1. Research question

Is the visual (VA) or sentence (SA) anticipation different from each other in terms of strength? The null hypothesis is not equality precisely, but a state where the difference between VA and SA is not statistically significant.
H1: VA != SA, H0: VA = SA

Note that we will be comparing only a particular type of sentence structure and limit ourselves to simple visual scenes (single objects). This may lead to generalizations to VA or SA overall but will be unfounded to real-life scenarios: sharing the whole immediate physical scene.

As a corollary finding, we inspect which of the two anticipations is stronger and by what margin.

2. Method

Self-paced reading, a common online method to measure how much mental load is happening after reading individual words. Images are first shown (they depict objects of the sentence) and then the sentence. We then measure the sentence reading times. We use picture-naming to make sure the participant interacted with the image. This also forces the participants to load in the linguistics of the object. The specific flavour of SPR is that we only show the current word in the middle of the screen.

Our working linking hypothesis is that higher reading times are caused by lower anticipation and vice versa. This is supported by the fact that sentences with high-anticipatory contexts are easier to read and process.

The users could cheat the experiment by not paying attention and just speeding through the self-paced reading part. This would clearly be an outlier and could be identified and removed from the data. More sophisticated cheating methods could be employed by the cheating participants, such as automated key pressing or just generally not paying enough attention as we ask them to. For this reason, we add an easy comprehension question that appears at the end of every sentence, such as: What did he write? There is nothing stopping the participants from failing this question as well, and they will still get paid if they do so. But them succeeding in both the picture naming and sentence comprehension is proof enough for us to consider this a valid data point.
Other anti-cheat measures were also considered, such as asking them to rate the plausibility of the shown sentence, which, however, is harder to verify. A grammatical maze would also be an option, but this would perhaps obscure the reading times more.

Images are **not** kept visible along the sentence to remove image reading time variability and potential distractions. The possibility of the issue of the image being more distant from the object than the verb is discussed in Confounds.

There is a possibility of the wrap-up effect distorting our data. Since our objects of interest are at the end in a simple sentence like: *He ate an apple.*, then apple would have extra time added for whole sentence comprehension. Because of this, we extend the stimuli to not just be verbs, but also adjuncts which attach at the end of the sentence, such as *He ate an apple for dinner.*

3. Variables

Independent variables: V1: image is related to the focus of the sentence or not, V2: sentence contains anticipation or not. Specifically, we manipulate the conditional probability of the object given the verb that precedes it (sentence anticipation) and whether the image depicts the object or not.

Dependent variables: reading times for individual words and the sentence in total. Specifically, we are interested in the reading time for the sentence object, although the overall sentence reading time may be useful as well.

4. Predictions

The working hypothesis predicts that there should be different processing difficulty for the Visual Anticipation condition (and therefore, different reading times) than for the Sentence Anticipation condition. Null hypothesis predicts that there are no significant differences between VA and SA conditions.
Furthermore, we expect the SA to be stronger than VA.

5. Design

2x2 design (2 factors with 2 levels) -> 4 conditions:

1. Anticipatory verb + visualised object (AV + VO)
2. Anticipatory verb + non-visualised object (AV + NO)
3. Non-anticipatory verb + visualised object (NV +VO)
4. Non-anticipatory verb + non-visualised object (NV + NO)

Another approach to this experiment would be to just use the anticipatory verbs (stimuli) and either show passing objects (*He ate an apple for dinner.*) or objects that do not fit (*He ate the bus for dinner.*) The advantage of this is that there are less stimuli to create comprehension questions for and that the sentences have less variability. The main disadvantage and the reason why we chose non-anticipatory verbs explicitly is that in the latter case, the produced questions are too bizarre and not plausible enough.

Each participant is tested in each condition - within subject design. This prevents individual factors to affect the results (baseline for every participant can be measured) and requires less subjects. Four lists are created with the Latin square design to prevent carry-over effects. We split participants into four different lists.

For a real experiment, we would use a higher number of stimuli to make better use of the participant time. For the stimuli number selection three factors need to be taken into consideration: ability to make inferences on such a number of different stimuli, participant warmup and participant fatigue.
We estimate that after carefully reading the instructions, every stimulus takes on average 25s±9s. Because of this, lists of length 48 should still be manageable to be done under half an hour, together with instruction reading and warmup.

