is acceptable to native speakers. We tested for the acceptability of different types of noun-verb dependencies by interposing a short prepositional phrase between them. Taking Karimi-Doostan's classification of complex predicates into account, 36 items from each of the three categories were selected and randomized to test 50 native speakers of Persian (these participants did not take part in any other experiments reported here). They were asked to rate the sentences from 1 (unacceptable) to 7 (completely acceptable). Every subject saw all items. The average acceptability ratings for predicative verbal nouns, predicative nouns and non-predicative nouns were 3.23 (first quartile 1, third quartile 5), 6.08 (first quartile 6, third quartile 7), and 3.12 (first quartile 1, third quartile 5) respectively. That is, items with predicative nouns were the most acceptable. We used all the 36 items of the predicative noun condition in our experiments 1, 2, and 32 items in experiments 3, 4 (see the Methods section of experiment 3 for an explanation).

4 EXPERIMENT 1

4.1 METHOD

4.1.1 Participants Forty-two subjects aged between 17-40 years old (mean 24 years) participated in this experiment in Tehran, Iran. All participants were native speakers of Persian and were unaware of the

[3] In fact, in our sentence completion data, as discussed in the General Discussion, entropy increases with distance.

purpose of the study. This study was carried out in accordance with the Helsinki Declaration, and letters of consent were obtained from all the participants.

4.1.2 Materials We created 36 experimental sentences with a 2×2 factorial design, manipulating predictability strength and distance between the object noun and verb. The short intervener was a prepositional phrase and the long intervener was a relative clause added before the prepositional phrase. In order to mask the experiment, we included 100 filler sentences with varying syntactic structures (see Supplementary materials). Here is an example of the target sentences:

(4) a. Strong predictability, short distance (PP)

Ali	a:rezouyee	bara:ye	man	kard	va...
Ali	wish-INDEF	for	1.S	do-PST	and...

‘Ali made a wish for me and...’

b. Strong predictability, long distance (RC+PP)

Ali	a:rezouyee	ke	besya:r	doost-da:sht-am	bara:ye	man	kard	va...
Ali	wish-INDEF	that	a lot	like-1.S-PST	for	1.S	do-PST	and...

‘Ali made a wish that I liked a lot for me and...’

c. Weak predictability, short distance (PP)

Ali	shokola:ti	bara:ye	man	xarid	va ...
Ali	chocolate-INDEF	for	1.S	buy-PST	and...

‘Ali bought a chocolate for me and ...’

d. Weak predictability, long distance (RC+PP)

Ali	shokola:ti	ke	besya:r	doost-da:sht-am	bara:ye	man	xarid	va...
Ali	chocolate-INDEF	that	a lot	like-1.S-PST	for	1.S	buy-PST	and...

‘Ali bought a chocolate that I liked a lot for me and....’

The critical region is the verb (*kard* and *xarid*).

4.1.3 Procedure Participants were tested individually using a PC. They were explained the task before they performed the self-paced reading (SPR) experiment. The participants were instructed to read for comprehension in a normal manner and had a practice session of five sentences. All the sentences were displayed on a single line and were presented in 22 pt Persian Arial font using Linger software (http://tedlab.mit.edue/ dr/Linger/). In order to read each word of a sentence successively in a moving window display, participants had to press the space bar; then the word seen previously was masked and the next word was shown. After each sentence, they were asked to answer a comprehension question to ensure that the participants paid attention to the complete sentence.

4.1.4 Data analysis The data analysis was conducted in the R programming environment (**R Development Core Team**, 2013), using linear mixed-effects models (LMMs; **Pinheiro and Bates** 2000; **Bates et al.** 2015). For large samples, the t-distribution approximates the normal distribution and an absolute value of t larger than 2 indicates a statistically significant effect at $\alpha = 0.05$. Sum contrasts were used to code main effects and interactions. In addition, a nested contrast was defined for a secondary analysis in order to look at the effect of distance in complex predicates vs the control conditions separately; these were also coded as sum contrasts. For the reading time data, the most complex model possible given the data and the design was chosen based on the `rePCA` function

(**Bates et al.**, 2015); see the package `RePsychLing` (https://github.com/dmbates/RePsychLing) for examples and more theoretical background. The `rePCA` function computes a principal components analysis of the variance covariance matrices for the random effects (subject and item), which allows the modeler to decide which variance components should be included. No attempt was made to fit correlations between intercepts and slopes, for subjects or for items. All data and code are available from http://www.ling.uni-potsdam.de/∼vasishth/code/SafaviEtAl2016DataCode.zip.

4.2 PREDICTIONS (EXPERIMENT 1)

Based on the **Husain et al.** (2014) results, in experiment 1, we expected that increasing noun-verb distance would lead to faster reading time at the verb in the strong predictable conditions, but slower reading time in the weak predictable conditions. Thus, we expected to obtain a cross-over interaction.

The memory based accounts (**Just and Carpenter**, 1992; **Gibson**, 2000; **Lewis and Vasishth**, 2005) predict that increasing distance should lead to a slowdown at the verb; these accounts make no predictions about the strength of predictability.

There are two alternative predictions possible for the expectation account, depending on how one operationalizes expectation. First, if sentence completion probabilities are a reasonable proxy for conditional probabilities—and the previous research reported above (**Husain et al.**, 2014; **Levy and Keller**, 2013; **Jäger et al.**, 2015) suggests that they may be—then we predict (a) no difference in reading time at the verb as a function of distance, and (b) faster reading time at the verb in the strong predictable conditions than the weak predictable conditions. Prediction (a) arises because, in the sentence completion data, we see no effect of distance on the predictability of the upcoming verb, in either the strong or weak predictability conditions; prediction (b) arises due to the difference in predictability of the exact verb that we see in the strong versus weak predictability conditions (see the results of the sentence completion studies).

An alternative possible prediction of the expectation account is that increasing distance should facilitate processing at the verb. Surprisal predicts facilitation with increasing distance whenever distance causes the number of possible parses to decrease; this decrease in the number of possible parses leads to the probability mass being reassigned among the remaining parses. In our materials, when the participant reads the noun in the noun-verb complex predicate, they are expecting the light verb with high probability (nearly 1). However, in the long distance condition, the next word begins a relative clause; this leads to an expectation that the light verb will appear *after* the relative clause verb. But what appears after the relative clause verb is a PP that modifies the upcoming light verb. For a facilitation to be predicted in this long-distance condition by the surprisal metric, it would have to be the case that the conditional probability of the light verb following the RC and PP would be higher than the conditional probability of the light verb in the short-distance (PP) condition. In order to get a sense of how the conditional probabilities change in the noun-light verb condition as a function of distance, we extracted all light verb sentences from a Persian corpus (**Seraji**, 2015) and then counted, for different numbers of modifying phrases, the proportion of cases that a verb followed the intervening phrase. For example, in a Persian sentence such as *John in the morning went*, there is one intervening phrase, the PP. As shown in Table 1, we find that the conditional probability of the verb appearing next is always high, but goes to 1 with increasing distance. This suggests that in general, increasing distance tends to sharpen the expectation for an upcoming verb. We also did this calculation using as a metric not the number of intervening phrases but the number of intervening words; the result, shown in Table 2, is substantially the same as in Table 1. Of course, these corpus counts don't give us any direct information about the predictions regarding our particular experiment design.

Regarding the strong vs weak predictability conditions, note that the expectation account of Hale and Levy does not predict that processing should be facilitated when the exact identity of the upcoming verb is predicted (strong predictability condition), compared to the case when just some verb is predicted (weak predictability condition). This is because the surprisal metric is usually calculated using the conditional probability of the part-of-speech (verb) given preceding context, and this will be the same in both the

Table 1. The conditional probability of a light verb appearing given the complex predicate noun and n intervening phrases between the noun and the light verb.

n intervening phrases	probability of verb
0	$3826/4003 = 0.95$
1	$131/133 = 0.98$
2	$28/31 = 0.90$
3	$5/5 = 1$
4	$2/2 = 1$
6	$1/1 = 1$

Table 2. The conditional probability of a light verb appearing given the complex predicate noun and n intervening words between the noun and the light verb.

n intervening words	probability of verb
0	3826/4003 = 0.96
1	104/104 = 1
2	36/39 = 0.92
3	4/5 = 0.8
4	9/10 = 0.9
5	3/3 = 1
6	3/3 = 1
7	1/1 = 1
8	2/2 = 1
10	2/2 = 1
12	1/1 = 1
13	1/1 = 1
14	1/1 = 1

strong and weak predictability conditions. However, it is possible to subsume the difference between strong and weak predictability under the surprisal account by reframing the conditional probabilities in terms of the exact identity of the verb. In this case, the expectation account would predict faster reading times in the strong predictability conditions compared to the weak predictability conditions, regardless of distance.

