

Adrian Ziółkowski (University of Warsaw)

adrian.a.ziolkowski@uw.edu.pl

Tomasz Zyglewicz (CUNY Graduate Center)

tzyglewicz@gradcenter.cuny.edu

Truth-conditional variability of color ascriptions: empirical results concerning the polysemy hypothesis¹

[forthcoming in *Oxford Studies in Experimental Philosophy*, vol. 5]

[penultimate draft – please cite the published version if possible]

Abstract

Recent experimental work has shown that the truth-value judgments of color predications, i.e. utterances of the form “the leaves on my tree are green” or “these walls are brown,” are influenced by slight changes in the context of utterance (Hansen and Chemla 2013, Ziółkowski, 2021). Most explanations of this phenomenon focus on the semantics of color adjectives. However, it is not clear if these explanations do justice to the nuances of the empirical data on context-sensitivity of color predications (Ziółkowski, 2021). In contrast to the adjectival explanations, Agustin Vicente (2015) has recently proposed that the context-sensitivity of color predications can be explained by invoking the polysemy of the noun. In this paper, we present the results of three studies designed to empirically test this

¹ The authorship is equal. The authors are grateful to two anonymous reviewers, the editors, and Agustin Vicente for very helpful comments on earlier drafts of this chapter. Thank-you to Sam Alxatib, Shannon Brick, Jesse Prinz, and Mieszko Tałasiewicz for helpful discussions. This project was financed by the Faculty of Philosophy at the University of Warsaw (research subvention granted to Adrian Ziółkowski).

hypothesis: a traditional survey experiment (Study 1), an exploratory correlational study inspired by the semantic integration paradigm (Study 2a), and a follow-up experiment (Study 2b) that was designed to mitigate possible shortcomings of Study 2a. The results of our studies present preliminary evidence against Vicente’s theory.

Keywords: polysemy, ambiguity, context-sensitivity, color adjectives, Travis cases

1. Introduction

Context-shifting experiments play a prominent role in contemporary philosophy of language. They are used to “elicit intuitions about uses of an expression *e* in different imagined contexts” (Hansen and Chemla 2013, 287; Cappelen and Lepore 2005). One of the most commonly discussed such experiments is Charles Travis’ *Leaves* case, which invites us to compare our intuitions about the truth value of the sentence “these leaves are green” across the following two contexts (Travis 1994, 172; 1997).²

Acceptance. *Pia has a Japanese maple tree in her backyard that has russet (reddish brown) leaves. She paints the leaves of the tree green and puts them in a woven basket. A friend of Pia’s who is making decorations for a play asks if Pia has any green leaves she can use in her stage set. Pia shows the contents of the basket to her friend and says ‘These leaves are green.’*

² The vignettes are slightly modified (for the purpose of our studies) versions of those used by Hansen and Chemla (2013).

Rejection. *Pia has a Japanese maple tree in her backyard that has russet (reddish brown) leaves. She paints the leaves of the tree green and puts them in a woven basket. A friend of Pia's who is conducting a study of green-leaf chemistry asks if Pia has any green leaves she can use in her study. Pia shows the contents of the basket to her friend and says 'These leaves are green.'*

Travis reports the intuition that Pia's utterance is true in the Acceptance context and false in the Rejection context. He uses this observation, as well as the intuitions elicited by a wide range of other context-shifting experiments, to launch an attack on traditional truth-conditional semantics [TCS], which rests on the assumption that the truth-value of a non-ambiguous and non-elliptical sentence is determined by its syntax and the meanings of the expressions that constitute it.³

A proponent of TCS has two main ways of responding to Travis' challenge. First, they can reject Travis' intuitions and insist that the two utterances of "These leaves are green" have the same truth-value after all (e.g. Sainsbury 2001). However, this maneuver has been put under significant pressure by some recent work in experimental philosophy, which has documented that the folk's truth-value judgments in at least some context-shifting experiments align with those of Travis (Hansen and Chemla 2013; Ziółkowski 2021).

The second possible response is to accept Travis' intuitions and try to accommodate them within the framework of TCS. With respect to the green leaves case, this would mean arguing that the variation of the truth-value of the two utterances of "These leaves are green" can be traced back to the semantics of one (or more) of the expressions that make up the sentence. The majority of advocates of TCS have focused attention on the semantics of the color-adjective

³ For more precise formulations of this assumption, see for instance Kennedy and McNally (2010, 80) or Neale (2016, 229).