6. Stimuli

Stimulus	Example Objects	Comprehension Question
Anticipation Verbs		
eat .. during a recess	apple	What does he eat during a recess?
drive .. fast	car	How fast does he go?
drink .. in the evening	water	What does he drink in the evening?
bake .. out of boredom	cookies	Why does he bake the X?
cook .. for friends	dinner	With whom does he eat the X?
play .. for fun	piano	Why does he play X?
wash .. with the detergent	laundry	What does he wash with the detergent?
wear .. the whole week	jeans	How long does he wear X?
smoke .. with friends	cigarette	What does he smoke with friends?
grow .. in the garden	tomatoes	Where does he grow X?
clean .. before dinner	dishes	What does he clean before dinner?
brush .. before bed	teeth	When does he brush her X?
No Anticipation Verbs		
hold .. for a few seconds	umbrella	What does he do with the X?
kick .. on the ground	bucket	What does he do?
break .. on purpose	branch	What does he break?
see .. through the window	sunset	What does he see?
leave .. tomorrow	home	When does he leave X?
cut .. with scissors	rope	What happened to the rope?
touch .. on the head	tiger	Where does he touch the X?
feel .. briefly	skin	How long does he feel the X?
smell .. in the other room	dinner	Where is the smell of X coming from?
grab .. from the table	cup	Where does he grab the X from?
discuss .. with a friend	book	What does he discuss?
close .. quietly	door	How does he close X?

Note that the Passing Objects here are only for an example and include our prior expectations. The correspondence for the test will be determined by the pre-testing results. For example, if the objects we proposed are not the top object the participants write (for the positive anticipation verbs), then we use the most popular one instead.

Example:

- 🍎 *He eats an apple during a recess.* High sentence and visual anticipation. The word *eat* suggests that the object will be something edible.
- ☂️ *He holds the umbrella for a few seconds.* Still high visual anticipation, though the class of objects that can be held is much greater, resulting in lower conditional probability.
- 🎷 *He eats an apple during a recess.* Sentence anticipation is preserved, but visual is not.

- 🚗 *He holds the umbrella for a few seconds.* Neither sentence nor visual anticipation is present.

7. Pre-testing

In an open cloze task scenario, the participants will be given the sentences in the stimuli without the object in the end and will be then asked to fill in the blank (with an inanimate object-noun). This reduces our prior bias regarding what objects have good anticipation together with the specific verb. For practical purposes so that the participant's time is used effectively, they are asked to produce up to 5 words that fit the gap, sorted by the plausibility in descending order. This also increases the probability of overlap. For anticipatory verbs, we expect greater overlap than for low-anticipatory verbs. If some of the verbs do not fulfil this condition, they are either removed or substituted.

The collected objects are then used in the actual experiment (for anticipatory verbs the ones with the highest overlap). Care is taken so that no object repeats itself for multiple stimuli.

There are two alternatives to this pre-testing:

1. A forced choice between the objects that we came up with. This introduces a bias as to what we (non-native English speakers) think regarding verb anticipation and other issues, such as multiple objects fitting the gap or none of them.
2. Rating the naturality of the already filled-in sentence. This, again, is more demanding for the participant as much more samples need to be collected.

Image anticipation can be pre-tested with picture-naming. A threshold has to be established as well (or some other outlier detection). This is not the same as picture-naming in the experiment, which will, however, also be used for removing invalid instances. Pictures of the objects from the first phase of pre-testing are shown and compared whether the responses "correctly" name the word which fills in the sentences. Again, multiple responses can be recorder.

In both of these pre-testing cases, the pre-testing participants can also cheat. Here we hope that most of them will not and that the outliers can be therefore identified by standard statistic tools.

We do not need to treat uncommon/bizarre objects because they are filled in by the participants in the pre-testing phase. We assume that the verbs are common enough.

Examples:

- Sentence anticipation: *Fill in the gap so that the sentence is as natural as possible. Use inanimate objects in the form of an article + noun. Use up to 5 responses.*
    - He eats _____ during a recess.
    - He smokes _____ with friends.
- Image anticipation (on objects from the first phase): *Name the object on the screen. Use up to 5 responses.*
    - 
    - 

8. Confounds

- Image "afterthoughts": Seeing a bizarre image can slow down further work. This could also be examined in the pre-testing but would add an extra phase, which would

further slow down the experiment. Furthermore, we believe the images to be common enough to not cause this.

- We may find that SA is stronger than VA, but this may be caused solely by the fact that the object immediately follows the verb, while the image was seen a longer time ago. This is hopefully compensated by the fact that the image was interacted with.
- As far as we know, there is no simple way to resolve this issue with the current setup of showing an image and then doing a self-paced reading.
- For longer lists, the participants may notice that the image sometimes corresponds and sometimes not, and they may choose to disregard it completely. This is perhaps not an issue if they are subconsciously forced to load it into memory by the picture-naming task anyway. If we needed to combat this specifically, we could make the lists imbalance in favour of VA condition. For the data processing phase, we would have to disregard some of these samples so that we again work with the same number of samples per every condition.
- Reading times may be longer simply because of the word length. This, we hope, will not significantly affect the results, as the word lengths are not that different from each other and are distributed evenly across all conditions.
- There will also be interactions between the variables. E.g. showing 🎂 and the verb *eats* will can perhaps make the verb itself more easily digestible. However, these effects can be examined post-experiment from the collected data.