To summarize, regarding the distance manipulation, the expectation account predicts either no effect or a facilitation at the verb as a function of distance; and regarding the predictability manipulation, the expectation account (appropriately formulated to include the conditional probability of the exact lexical item predicted) would predict a main effect of predictability.

4.3 RESULTS

4.3.1 Comprehension accuracy Participants answered correctly on average 92.73 percent of all comprehension questions (excluding fillers). Accuracy was 91, 94, 95 and 91 percent respectively for the four conditions in (1). A generalized linear mixed model of the binary responses showed an interaction (coef=-0.25, SE=0.10, z= -2.37) between predictability and distance. A nested contrast suggests that this interaction is driven by the weak predictability condition, such that response accuracy is lower in the long condition compared to the short condition.

4.3.2 Reading time Reading times (RTs) were analyzed at the verb; plots of the other regions are shown in the Supplementary materials. As shown in Table 3 and Figure 1, there was a main effect of distance, such that increasing distance led to longer reading times. There was also a main effect of predictability: the complex predicate conditions were read faster overall. A marginal interaction is also seen: stronger locality effects are seen in the control condition than in the complex predicate condition. A nested analysis shows that the distance effect was driven by the control (weak predictability) condition (strong predictability: coef.=0.02, SE=0.015, t=1.53; weak predictability: coef.=0.06, SE=0.015, t=3.87).

Table 3. Coefficients, standard errors, and t-values for the main effects and interactions in Experiment 1.

Comparison	Coefficient	SE	t-value
(Intercept)	6.24	0.04	151.99
Distance	0.04	0.01	**3.88**
Predictability	-0.03	0.01	**-2.94**
Distance x Predictability	0.02	0.01	1.70

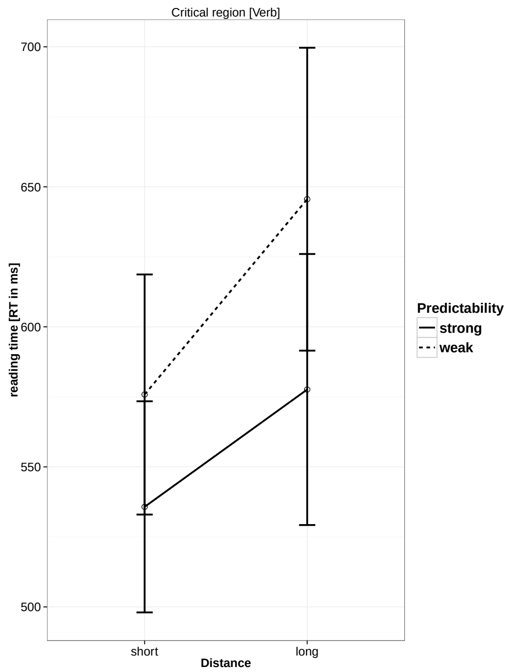

Figure 1. Reading times at the critical verb in Experiment 1.

4.4 DISCUSSION

Experiment 1 found a main effect of predictability such that the strong predictability conditions were read faster than the weak predictability conditions, and a main effect of distance, such that the short conditions were read faster than the long conditions. A nested contrast showed that this effect of distance was driven by the weak predictability conditions, i.e., reading time at the verb in condition c was faster than the reading time in condition d. A marginal interaction suggests that the locality effect may be somewhat stronger in the weak predictability condition. The marginal interaction seems to provide only weak support for the idea that strong predictability can at least attenuate locality effects (**Husain et al.**, 2014). The results are partly consistent with memory-based accounts, which correctly predict a slowdown at the verb in the long conditions, i.e., a main effect of distance. However, as the nested comparison shows, the main effect of distance is driven only by the weak predictability (non-complex predicate) conditions. Memory-based theories would be unable to explain this because they predict a slowdown in long conditions irrespective of predictability strength. The expectation account's prediction regarding distance, that increasing the argument-verb distance would either have no effect or result in a facilitation, was not validated; however, the main effect of predictability is consistent with a version of the expectation account that uses the conditional probability of the exact lexical item (verb) appearing given the preceding context.

Our original motivation for this study was to attempt a replication of the **Husain et al.** (2014) findings. The results are not entirely inconsistent with those of **Husain et al.** (2014), but they are also not a strong validation of the expectation-memory cost tradeoff posited in that paper. As in the Husain et al. study, we see a main effect of predictability driven by the complex predicate condition. This effect could be explained in terms of reduced retrieval cost at the verb due to its high expectation. An obvious confounding factor here is that the verbs in the strong vs weak predictability conditions are not identical; this prevents us from ruling out the possibility that low-level differences in the verbs might be responsible for the facilitation due to prediction strength.

We turn next to experiment 2, in which we manipulate the type of intervener. Here, in the long distance condition, instead of a relative clause and PP intervener, a long PP intervenes. The motivation was to increase distance without having different types of interveners in the short vs long conditions, as this might be a fairer comparison.

5 EXPERIMENT 2

5.1 METHOD

5.1.1 Participants Forty-three subjects, with the same criteria as in experiment 1, participated in this experiment in Tehran, Iran. This study was carried out in accordance with Helsinki Declaration, and consent forms were obtained from all the participants.

5.1.2 Materials The stimuli and fillers were the same as in experiment 1 except, for the long conditions (b and d) where the intervener was a longer prepositional phrase (PP) instead of the combination of a relative clause and a PP as in the previous experiment. The PP was lengthened using several different structures, all of which had one or more instance of the ezafe possessive marker (**Samvelian**, 2007):

1. N-ez N-ez N/pronoun/proper name
2. N-ez adj-ez N/pronoun/proper name
3. N-ez adj-ez N
4. N-ez N-ez adj
5. N adj-ez adj
6. superlative adj N N/pronoun/proper name
7. N-ez pronoun

One set of examples using the first type of PP shown above is as follows :

(5) a. Strong predictability, short distance (PP)

Ali a:rezouyee bara:ye man kard va. . .
Ali wish-INDEF for 1.S do-PST and. . .

'Ali made a wish for me and. . . '

b. Strong predictability, long distance (longer PP)
Ali a:rezouyee bara:ye doost-e xa:har-e man kard va. . .
Ali wish-INDEF for friend-EZ sister-EZ 1.S do-PST and. . .

'Ali made a wish for my sister's friend. . . '

c. Weak predictability, short distance (PP)
Ali shokola:ti bara:ye man xarid va. . .
Ali chocolate-INDEF for 1.S buy-PST and. . .

'Ali bought a chocolate for me and. . . '

d. weak predictability, long distance (longer PP)
Ali shokola:ti bara:ye doost-e xa:har-e man xarid va. . .
Ali chocolate-INDEF for friend-EZ sister-EZ 1.S buy-PST and. . .

'Ali bought a chocolate for my sister's friend and. . . '

More details about the PPs are provided in the Supplementary materials.

5.1.3 Procedure and Data Analysis The procedure and data analysis methodology was the same as experiment 1.

5.2 PREDICTIONS (EXPERIMENT 2)

In experiment 2, the distance manipulation involves lengthening the PP. There are two possible predictions of surprisal. One is that surprisal may predict no difference at the verb; this would be because the end of the PP raises a strong expectation for a verb, and this strong expectation for a verb would be the same in both the short and long PP conditions. Another alternative possible prediction of surprisal is that lengthening the PP could lead to a facilitation. This prediction could hold if increasing distance, counted in terms of the number of intervening words, generally increases the predictability of the upcoming verb; this is a possibility given the corpus counts in Table 2.

5.3 RESULTS

5.3.1 Comprehension Accuracy Participants answered 92.75 percent of all comprehension questions correctly on average (excluding fillers). The accuracies by condition were 96, 92, 94, and 89 percent respectively for the four conditions in (2). The generalized linear mixed models of the responses showed a main effect of distance (coef=-0.35, SE=0.10, z=-3.39) such that accuracies were lower in the long conditions. No effect of predictability strength, and no interaction between predictability strength and distance were found.

5.3.2 Reading Time As shown in Table 4 and Figure 2, the results showed a main effect of distance, with long distance conditions being read slower. There was also an effect of predictability, with the strong predictability condition being read faster than the weak predictability condition. No interaction was found between predictability and distance. A nested contrast showed that the distance effect is seen in both strong predictability (coef.=0.06, SE=0.02, t=3.63) and weak predictability (coef.=0.05, SE=0.02, t=2.62) conditions.

Table 4. Coefficients, standard errors, and t-values for the main effects and interactions in Experiment 2.