“green” (Szabó 2001; Rothschild and Segal 2009; Kennedy and McNally 2010; Hansen 2011; cf. Clapp 2012). Much in contrast to these approaches, Agustín Vicente has recently suggested that it is not the adjective “green” but rather the noun “leaves” that is responsible for the truth-value variability in the *Leaves* case (Vicente 2015; cf. 2012; 2017; 2018; 2021). It is our goal in this paper to put his hypothesis to an empirical test.

2. A polysemy account of the *Leaves* case

Like Christopher Kennedy and Louise McNally (2010), Vicente believes that the sentence “These leaves are green” is ambiguous and that this ambiguity is responsible for the truth-variability of the sentence across the Acceptance and the Rejection contexts. However, unlike them, he thinks that this ambiguity is generated by the ambiguity of the noun “leaves.”⁴

Crucial to his account is a distinction between two types of lexical ambiguity, namely homonymy and polysemy. An expression is ambiguous iff it has two or more senses. An ambiguous expression is homonymous with respect to its two senses, S1 and S2, iff S1 and S2 are semantically unrelated, for instance *bass*_[fish] and *bass*_[voice]; it is polysemous iff S1 and S2 are semantically related, as are for instance *book*_[text] and *book*_[tome].⁵

According to Vicente, nouns are polysemous. More specifically, he argues that they are associated with rich conceptual structures (Pustejovsky 1995; cf. Fodor and Lepore 1998) that encode information pertaining to the various aspects of their denotation (Vicente 2015, 55–56;

⁴ It is worth pointing out that recent empirical work lends support to the claim that the *reference* of natural kind terms is ambiguous (e.g. Nichols, Pinillos, and Mallon 2016; Tobia, Newman, and Knobe 2020; cf. Devitt and Porter 2021).

⁵ Spelling out what sense relatedness consists in is a notoriously difficult task. For a recent proposal, see: Liu (2022). For the ratings of semantic similarity of a number of ambiguous words, see the supplemental material of Brocher, Foraker, and Koenig (2016), as well as appendix B of Klepousniotou, Titone, and Romero (2008).

cf. Cruse 2004; Paradis 2004; Vicente 2012). Furthermore, he thinks that contexts can differentially activate these aspects, which in turn can influence our truth-value judgments in specific cases.

The two aspects of the noun “leaves” that are relevant to Vicente’s analysis of the *Leaves* case have to do with the leaves’ external appearance and their essence (“leaves-as-they-look” vs “leaves-as they-are”). He argues that the Acceptance context highlights the external appearance aspect, which explains the intuition that what Pia says is true. The intuition of Pia’s judgment in the Rejection context is explained by the fact that the context brings out the properties that are causally connected to the leaves’ essence, in particular their natural color.

Even though in this paper we focus exclusively on the *Leaves* case, it is important to note that Vicente’s explanation is intended to be more general than that. Indeed, he formulates the following prediction:

(Vicente’s Hypothesis; VH) If we have an object O of kind K, and a property P which is causally linked to the essence that O has in virtue of being a K, then ‘Det K is P’—where ‘Det K’ refers to O—is ambiguous (Vicente 2015, 63).

The view has an important advantage over the adjectival explanations of Travis’ intuitions mentioned in Section 1. In a recent series of experiments Ziółkowski (2021) tested the folk intuitions about four Travis-style context-shifting experiments involving color-predicates, called *Leaves*, *Walls*, *Kettle*, and *Apples*.⁶ Somewhat surprisingly, his participants assigned different truth values across Acceptance and Rejection only in the *Leaves* scenario. It is not clear that the adjectival explanations can account for this result.⁷ VH, by contrast, suggests a fairly intuitive

⁶ His vignettes were borrowed from Hansen and Chemla (2013).

⁷ See also Clapp (2012) for further criticisms of the adjectival explanations.

explanation of it. With respect to *Walls* and *Kettle*, it can be pointed out that even if we conceptualize some artifacts as having an essence, in most cases their color is not conceptualized as causally connected to it (a possible exception: paint). It is less obvious whether VH has the resources to shed light on the lack of truth-variability in the pair of scenarios concerning *Apples*. However, there are independent reasons for thinking that Ziółkowski’s participants had trouble grasping the difference between the Acceptance and Rejection contexts in this case (see section 4.5 of his paper).