9. Lists

Only a few images are provided, for an illustration. We chose emojis, because they have been specifically designed to be easily recognized.

	List 1	List 2	List 3	List 4
1	He eats an apple during a recess. (AV + VO) 🍎	He smokes a cigarette with friends. (AV + NO) 🚬	He kicks a bucket on the ground. (NV + NO) ⏲️	He sees sunset through the window. (NV +VO) 🌇
2	He drives a car fast. (AV + NO) 🧁	He grabs a cup from the table. (NV +VO) ☕	He drinks water in the evening. (AV + VO) 🥤	He drinks water in the evening. (AV + NO) 🔥
3	He breaks the branch on purpose. (NV +VO) 🎋	He discusses a book with a friend. (NV +NO) 📯	He holds an umbrella for a few seconds. (NV +VO) ☂️	He drives a car fast. (AV + VO) 🚗
4	He sees sunset through the window. (NV + NO)	He brushes teeth before bed. (AV + VO)	He bakes cookies out of boredom. (AV + NO)	He holds an umbrella for a few seconds. (NV + NO)
5	He plays the guitar for fun. (AV + NO)	He cuts a rope with scissors. (NV +VO)	He leaves home tomorrow. (NV +VO)	He plays the guitar for fun. (AV + VO)
6	He feels skin briefly. (NV + NO)	He cooks a salad for friends. (AV + NO)	He cuts a rope with scissors. (NV + NO)	He feels skin briefly. (NV +VO)
7	He touches the tiger on the head. (NV +VO)	He wears jeans the whole week. (AV + VO)	He washes laundry with the detergent. (AV + VO)	He leaves home tomorrow. (NV + NO)
8	He cooks a salad for friends. (AV + VO)	He touches the tiger on the head. (NV +NO)	He wears jeans the whole week. (AV + NO)	He washes laundry with the detergent. (AV + NO)
9	He closes the door quietly. (NV + NO)	He bakes cookies out of boredom. (AV + VO)	He grabs a cup from the table. (NV + NO)	He smells perfume in the other room. (NV + NO)
10	He smokes a cigarette with friends. (AV + VO)	He breaks the branch on purpose. (NV + NO)	He brushes teeth before bed. (AV + NO)	He closes the door quietly. (NV +VO)
11	He grows tomatoes in the garden. (AV + NO)	He kicks a bucket on the ground. (NV +VO)	He cleans dishes before dinner. (AV + VO)	He cleans dishes before dinner. (AV + NO)
12	He discusses a book with a friend. (NV +VO)	He eats an apple during a recess. (AV + NO)	He smells perfume in the other room. (NV +VO)	He grows tomatoes in the garden. (AV + VO)

10. Fillers

The issue is that the task is already quite complex for users and has high structure to it. They first have to do picture naming, then self paced reading and finally a comprehension question. Since we limit ourselves to only sentences with a simple structure with only two slots being exchanged, then the participants can soon understand what is going on.

This however is not an issue per se. If the participants wanted to be “too helpful” or cheat, they can’t do much to affect the process. They still have to read at their maximum reading rate but are limited by their ability to answer comprehension questions.

Fillers are therefore probably not needed. They could even be harmful, as they could distract the participants from the already mildly complex task.

12. Abstract

It has been shown that the speaker's gaze at an object produces anticipation that results in the said object usually being mentioned in the following sentences, even with a lack of semantical context. The semantics of a sentence can by themselves provoke strong anticipation as well. This study is of interest in that we examine whether visual (VA), and sentence (SA) anticipations occur in different intensities, and if so, which one is the stronger in a context where both types of anticipations are present. In order to answer this question, a self-paced reading study was performed. Participants were shown images followed by a sentence, provoking either only SA, only VA, SA and VA, or neither.
In this specific setup, a two-tailed t-test showed a strong statistical difference between SA and VA ($p < 0.0001$). Furthermore, SA scenarios were found to have significantly ($p < 0.05$ with a one-tailed t-test) lower reading times than stimuli with VA by a margin of 2.1 seconds (8% of the overall average reading time). In all cases, the presence of either SA or VA reduced the reading from cases where they were not present, which is in accord with the previous work.