Comparison	Coefficient	SE	t-value
Intercept	6.27	0.04	147.86
Distance	0.06	0.01	**3.99**
Predictability	-0.02	0.01	**-2.28**
Distance x Predictability	-0.01	0.01	-0.67

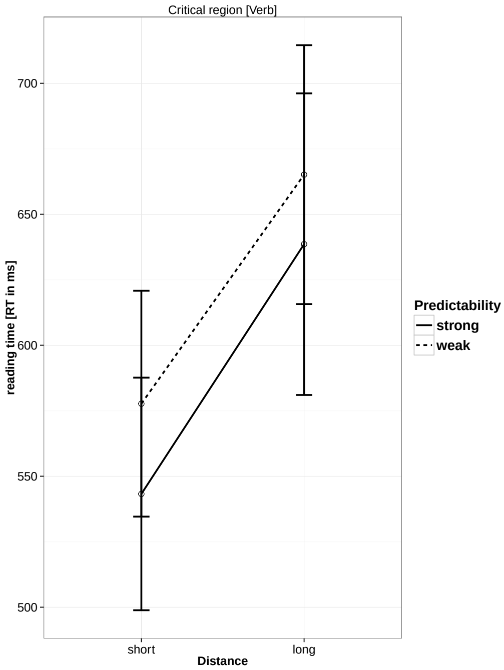

Figure 2. Reading times at the critical verb in experiment 2.

5.4 DISCUSSION

In this experiment, we replicated the locality effects found in Experiment 1, but we no longer see a weakening of the locality effect that was seen in Experiment 1 (a marginal interaction was found in Experiment 1). Nested contrasts showed that locality effects are equally strong in both the strong and weak predictability conditions. In Experiment 2, we also see an effect of predictability, with the strong predictable verb being read faster. Thus, regarding the distance manipulation, the working-memory account's prediction is validated, and the expectation-based account's prediction is not supported. The main effect of predictability does furnish evidence consistent with the expectation-based account.

A secondary analysis was conducted to compare the strength of the locality effect in the two experiments, and to determine whether the interaction between distance, predictability and experiment is significant.

The between-subject factor experiment was coded using sum coding: experiment 1 was coded -1, and experiment 2 was coded +1 (further details are available in the Supplementary materials). The results are shown in Table 5. An interaction between distance and experiment is seen: the locality effect was stronger in experiment 2. This confirms the suggestion that the locality effect is strengthened in experiment 2 compared to experiment 1.

Table 5. Comparison of experiments 1 and 2.

Comparison	Coef.	SE	t-value
Intercept	6.23	0.03	231.29
Distance	0.03	0.01	3.72
Predictability	-0.02	0.01	-2.25
Expt	0.04	0.03	1.67
Distance×Predictability	-0.00	0.01	-0.55
Distance× Expt	0.03	0.01	5.58
Predictability× Expt	-0.01	0.01	-1.27
Pred× Dist × Expt	0.01	0.01	1.66

In experiment 2, the intervener was a long, uninterrupted prepositional phrase whereas in experiment 1, the intervener consisted of a short RC followed by a PP. One can speculate about why experiment 2 shows equally strong effects in both predictability conditions: Processing a single long intervening phrase may be harder than processing two different phrases because it may be harder to chunk a single long phrase compared to two shorter phrases; this is predicted by the Sausage Machine proposal of **Frazier and Fodor** (1978). If this is correct, then the complexity of the intervener may indeed be a relevant factor in determining whether strong expectation can weaken locality effects. It is possible to test this claim by using an intervener that is much easier to process; an example would an adverb containing no noun phrases.

We were motivated by the recent replication crisis in psychology (**Open Science Collaboration**, 2015) to attempt to replicate our results using a different method. Furthermore, replications using eye-tracking would be very informative because it is possible that self-paced reading overburdens the working-memory system in an unnatural manner. If this is the case, one prediction would be that the eye-tracking data would not necessarily show locality effects. We describe these experiments next.

6 EXPERIMENT 3

6.1 METHOD

6.1.1 Participants Forty subjects, with the same criteria for inclusion as in the previous experiments, participated in the eye-tracking study in University of Potsdam, Germany.

6.1.2 Materials The experimental items were exactly the same as experiment 1 (self-paced reading), except that four items from experiment 1 were removed. The following four items were removed: item id 5, *sheka:yat kardan* (complain + to do), item id 9, *sahm bordan* (share + to win), item id 26, *pishraft kardan* (progress + to do), and item id 32, *hes kardan* (feel + to do). The reason was that the results of the sentence-completion studies suggested that these light verbs had lower predictability than the other light verbs in the stimuli. It could be that this lower predictability is due to the existence of some other alternative light verbs with which the nominal part can combine to make other possible complex predicates. The last two CPs also had a lower acceptability rating (item 26 had 4.7, and item 32 had 3.5).

As a consequence, in our eye-tracking study, we had thirty-two experimental items and sixty-four fillers. All items, including fillers are available in the Supplementary materials.[4]

6.1.3 Procedure An eye-tracking study was prepared using Experiment-Builder software, and participants' eye-movements were recorded using an EyeLink 1000 tracker, with a connection to a PC. Before the experiment started, the participants were instructed to read the sentences silently at a normal pace and had a practice block consisting of five sentences. After answering the comprehension questions of the practice block, they were provided with feedback indicating whether or not the answer was correct. A 21-inch monitor was placed 60 centimeters from the participants' eyes. In order to reduce head movements, the participants were asked to use the chin-rest. They viewed the sentences with both eyes, but only the right eye was recorded. The items were presented in one line and in 18 points Persian Arial font (from right to left). First, they had to fixate on a dot at the right edge of the screen so that the sentence appeared. After they finished reading, they had to fixate on the dot in the bottom left corner of the screen; once they fixated on the dot, the comprehension question was presented. Unlike the practice items, they were not provided with any feedback. Calibration was performed at the beginning of the experiment, after their 5-minute break (which occurred after they had read half the items), and whenever it was necessary.

6.1.4 Data analysis Raw gaze duration data was obtained using the Data Viewer software.[5] This data was then processed to get different eye-tracking measures using em2 (**Logačev and Vasishth**, 2014). As discussed earlier, linear mixed models were used for the analysis. All analyses were carried out using log-transformed data. For each reading time measure, zero ms reading times were removed.

6.2 RESULTS

6.2.1 Comprehension accuracy On average, participants correctly answered 91.05 percent of the target comprehension questions. Also, per condition, they had 91 percent response accuracy for condition a, 91 percent for condition b, 95 percent for condition c, and 89 percent for condition d. The generalized linear mixed models of the responses showed a main effect of distance (coef=-0.26, SE= 0.11, z= -2.39, p=0.016). An interaction was also found between predictability and distance (coef=-0.26, SE= 0.11, z=-2.38, p=0.016), and the nested analysis suggested that the interaction derives from the weak predictability conditions showing lower accuracies in the long vs short distance conditions (coef=-0.52, SE=0.16, z=-3.14, p=0.001).

6.2.2 Reading time The critical region was the verb, as in Experiments 1 and 2. The same sum contrast coding was used as in experiments 1 and 2; in addition, nested contrast coding was used to investigate the effect of distance within the two predictability conditions. We present results for first-pass reading time, regression path duration, and total reading time.

The effect of predictability, seen in Experiments 1 and 2, is also present in first-pass reading time and total reading time; the strong-predictability conditions had shorter reading times. Also, as in Experiments 1 and 2, there was an effect of distance; the long-distance conditions has longer reading times. Table 7 shows the details of the analyses. A nested contrast showed that in first-pass reading time the distance effect was present in both the strong- and weak-predictability conditions (strong predictability: coef.=0.044, SE=0.02, t=2.19; weak predictability: coef.=0.06, SE=0.02, t=2.57). Regression path duration did not show any distance effects with the two predictability conditions (strong predictability: coef.=0.03, SE=0.03, t=1.23; weak predictability: coef.=0.04, SE=0.03, t=1.49). The nested contrast in

[4] We also reanalyzed Experiments 1 and 2 after removing these four items; this did not change the results reported above for the experiments.

[5] http://www.sr-research.com/dv.html

total reading time showed no effect of distance in the strong-predictability condition (coef.=0.03, SE=0.03, t=1.27), but a distance effect was seen in the weak predictability condition (coef.=0.07, SE=0.026, t=2.52).

Table 6. Coefficients, standard errors, and t-values for the main effects and interactions in Experiment 3.