Furthermore, VH makes a more specific prediction about the *Leaves* case. Recall that the view explains Travis’ intuition by tracing the ambiguity of “these leaves are green” to the noun “leaves.” This means that the sentence should cease to be ambiguous, if the noun “leaves” is dropped from it (Vicente 2015, 60). In other words, VH entails the following:

(Noun triggers; NT) It is the presence of the noun “leaves” in Pia’s utterance that triggers the truth-value variation across the Acceptance and the Rejection contexts.⁸

In Study 1, we approached NT in a straightforward manner and wanted to resolve the question whether removing the noun “leaves” from the crucial utterance cancels out the impact of contextual shift on intuitive truth evaluations, or at least diminishes its impact when compared to the utterance which contains the noun “leaves”. However, as we will argue later in Section 3.5, we are aware that NT might have alternative readings which require more subtle treatment than our Study 1 can offer. We will try to address these issues in our studies 2a and 2b.

⁸ "This in turn means that, unless we enrich its meaning by introducing a kind concept (i.e., the meaning of a sortal noun), an utterance of ‘these are green’ will not have the ambiguity detected in the original the-leaves-are-green example” (Vicente 2015, 60). To keep things simple, we bracket out the worry about enriching the utterance by the concept of leaves in Study 1. We will discuss this issue in more detail in section 3.5.

3. Study 1

3.1 Research Objectives and Hypotheses

NT entails two empirically testable hypotheses:

(H1) Subjects should be more inclined to agree with the color-ascription claim in the Acceptance context than in the Rejection context when the noun “leaves” occurs in the target utterance.

(H2) Subjects should be equally likely to agree with the color-ascription claim in the Acceptance context and in the Rejection context when the noun “leaves” does not occur in the target utterance.

We wanted to test these hypotheses in a large-scale, preregistered setting. The preregistration form can be found on the OSF platform (<https://osf.io/m83v4>). Since H2 is a negative prediction (i.e. we did not expect to find significant differences between the Acceptance and Rejection conditions if the noun is removed from the target utterance), we wanted to achieve enough statistical power to detect the effect in question, if it really exists (for details, see Section 3.2).

3.2 Materials and Procedure

We conducted our first experiment in a 2x2 between-subject design. Each subject was randomly assigned to one of four conditions: either the Acceptance or Rejection context, and either the noun-present or noun-absent condition (the target sentence “These leaves are green” or “These are green”).

The study was designed with Qualtrics and carried out online. After reading the consent form and agreeing to participate in the study, participants were presented with a variant of the *Leaves* scenario depending on the random assignment to an experimental condition (see Section

1. for precise formulation). The main dependent variable was measured identically in all conditions: respondents were asked “To what extent do you agree that the character's claim appearing in boldface is true?” and offered a 5-point Likert scale ranging from ‘Strongly disagree’ (numbered 1) and ‘Strongly agree’ (numbered 5); only the edge-points were described.

Afterwards, each subject had to pass a short, standard attention check (for details, see the preregistration form⁹). Only those subjects who passed the attention check were included in the final analysis.

At the end of the survey, we collected some basic demographic data about our participants: their age, gender, and education. We also asked them if they hold a degree in philosophy: those subjects who reported having a BA, MA or PhD in philosophy were excluded from the final analysis.

In our preregistration, we assumed that we need to collect at least 88 valid responses for each experimental condition (352 in total) in order to achieve 95% power to detect a medium-sized (Cohen’s $d = 0.5$) difference in average ratings between the Acceptance and Rejection context in the noun-present and noun-absent conditions (t test, one tailed).