ET 1 measures	Comparison	Coef	SE	t value
Log FPRT	Intercept	5.62	0.03	175.88
	Distance	0.05	0.02	**2.67**
	Predictability	-0.053	0.02	**-3.08**
	Distance× Predictability	0.01	0.01	0.79
Log RPD	Intercept	5.73	0.04	128.98
	Distance	0.04	0.02	1.47
	Predictability	-0.08	0.02	**-3.21**
	Distance× Predictability	0.004	0.02	0.26
Log TRT	Intercept	5.77	0.05	124.04
	Distance	0.05	0.02	**2.17**
	Predictability	-0.10	0.02	**-4.17**
	Distance× Predictability	0.02	0.02	0.99

6.3 DISCUSSION

In the eye-tracking Experiment 3, we replicated the locality effects found in the Experiment 1 in first-pass reading time and total reading time. Nested contrasts showed that the locality effect tends to appear in weak-predictability conditions, which is similar to the result in Experiment 1. In first-pass reading time, the locality effect appeared in both the strong and weak-predictability conditions, but the magnitude of the effect was stronger in the weak-predictability condition. A main effect of predictability was found in all three dependent measures, replicating the effect in Experiment 1.

Since we failed to find any interaction between predictability and distance, we cannot conclude, as **Husain et al.** (2014) did, that expectation effects can cancel locality effects. The locality effects are consistent with the working memory accounts (**Gibson**, 2000; **Lewis and Vasishth**, 2005) and inconsistent with the distance-based predictions of the expectation account. As in the SPR experiments, we have evidence consistent with a version of the expectation account that predicts that strong predictability conditions will be read faster than the weak predictability conditions.

In the strong-predictability conditions, the somewhat weaker locality effect seen in first-pass reading time, and the absence of the effect in total reading time could be taken to be weakly consistent with the claims in **Husain et al.** (2014), but without an interaction between distance and predictability, these patterns are not really convincing.

In sum, the main result in experiment 3 is that we have replicated the locality effect and the facilitation due to strong predictability. There is some weak evidence that the locality effect may be reduced in the strong-predictability condition; but the absence of an interaction does not support the claim in **Husain et al.** (2014), that strong expectations cancel locality effects; the most we can say from the eye-tracking data is that strong expectations may weaken locality effects.

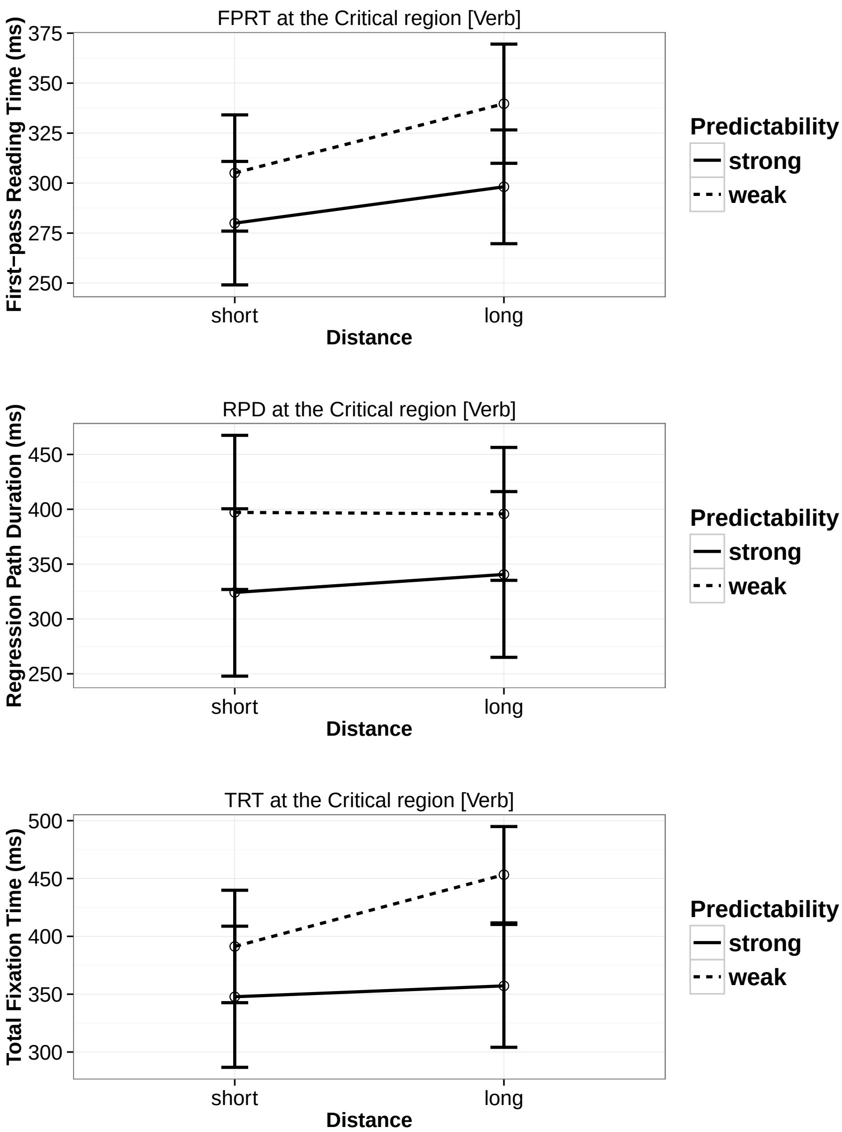

Figure 3. First-pass reading time, regression path duration, and total reading time in Experiment 3 at the critical verb.

7 EXPERIMENT 4

7.1 METHOD

7.1.1 Participants Forty participants, with the same criteria as in the previous experiments, participated in the eye-tracking study in Golm campus, University of Potsdam, Germany.

7.1.2 Materials The experimental items were exactly the same as experiment 2 (self-paced reading), but with 32 items (see the explanation for Experiment 3 regarding the four items that were removed). The experimental items were complemented with 64 filler sentences with varying syntactic structures (see Supplementary materials).

7.1.3 Procedure and Data Analysis The procedure and data analysis were exactly the same as experiment 3 (eye-tracking).

7.2 RESULTS

7.2.1 Comprehension Accuracy On average, participants answered 90.05 percent of comprehension questions correctly. They had 94 percent response accuracy for condition a, 88 percent for condition b, 94 percent for condition c, and 86 percent for condition d. The generalized linear mixed models of the responses showed a main effect of prediction (coef=0.86, SE=0.40, z=2.13, p= 0.033). Also there was an interaction between predictability and distance (coef=-1.034, SE=0.50, z=-2.04, p=0.041), and the nested analysis shows that this interaction derives from lower accuracy in the long vs short distance conditions in the weak-predictability conditions (coef=-0.52, SE=0.19, z=-2.65, p=0.007).

7.2.2 Eye-tracking measures Unlike experiment 3, in the current experiment, we found effects of distance and predictability in all the three measures (see Table 7). In other words, in the three measures example, the long conditions (b and d) were read slower than the short conditions (a and c), and the weak predictability conditions (c and d) were read slower than the strong predictability conditions (a and b). None of the measures showed any interaction between predictability and distance.

Nested comparisons showed that in first-pass reading time, the locality effect was seen in the strong-predictability condition (coef.=0.05, SE=0.02, t=2.33), but there was a weaker tendency towards a locality effect in the low-predictability condition (coef.=0.06, SE=0.03, t=1.86). In regression-path duration, both strong- and weak-predictability conditions showed a locality effect (strong-predictability: coef.=0.086, SE=0.02, t= 3.57; low-predictability: coef.=0.07, SE=0.03, t=2.15). In total reading time, the strong-predictability condition showed a locality effect (coef.=0.10, SE=0.02, t=4.06), but in the low-predictability condition, only a tendency towards a locality effect was seen (coef.=0.06, SE=0.04, t=1.59).

Table 7. Coefficients, standard errors, and t-values for the main effects and interactions in Experiment 4.

ET measures	Comparison	Coef	SE	t value
Log FPRT	Intercept	5.67	0.04	147.60
	Distance	0.06	0.02	**2.46**
	Predictability	-0.11	0.02	**-5.10**
	Distance × Predictability	-0.01	0.02	-0.30
Log RPD	Intercept	5.79	0.05	125.45
	Distance	0.08	0.02	**3.63**
	Predictability	-0.11	0.02	**-4.76**
	Distance × Predictability	0.01	0.02	0.43
Log TRT	Intercept	5.80	0.05	122.09
	Distance	0.08	0.03	**3.06**
	Predictability	-0.15	0.02	**-6.69**
	Distance × Predictability	0.02	0.02	1.27

7.3 DISCUSSION

The eye-tracking Experiment 4 replicated the results of the self-paced reading study (Experiment 2): a main effect of distance and a main effect of predictability, with no evidence for an interaction. The effects

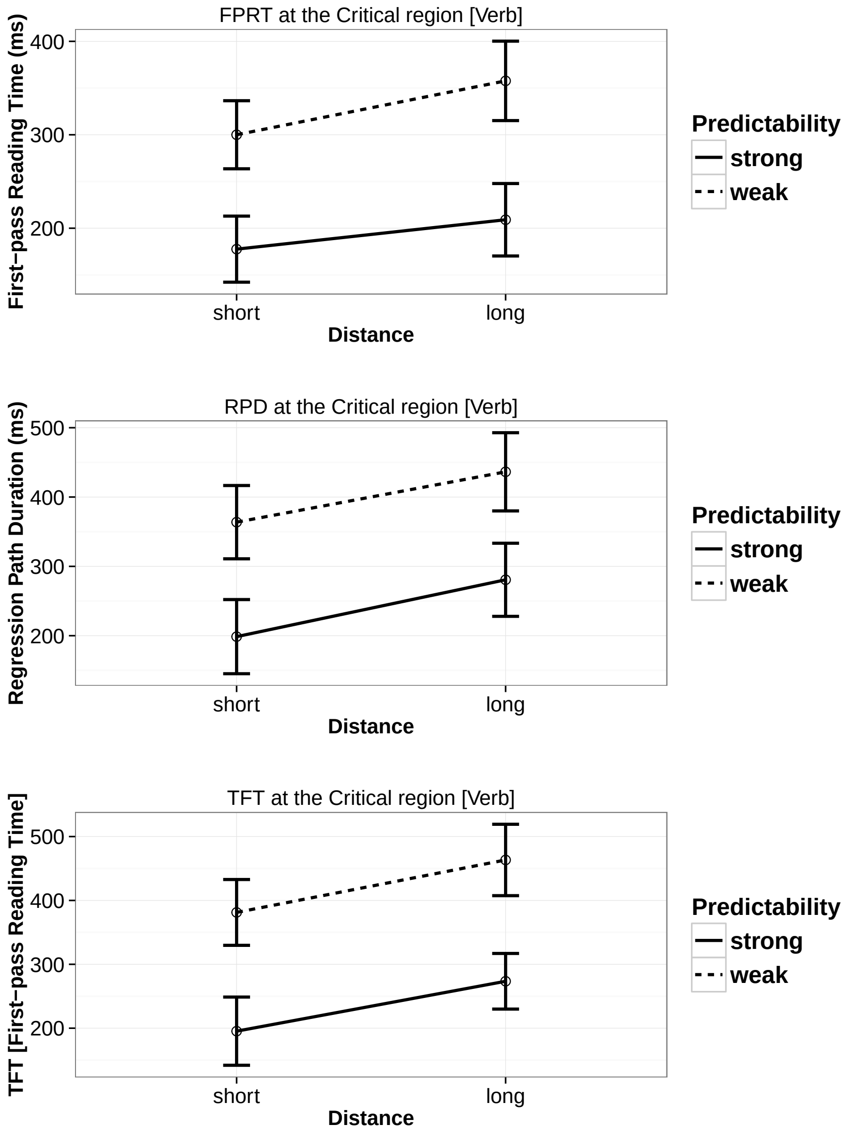

Figure 4. First-pass reading time, regression path duration, and total reading time in Experiment 4 at the critical verb.

in early (FPRT), regression (RPD), and late (TFT) measures showed the same patterns as in the first eye-tracking study. However, the locality effects were even stronger, in the same way that the second self-paced reading study showed stronger locality effects. Also, these effects are equally strong in both strong and weak predictability conditions, mirroring our finding in the second self-paced reading study. Evidence that the long PP in experiment 4 leads to stronger locality effects than in experiment 3, which has an RC+PP intervener, comes from a combined analysis of the two eye-tracking experiments, including experiment as a between subjects factor; this showed significant interactions between distance and experiment in regression path duration. In other words, the locality effect was stronger in experiment 4 than experiment 3 (coef. 0.02, SE 0.01, t= 2.14).

Overall, regarding the distance manipulation, the results are consistent with memory-based accounts, and inconsistent with the expectation account. The main effect of predictability is consistent with the expectation account, as discussed earlier. In Experiment 4, we don't see any evidence consistent with the **Husain et al.** (2014) proposal; if anything, the locality effect is *stronger* in the strong-predictability conditions.

8 GENERAL DISCUSSION

8.1 EVALUATING THE PREDICTIONS OF THE MEMORY-BASED ACCOUNTS AND THE EXPECTATION-BASED ACCOUNT

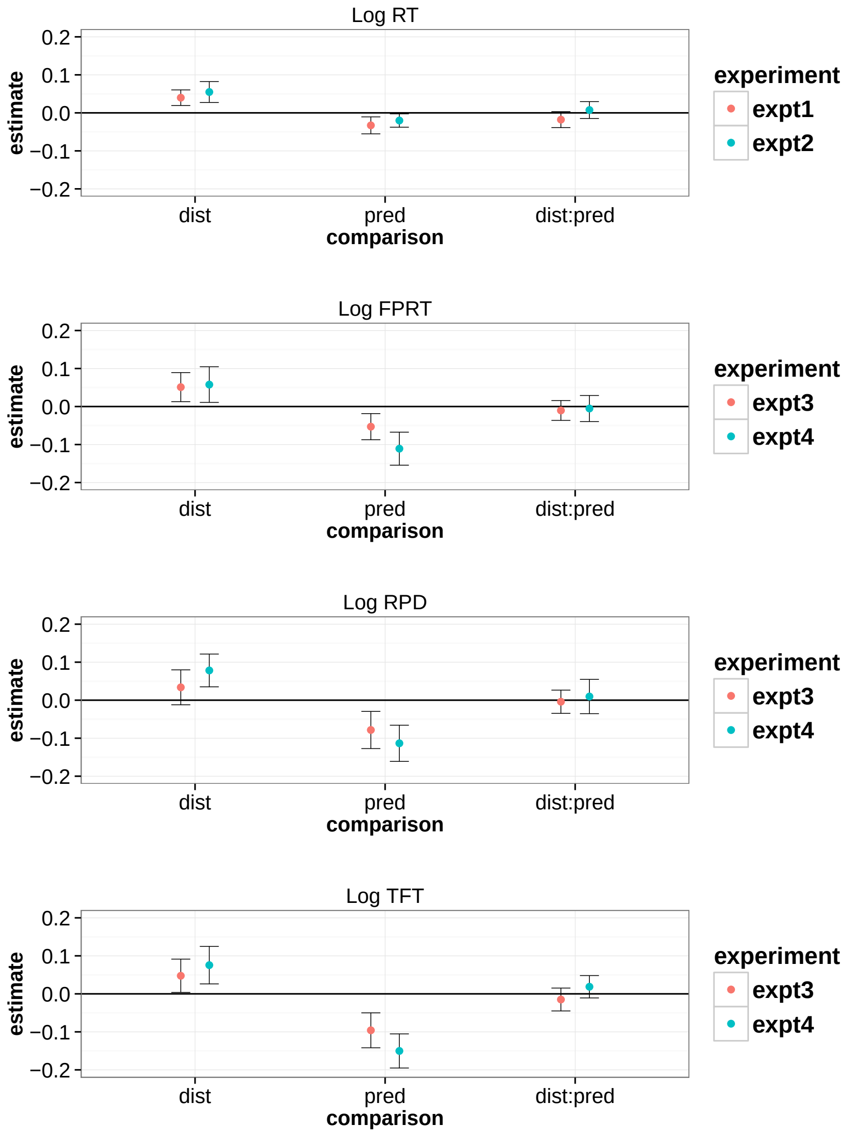

Figure 5. Summary of the magnitudes of effects (derived from the linear mixed models) across the four experiments. The error bars show 95% confidence intervals.

As summarized graphically in Figure 5, Our main finding from the four Persian studies is that the locality effect predicted by memory accounts is upheld, but there is no evidence for the expectation-based account's prediction of facilitation in longer distance conditions. We consistently see a main effect of predictability, which is consistent with expectation accounts. Finally, there is no compelling evidence in the Persian data that strong expectations cancel locality effects.

There is also suggestive evidence that the complexity of intervening material could strengthen the locality effect: when the intervener is an RC followed by a PP, we see a marginal interaction between distance and predictability, but when the intervener is a single long PP, we see no evidence for an interaction between distance and predictability strength, and we tend to see stronger effects. The two combined analyses of the SPR studies and of the eye-tracking studies show that the locality effect is stronger in the experiments with the long PP. Of course, a definitive test of such a difference would be a new design where we directly compare intervener types in a within-subjects design; such a follow-up study is currently being planned.