3.3 Participants

Participants were recruited via Prolific, a professional platform for online surveying for academic purposes. They received £0.45 for taking the survey. Overall, 501 respondents completed the survey, 70 of whom were excluded from the analysis (for the exclusion criteria, see the section above), which yields the final sample size of $N = 431$. This exceeds the sample size we assumed

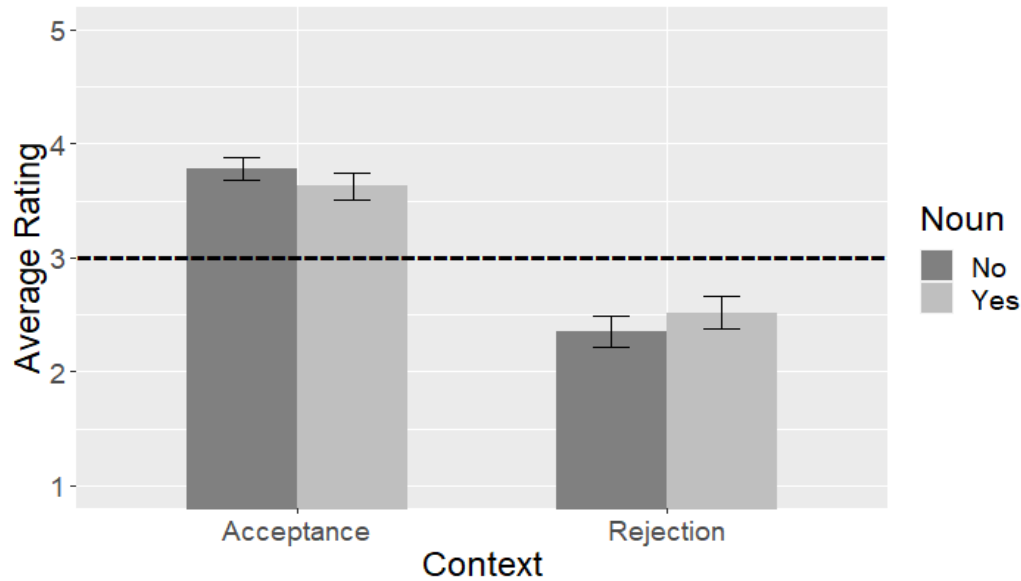
⁹ The data we collected in our three studies and supplementary materials are publicly accessible on the OSF platform: <https://osf.io/82qny/>

in the preregistration (over 100 participants fell in each condition). 49.2% of participants self-identified as female, 48.7% as male, and 2.1% as non-binary. The average age of respondents was 45.11 ($SD = 16.06$).

3.4 Results

In order to test H1 and H2, we conducted an independent-samples t test for each of these hypotheses. Unsurprisingly, we found a significant difference between participants' ratings in the Acceptance and Rejection contexts in the noun-present condition: $t(213) = 6.18, p < 0.001$. Respondents were less likely to agree with the target utterance in the Rejection context ($M = 2.52; SD = 1.44$) than in the Acceptance context ($M = 3.63; SD = 1.17$); $d = 0.84$. This confirms H1 and is in line with the previous findings reported by Hansen and Chemla (2013) and Ziółkowski (2021). However, contrary to Vicente's (2015) predictions, we also found a significant effect of context (similar in size) in the noun-absent condition: $t(214) = 8.49, p < 0.001$. Again, participants were more inclined to agree with the utterance in the Acceptance context ($M = 3.78; SD = 1.04$) than in the Rejection context ($M = 2.36; SD = 1.4$); $d = 1.15$. Therefore, we did not find evidence in support of H2. The results are summarized in Figure 1.

Figure 1. Average truth evaluations in Study 1 depending on the experimental condition. Error bars represent standard error of mean. The midpoint is marked with a dashed line.



3.5 Discussion

The most interesting finding of Study 1 is that the folk share Travis’ intuitions about the *Leaves* case, even when the noun “leaves” is removed from Pia’s utterance. *Prima facie*, this result might seem to speak against NT. However, we think that this conclusion would be too hasty.

It is important to distinguish between two readings of NT. On the first one, NT is a thesis about the surface form (SF), i.e. about whether the sound /li:vz/ was actually made, or the word “leaves” was actually inscribed:

(NT-SF) It is the presence of the noun “leaves” in the surface form of Pia’s utterance that triggers the truth-value variation across the Acceptance and the Rejection contexts.

Indeed, the results of Study 1 constitute compelling evidence against (NT-SF). Nevertheless, linguists routinely posit the existence of deeper levels of linguistic representation, the structure of which is not always perfectly mirrored by the SF. In particular, the absence of a word from the

surface form does not mean that it is absent from some underlying level of representation (UR).¹⁰

Hence the second reading of NT:

(NT-UR) It is the presence of the noun (concept) “leaves” in the underlying representation of Pia’s utterance that triggers the truth-value variation across the Acceptance and the Rejection contexts.