We consistently found a main effect of predictability in all four experiments: the strong predictability conditions were read faster at the verb than the weak predictability conditions. This is consistent with the expectation-based account. Since the verbs in the strong and weak predictability conditions are not identical, we cannot rule out the possibility that word frequency or other such low-level factors are responsible for these effects. However, it is plausible that the highly predictable verb is processed faster than the less predictable verb. Thus, the main effect of predictability can be seen as evidence for expectation-based accounts, operationalized in terms of the conditional probabilities of the appearance of the exact verb given the preceding context.

It is possible that we were unable to replicate the Husain et al findings because of the nature of the intervener used in the Persian studies. Unlike **Husain et al.** (2014) where the long distance condition had extra adverbials compared to the short condition, in Experiment 1 we have a more complex intervener, a relative clause. Another reason for finding the effects which are different from the study by **Husain et al.** (2014) could be that in Persian, separating the nominal part of the CP from the light verb occurs relatively rarely, compared to Hindi. There is some support for this in corpus data. Based on the Hindi dependency treebank (**Bhatt et al.**, 2009), the average distance, counted as the number of intervening phrases, between an object and its (heavy) verb in Hindi is 0.82 (with minimum 0 and maximum 15, and first and third quantiles 0 and 1), and the average distance between a noun and light verb is .07 (minimum 0 and maximum 18, with first and third quantiles 0 and 0). In the Persian dependency treebank (**Seraji**, 2015), the average distance between an object and (heavy) verb is 2.48 (with minimum 0 and maximum 9, and first and third quantiles 1 and 3), while the average distance between a noun and light verb is 0.05 (with minimum 0, and maximum 6, and first and third quantiles 0 and 0). Thus, the adjacency of CPs in Persian is strongly preferred (maximum 6 vs Hindi's maximum 18), although as validated in the acceptability rating norming study, this separability is acceptable and not considered ungrammatical.[6]

8.2 AN ALTERNATIVE EXPLANATION OF LOCALITY EFFECTS IN TERMS OF ENTROPY

Could there be an alternative explanation for the locality effect seen in the four experiments, one that does not invoke greater memory cost in the long-distance conditions? One possibility is that entropy (uncertainty) increases with increasing distance. Entropy is an information-theoretic measure that essentially represents how uncertain we are of the outcome (**Shannon**, 2001). In the present case, this would translate to our uncertainty about the upcoming verb. If there are n possible ways to continue a sentence, and each of the possible ways has probability p_i, where $i = 1, \ldots, n$, then entropy is defined

[6] These intervening phrases have been computed using dependency treebanks. Consequently, phrasal boundaries are approximations. Also, because of annotation differences between the two treebanks, phrase boundary criteria sometimes differ for the two languages. The phrasal counts lead to the same conclusions regardless of whether one counts intervening phrases or words.

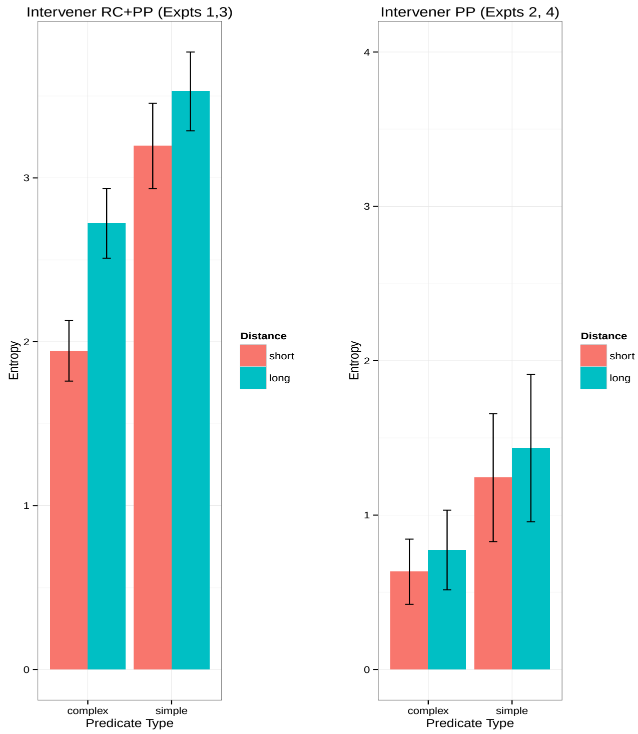

Figure 6. The estimated entropy (with 95% confidence intervals), computed using the sentence completion data, for the two experiment designs.

568 (**Shannon**, 2001) as $-\sum_i p_i \times \log_2(p_i)$. The entropy associated with the upcoming verb can be calculated
569 using our sentence completion data.[7]

570 *8.2.1 Evaluating the effect of entropy* In order to evaluate whether entropy could explain the locality
571 data, we computed entropy for each item in each condition for both experiments. The estimated entropies
572 for each condition in the two experiment designs are shown in Figure 6. It is important to note here that
573 entropy for each condition in Figure 6 is based only on nine data points per condition (we only have
574 $9 \times 4 = 36$ items); for different items, there is substantial variability in the entropy patterns by condition.
575 Nevertheless, in the figure we can see that in the items used for Experiments 1 and 3, the entropy is higher
576 in the long-distance conditions. The effect of entropy is less clear for the items used in Experiments 2
577 and 4, because of the relatively wider confidence intervals. Clearly uncertainty is higher in the RC+PP
578 conditions. A closer look at the high predictability conditions shows that the entropy difference between
579 the long and short distance conditions is larger in the RC+PP intervener items than the entropy difference
580 in the long PP intervener items (it is larger by 0.28, with 95% confidence intervals -0.01 and 0.57). This
581 suggests that the intervening RC may be responsible for creating a greater degree of uncertainty regarding
582 the upcoming verb.

583 In order to investigate whether entropy affects reading times at the verb, we fit linear mixed models with
584 predicate type, distance, as sum-coded factors, and entropy (centered) as a continuous factor; all higher
585 order interactions were also included. Varying slopes for entropy were always included with varying
586 intercepts for item, and other varying slopes were also included when these were justified (**Bates et al.**,

[7] See **Linzen and Jaeger** (2015) for a recent empirical investigation of entropy in sentence comprehension using corpus data instead of sentence completion data. Linzen and Jaeger calculated entropy in several ways, and also evaluated another metric called entropy reduction (ER); however, we cannot evaluate ER here because that would require knowing the entropy for the word preceding the verb.

2015); no attempt was made to estimate intercept-slope correlations. The dependent variable was log reading time at the critical verb. In Experiment 1, in addition to the effects of predictability and distance, we find an effect of entropy (coef.=0.05, SE=0.02, t=2.8), and an interaction between distance and entropy (coef.=0.04, SE= 0.02, t= 2.3), such that long distance conditions lead to a greater effect of entropy. None of the other experiments showed any effects of entropy. Thus, although the evidence in favor of entropy is not overwhelming, a potentially important finding here is that entropy could explain locality effects, at least in our experiment 1. To our knowledge, this is the first demonstration that locality effects may arise due to factors other than memory costs.

8.2.2 Does predictability of an upcoming verb increase with distance? We showed above that uncertainty about the upcoming verb may explain locality, at least in experiment 1. One important question that arises is, especially in the strong predictability conditions, is increasing distance nevertheless sharpening the expectation for the verb, as suggested by **Konieczny** (2000)? In order to address this question, we fit a Bayesian generalized linear mixed model (GLMM) with a logistic link function that investigated the change in probability mass for the target verb as a function of distance in the strong predictability conditions. We used a Bayesian model because the "maximal" frequentist model would not converge and we wanted to obtain the most conservative result possible from the data (**Barr et al.**, 2013; **Matuschek et al.**, 2015). The predictor (short vs long condition) was coded using sum contrasts, with the long condition coded as 1 and the short condition as -1; the dependent variable was binary and represented whether a target verb was produced by the participant for a particular item-condition combination or not. Participants and items were specified as crossed random factors, and a full variance-covariance matrix was fit for both random effects. The priors for the intercept and slope were the Student's t-distribution with 2 degrees of freedom, allowing a range of approximately -10 to 10 on the log odds scale, with 0 the most likely value. The prior on the variance-covariance matrices was defined via the LKJ prior (**Stan Development Team**, 2013, 2014) on the correlation matrix; see **Sorensen et al.** (2015) for a tutorial intended for psycholinguists and cognitive scientists. The model was fit using the `stan_lmer` function from the `rstanrarm` package (**Gabry and Goodrich**, 2016). The results showed that in the long distance condition, the probability of producing the target verb fell: On the log-odds scale, the mean and 95% uncertainty interval were $-0.29, [-0.78, 0.17]$, and the posterior probability of the reduction being less than 0 was 0.89. The odds ratio of producing a target verb was 0.75, with 95% uncertainty interval $[0.46, 1.2]$; this means that in the short condition, participants are more likely to produce the target verb, but since the uncertainty interval for the odds ratio includes 1, the reduction in probability of target verb production is possibly unchanged in the short vs long distance conditions. The posterior distribution is shown in Figure 7 in order to give a visual sense of the dispersion in the probability. Thus, our sentence completion data for experiment 1's strong predictability condition shows that increasing distance tends to reduce the proportion of target verbs produced, although the evidence for this reduction is rather weak. Our data from Persian therefore seem to go against the suggestion by **Konieczny** (2000) that increasing distance might lead to narrowing down the prediction to the target verb.

It may be helpful to also see the by-item adjustments to the mean effect of distance (in the frequentist framework, these would be called Best Linear Unbiased Predictors or BLUPs). This is shown in Figure 8. As this figure shows, across the items the change in log odds of producing the target verb is very variable, with wide 95% uncertainty intervals. In fact, in some items there is, at least numerically, an increase in log odds in the long conditions, corresponding to the sharpening of expectations, as suggested by **Konieczny** (2000). It is possible that a much larger study (with more items as well as participants) using sentence completion to compute target-verb production can reveal an a more nuanced picture of sharpened expectations.

We were concerned that the reduction in correct completions in the long-distance condition could be an artefact of a peculiarity of the Persian language. Participants in the sentence completion study tended to produce synonymous noun-verb sequences as a completion (in place of the single target verb); such noun-verb sequences are grammatical continuations and represent a different (more formal) register. In the short condition the proportion of such completions was 0.05, but in the long condition this proportion

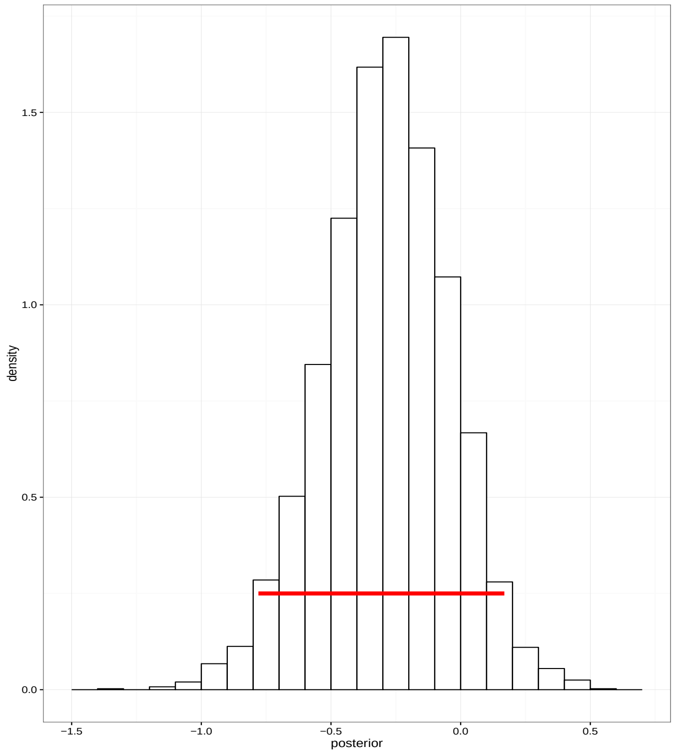

Figure 7. The posterior distribution computed using a Bayesian generalized linear mixed model representing the reduction in the log odds of producing the target verb in the long vs short distance condition in the first sentence completion study (strong predictability conditions only). The red horizontal line marks a 95% uncertainty interval, the range over which we can be 95% certain that the true parameter lies given the data.

was higher, at 0.14. In our original analysis above, we had considered a completion correct if the exact verb was produced. If we relax this criterion and recode the sentence completion data so that a synonym is considered a correct completion, then we still find weak evidence for a reduction in the probability of producing a correct continuation in the long vs short distance condition. The log odds are $-0.21[-0.71, 0.23]$, with the probability of the log odds being less than 0 being 0.82. Thus, the reduction in correct completions does not seem to be due to this property of Persian, that complex predicate verbs have synonyms. Nevertheless, caution is necessary in interpreting this weakening of the expectation for the target verb with increasing distance; it would be very informative to carry out sentence completion studies for experiments such as those of**Konieczny** (2000); **Grodner and Gibson** (2005); **Bartek et al.** (2011); **Vasishth and Lewis** (2006); **Vasishth and Drenhaus** (2011); **Levy and Keller** (2013) in order to establish whether increasing distance can weaken expectation cross-linguistically.

In summary, the RC intervener may be the cause for the greater entropy in the long-distance conditions, but the reason for the RC causing an increase in entropy is not clear to us and needs further study.

Apart from revisit the existing locality effects in English, German, and Hindi from the perspective of entropy, a further possibility worth investigating is whether entropy reduction (**Hale**, 2006) can explain the locality effects cross-linguistically. In our Persian experiments, it is possible that the entropy at the word preceding the verb is higher than the entropy at the verb, and it is possible that the reduction in entropy is larger in the long-distance condition. Unfortunately, we have no way to test this in the present design, but future studies could compute entropy reduction empirically in the same way that we computed entropy using sentence completion data. Thus, in principle it is possible that entropy reduction could explain locality effects as well. A related issue that would then arise is whether entropy or entropy reduction furnishes a better explanation for locality effects. This would be a very productive line of research.

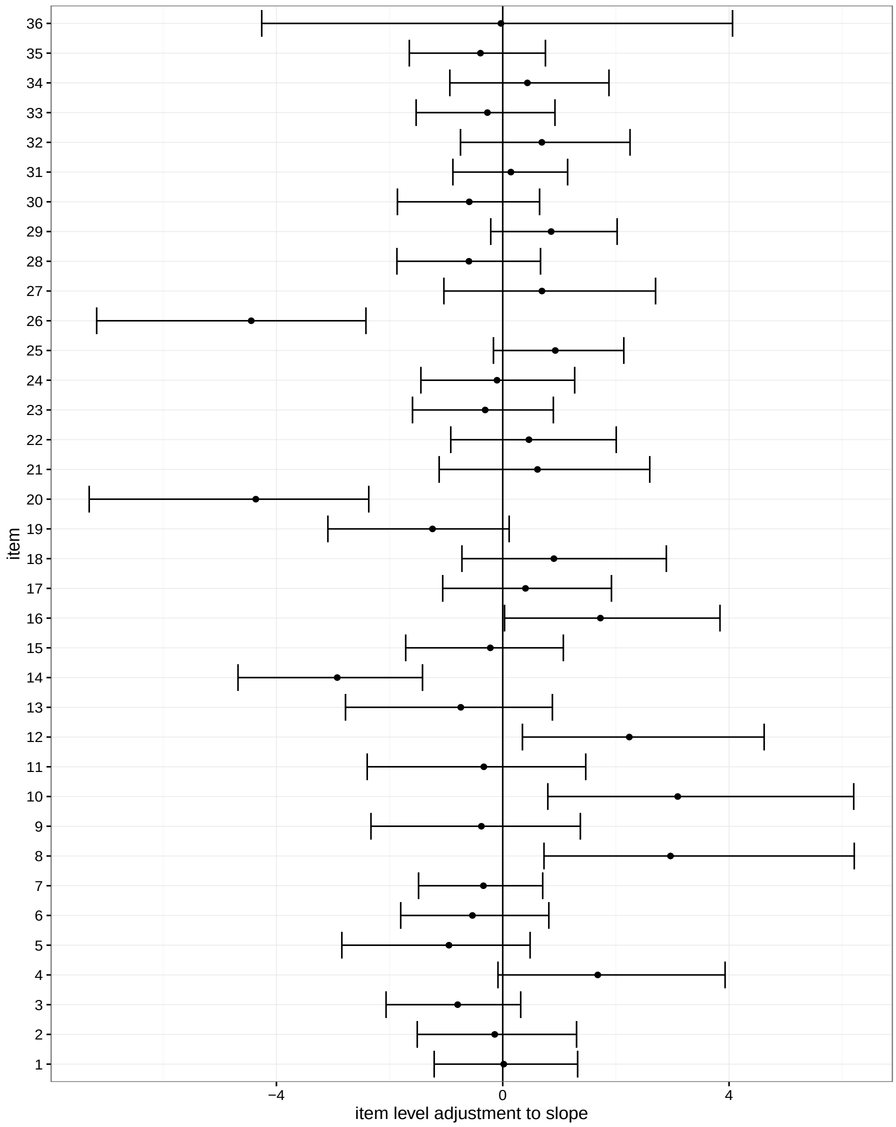

Figure 8. By-item distribution of random effects adjustments to the effect of distance in the experiment 1 sentence completion study (strong predictability conditions only). The error bars show 95% credible intervals.