Now, it bears emphasizing that most semantic theorizing concerns the UR, not the SF. Indeed, NT-UR is presumably the intended reading of NT. Nonetheless, in our first experiment we focused on NT-SF. This decision was prompted by the fact that while the SF of the target utterance is fully under the experimenter’s control, it is much more difficult to reliably manipulate how participants actually represent the utterance. Furthermore, if it turned out that the manipulation of the SF (“these are green” vs “these leaves are green”) leads participants to assign different truth values to the target sentence, this would provide very good reasons for thinking that the two sentences differ at the level of UR as well. For, this difference would presumably have to do with the absence of the noun “leaves” in the UR of “these are green.” To sum up: if NT-SF were true, very likely so would be NT-UR. However, the falsity of NT-SF does not entail the falsity of NT-UR.

In fact, there are good reasons for thinking that the UR of “these are green” contains the noun “leaves.” First, the demonstrative “these” is capable of picking out different objects in different situations. In particular, it is possible that what “these” contributes to the UR is not just the contents of Pia’s basket but the contents of Pia’s basket-*qua*-leaves. Second, several authors have suggested that sentences with simple demonstratives, like the one uttered by Pia, actually

¹⁰ For the sake of simplicity, henceforth we will speak of *the* underlying level of representation.

contain an ellipted nominal complementizer. According to this analysis, the UR of Pia's utterance is something along the lines of: $[[_{DP} \text{these } e][_{VP} \text{are green}]]$ (e.g. Abney 1987; Cohan 1998), where e stands for the ellipted material.¹¹ If this is correct, then in interpreting Pia's utterance participants look to the linguistic and extra-linguistic context to find the antecedent of e (Llombart-Huesca 2002, 64; Lobeck 1995, 25–26; cf. Chao 1987). And indeed, given the contents of our vignette, leaves seem to be the most plausible candidate for it.¹²

Given the above, the following is a plausible alternative explanation of our data that is consistent with NT-UR:

(Contextual Supplementation; CS) Even though in the no-noun conditions Pia says “These are green,” participants nonetheless tended to contextually supply the noun/concept “leaves,” which in turn led them to evaluate the sentence “these leaves are green” when rating the utterance (Vicente 2015, 60, fn. 15).

Now, if CS is true, then the no-noun condition, as designed in our first study, functions – as far as the UR is concerned – just like the noun condition; but if this is so, then our experiment is an inadequate test of NT-UR. In our second, exploratory study we wanted to put this alternative explanation to an empirical test.

4. Study 2a

4.1 Research Objectives and Hypotheses

¹¹ This analysis has at least two advantages. First, it avoids positing different syntactic trees for “these” as it occurs in “these are green” and “these leaves are green.” Second, it meshes well with the claim that both demonstratives and pronouns belong to the category of determiners. On this analysis, personal pronouns like “they” obligatorily take a null noun complement, hence the non-acceptability of sentences like “they leaves are green.”

¹² We are grateful to the editors for encouraging us to include a discussion of ellipsis.

In the second study, our aim was to look for evidence that a combination of NT-UR and CS is an empirically adequate explanation of the results of our first experiment. This time around, we focused exclusively on the no-noun/rejection condition, as it is the only condition whose results pose a problem for NT-UR (which predicts that the utterance of “these are green” will be judged true in both Acceptance and Rejection contexts).

We decided to use a combination of the truth-evaluation task from the first experiment and a recognition task that was meant to be an implicit measure of the activation of the concept “leaves”. Our recognition task was inspired by the semantic integration paradigm (Bransford and Franks 1971; Sulin and Dooling 1974; Gentner 1981; Powell et al. 2013; Waskan et al. 2014; Powell et al. 2015; Harmon and Horne 2016). The central idea behind the recognition task was that if a participant falsely recalls that the sentence “These leaves are green” was present in the story, that counts as evidence that they contextually supplied the concept “leaves” when evaluating the utterance “These are green.” We predicted that the more consistent a participant’s evaluation of the target sentence with NT-UR (i.e. the higher their answer on the Likert scale), the less likely they will be to falsely recall that the sentence “these leaves are green” was present in the scenario.

4.2 Materials and Procedure

The second study was exploratory and was aimed to detect a correlation between folk truth-value judgments in the *Leaves* scenario and false recognitions of the sentence “These leaves are green”; the design was correlational, not experimental.

Again, the survey started with the consent form; after agreeing to participate in the study, respondents read the Rejection, noun-absent variant of the *Leaves* scenario (with the utterance

“These are green”) and answered the same truth-evaluation question as in Study 1. They were explicitly asked to memorize as many details about the vignette as possible. Afterwards, they were presented with a series of four simple mathematical tasks that were distractors used to erase participants’ short-term memory, and also served as an attention check. They were modeled on standard distractors employed in psychological studies concerning memory (sometimes called “complex span tasks,” see, e.g., Waris et al. (2017)). The tasks were presented separately in random order (see supplementary online materials for details: <https://osf.io/p3r7h>). Only these respondents who solved all four tasks correctly were included in the final analysis.

After completing the mathematical tasks, respondents participated in the recognition task, which consisted of three questions, each starting with the following prompt: “Was the following sentence present in the scenario about Pia?” (binary answer options: “yes”; “no”). In the critical question the sentence was: “Pia shows the contents of the basket to the friend and says ‘These leaves are green.’” Besides that, they were asked to evaluate two other sentences: “Pia has a Canadian larch tree in her backyard that has light green needles.” (which did not appear in the scenario) and “A friend of Pia’s who is conducting a study of green-leaf chemistry asks if Pia has any green leaves she can use in her study.” (which in fact appeared in the scenario). These two additional recognition questions allowed us to check to what degree our participants memorized the vignette.

We collected basic demographic information about the respondents at the end of the survey: their age, gender, and education. Again, participants who reported holding a degree in philosophy were not included in the final analysis.

4.3 Participants

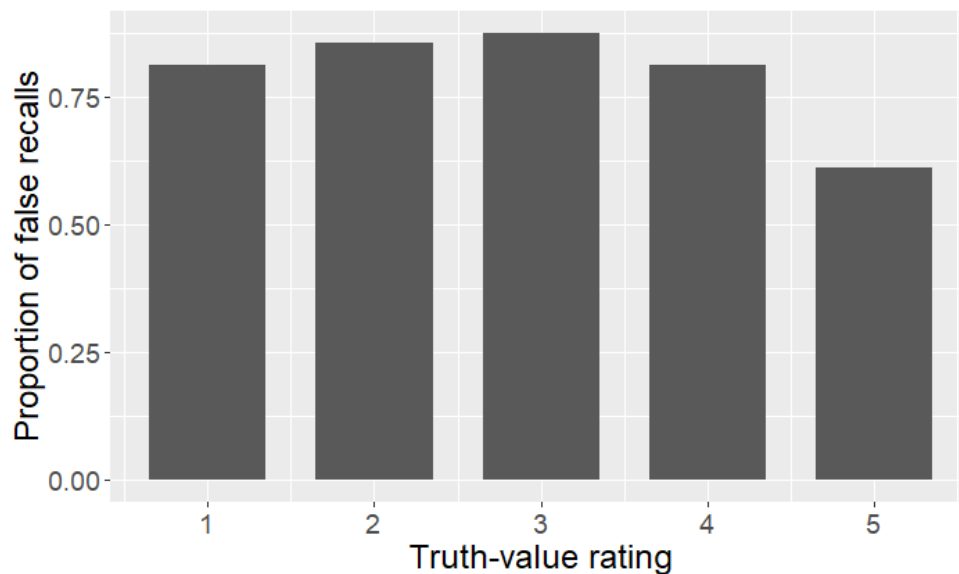
Similarly as in Study 1, the participants were recruited via Prolific and were paid £0.7 for taking the survey. In total, 201 respondents participated in the study, but nineteen were excluded from the final analysis, therefore the sample size was $N = 182$. 49.5% self-identified as female, 48.9% as male, and 1.6% as non-binary. Their average age was 36.18 ($SD = 14.76$).

4.4 Results

The participants' truth-value ratings were highly consistent with what was observed in Study 1: respondents were rather unlikely to agree with the target utterance in the Rejection, noun-absent condition ($M = 2.49$; $SD = 1.34$).