8.3 CONCLUDING REMARKS

In conclusion, as regards the distance manipulation, the evidence from Persian is in favor of working-memory accounts, although entropy is also a candidate explanation. There is not much evidence from Persian that strong-predictability conditions cancel locality effects, as Husain and colleagues had suggested. Interestingly, there is no evidence in these experiments for the prediction of the expectation account regarding the distance manipulation, that increasing argument-verb distance facilitates processing due to increasing conditional probabilities of the upcoming verb. The suggestion in (**Levy et al.**, 2013) that "the verb-medial languages tend to exhibit the general patterns predicted by memory-based theories, whereas verb-final languages tend to exhibit the general patterns predicted by expectation-based theories" seems to be difficult to maintain (also see **Husain et al.** (2015), for locality effects in Hindi). One implication of our findings from Persian is that locality and expectation effects observed across studies seem to be highly conditional on the language and syntactic construction being considered—broad cross-linguistic generalizations may be difficult to make.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SUPPLEMENTARY MATERIALS

All items, data, and R code associated with this paper are available from http://www.ling.uni-potsdam.de/∼vasishth/code/SafaviEtAl2016DataCode.zip.

ACKNOWLEDGMENTS

Thanks to Prof. Dr. Shahla Raghibdoust in Allameh Tabataba'i University who helped the first author to recruit the participants in Iran. We are grateful to Carla Kessler in University of Potsdam for her help in designing the eye-tracking study using Experiment-Builder software. Many thanks to Lena Jaeger who did a precise sanity check for the eye-tracking results. Finally, We would like to thank the audience in the 28th CUNY conference of human sentence processing in University of Southern California for their insightful feedback. Finally, we thank an anonymous reviewer for his/her comments, which led us to investigate entropy as a possible explanation.

Funding: This work was supported by Erasmus Mundus Joint Doctoral program of European Union, International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), and the University of Potsdam, Germany.

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013), Random effects structure for confirmatory hypothesis testing: Keep it maximal, *Journal of memory and language*, 68, 3, 255–278

Bartek, B., Lewis, R. L., Vasishth, S., and Smith, M. R. (2011), In search of on-line locality effects in sentence comprehension., *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 5, 1178

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015), Parsimonious mixed models, arXiv e-print

Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., and Xia, F. (2009), A multi-representational and multi-layered treebank for Hindi/Urdu, in Proceedings of the Third Linguistic Annotation Workshop (Association for Computational Linguistics), 186–189

Caplan, D. and Waters, G. (2013), Memory mechanisms supporting syntactic comprehension, *Psychonomic bulletin & review*, 20, 2, 243–268

Chomsky, N. (1965), Aspects of the Theory of Syntax (MIT, Cambridge, MA)

Dabir-Moghaddam, M. (1997), Compound verbs in persian, *Studies in the Linguistic Sciences*, 27, 2, 25–59

Frazier, L. and Fodor, J. D. (1978), The sausage machine: A new two-stage parsing model, *Cognition*, 6, 4, 291–325

Gabry, J. and Goodrich, B. (2016), rstanarm: Bayesian Applied Regression Modeling via Stan, r package version 2.9.0-1

Gibson, E. (1998), Linguistic complexity: Locality of syntactic dependencies, *Cognition*, 68, 1, 1–76

Gibson, E. (2000), The dependency locality theory: A distance-based theory of linguistic complexity, *Image, language, brain*, 95–126

Grodner, D. and Gibson, E. (2005), Consequences of the serial nature of linguistic input for sentenial complexity, *Cognitive Science*, 29, 2, 261–290

Hale, J. (2001), A probabilistic earley parser as a psycholinguistic model, in Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (Association for Computational Linguistics), 1–8

Hale, J. T. (2006), Uncertainty about the rest of the sentence, *Cognitive Science*, 30, 4

Hsiao, F. P.-F. and Gibson, E. (2003), Processing relative clauses in Chinese, *Cognition*, 90, 3–27

Husain, S., Vasishth, S., and Srinivasan, N. (2014), Strong Expectations Cancel Locality Effects: Evidence From Hindi, *PloS one*, 9, 7, e100986

Husain, S., Vasishth, S., and Srinivasan, N. (2015), Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus, *Journal of Eye Movement Research*, 8(2), 1–12

Jaeger, F. T., Fedorenko, E., Hofmeister, P., and Gibson, E. (2008), Expectation-based syntactic processing: Antilocality outside of head-final languages, in CUNY Sentence Processing Conference (North Carolina)

Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., and Vasishth, S. (2015), The subject-relative advantage in Chinese: Evidence for expectation-based processing, *Journal of Memory and Language*, 79–80, 97–120, doi:10.1016/j.jml.2014.10.005

Jurafsky, D. (1996), A probabilistic model of lexical and syntactic access and disambiguation, *Cognition*, 20, 137–194

Just, M. and Carpenter, P. (1992), A capacity theory of comprehension: Individual differences in working memory, *Psychological Review*, 99, 1, 122–149

Karimi-Doostan, G. (1997), Light verb constructions in Persian, Ph.D. thesis, University of Essex

Karimi-Doostan, G. (2005), Light verbs and structural case, *Lingua*, 115, 12, 1737–1756

Karimi-Doostan, G. (2011), Separability of light verb constructions in Persian, *Studia Linguistica*, 65, 1, 70–95

Konieczny, L. (2000), Locality and parsing complexity, *Journal of Psycholinguistic Research*, 29, 6, 627–645

Kutas, M. and Hillyard, S. (1984), Brain potentials during reading reflect word expectancy and semantic association, *Nature*, 307, 161 – 163

Levy, R. (2008), Expectation-based syntactic comprehension, *Cognition*, 106, 3, 1126–1177

Levy, R., Fedorenko, E., and Gibson, E. (2013), The syntactic complexity of Russian relative clauses, *Journal of memory and language*, 69, 4, 461–495

Levy, R. P. and Keller, F. (2013), Expectation and locality effects in German verb-final structures, *Journal of memory and language*, 68, 2, 199–222

Lewis, R. L. and Vasishth, S. (2005), An activation-based model of sentence processing as skilled memory retrieval, *Cognitive science*, 29, 3, 375–419

Linzen, T. and Jaeger, T. F. (2015), Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions, *Cognitive science*

Logačev, P. and Vasishth, S. (2014), The em2 package for computing eyetracking measures, University of Potsdam, Potsdam, Germany

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2015), Balancing type I error and power in linear mixed models, *arXiv preprint arXiv:1511.01864*

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., and Kliegl, R. (2015), Working memory differences in long-distance dependency resolution, *Frontiers in Psychology*, 6

Open Science Collaboration (2015), Estimating the reproducibility of psychological science, *Science*, 349, 6251, aac4716

Pinheiro, J. and Bates, D. (2000), Mixed-Effects Models in S and S-PLUS (Springer)

R Development Core Team (2013), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria

Samvelian, P. (2001), The Syntax of Iranian Languages, Canadian Journal of Linguistics. Vol. 46: 1-2,(March-June 2001), Toronto, University of Toronto Press., *Abstracta Iranica. Revue bibliographique pour le domaine irano-aryen*, , Volume 26

Samvelian, P. (2007), A (phrasal) affix analysis of the Persian Ezafe, *Journal of Linguistics*, 43, 3, 605–645

Seraji, M. (2015), Morphosyntactic Corpora and Tools for Persian, Ph.D. thesis, Uppsala University

Shannon, C. E. (2001), A mathematical theory of communication, *ACM SIGMOBILE Mobile Computing and Communications Review*, 5, 1, 3–55

Sorensen, T., Hohenstein, S., and Vasishth, S. (2015), Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists, arXiv e-print

Stan Development Team (2013), Stan: A C++ library for probability and sampling, version 2.1

Stan Development Team (2014), Stan Modeling Language Users Guide and Reference Manual, Version 2.2

Staub, A. (2010), Eye movements and processing difficulty in object relative clauses, *Cognition*, 116, 1, 71–86

Vahedi-Langrudi, M.-M. (1996), The syntax, semantics and argument structure of complex predicates in modern Farsi. (University of Ottawa (Canada).)

Vasishth, S. (2003), Working memory in sentence comprehension: Processing Hindi center embeddings (Routledge)

Vasishth, S. and Drenhaus, H. (2011), Locality in German, *Dialogue & Discourse*, 2, 1, 59–82

Vasishth, S. and Lewis, R. L. (2006), Argument-head distance and processing complexity: Explaining both locality and antilocality effects, *Language*, 767–794