When it comes to the recognition task, the participants were highly efficient in recalling the sentence that in fact occurred in the vignette (96.7% correct answers) and in denying that the sentence "Pia has a Canadian larch tree in her backyard that has light green needles." occurred in the vignette (100% correct answers). However, the ratio of false recognitions regarding the sentence "Pia shows the contents of the basket to the friend and says 'These leaves are green'." was surprisingly high: 81.3%. In such circumstances, it was natural that high correlation between participants' truth-value ratings and false recognitions of the crucial sentence could not be observed: $r = 0.099$; but the direction was consistent with CS. If we look at the percentages of false recalls depending on the rating in the truth-value question, we can see a trend: participants who gave the highest possible rating ("Strongly agree") were less likely to falsely recall the critical sentence (see Figure 2).

Figure 2. Percentages of false recalls of the sentence “Pia shows the contents of the basket to the friend and says “These leaves are green” depending on the truth-value judgment (1 = “Strongly disagree”; 5 = “Strongly agree”) in Study 2a.



4.5 Discussion

Although our second study does not provide direct, clear evidence in favor of CS – the method we used can only indirectly hint at a possible effect of contextual supplementation – the large number of false recalls of the sentence “Pia shows the contents of the basket to the friend and says “These leaves are green” alone can suggest that contextual supplementation of the noun/concept “leaves” might have played a role in shaping the results of Study 1 in the noun-absent condition. Moreover, the trend which we observed – that participants who gave high ratings in the truth-value question were less likely to falsely recall the crucial sentence – provides further reasons to think that CS might be a plausible explanation of our failure to support H2 in Study 1.

It is also possible the high rate of false recalls of the critical sentence, but not of the two control sentences, can be explained by the comparative difficulty of the critical recognition task.

As opposed to the control sentence that did not occur in the scenario, the critical sentence contained only a subtle modification of a sentence that actually occurred in the text (a removal of a single word). Relatedly, one might worry that by not offering the participants the option to select a sentence that actually occurred in the text, we biased them in favor of a sentence very similar to it, which for most intents and purposes might be considered equivalent to it.¹³ We decided to run a follow-up study to control for these potential confounds.

5. Study 2b

5. 1 Research Objectives and Hypotheses

Study 2b had the same goal as Study 2a; we also tried to find evidence for CS using the semantic integration paradigm, but we decided to simplify the crucial recognition task, since we worried that the task used in Study 2a might have been too difficult for our participants. Our hypothesis was similar to the one we tested in Study 2a: we predicted that participants who give a negative truth-value verdict in the Rejection, noun-absent variant of the *Leaves* scenario, will be more likely to falsely recall that the noun “leaves” occurred in the crucial utterance of the protagonist in the scenario than participants who give a positive truth-value judgment.

5. 2 Materials and Procedure

Just like in Study 2a, after agreeing to participate in the study, the participants of Study 2b read the Rejection, noun-absent version of the *Leaves* scenario. However, we decided to simplify the answer format for the question concerning the truth-value of the critical utterance, “These are green”; instead of a Likert scale, which was used in Study 2a, we used a simple dichotomous scale.

¹³ We are grateful to an anonymous referee for pointing this out.

After seeing the question “Is the character's claim appearing in boldface true or false?” the respondents expressed their verdict by choosing between “True” and “False”. After having answered this question, they were presented, in randomized order and on separate screens, with the same series of filler tasks (aimed at erasing participants’ short-term memory) as in Study 2a.

The recognition task was simplified in comparison to Study 2a. It consisted of three questions, each of which started with the prompt “Which of the following sentences was present in the scenario about Pia?”. The sentence in the critical recognition task was shorter; it did not include the phrase “Pia shows the contents of the basket to the friend and says.” Participants were offered two options: “These are green” (correct answer) and “These leaves are green” (false recall). Additionally, we used two control questions that were aimed to check whether our respondents memorized other relevant details of the scenario. In one, they could choose from “A friend of Pia’s who is making decorations for a play asks if Pia has any green leaves she can use in her stage set” (correct answer) and “A friend of Pia’s who is conducting a study on green-leaf chemistry asks if Pia has any green leaves she can use in her study” (false recall). In the other one, the options were: “Once the paint dries, Pia collects the leaves and puts them in a woven basket” (correct answer) or “Once the paint dries, Pia collects the leaves and puts them in a metal bucket” (false recall). The questions were presented on the same screen in randomized order.

Similarly as in Study 2a, the survey ended with a short demographic questionnaire. Same exclusion criteria were applied as in Study 2a.

5. 3 Participants

We recruited our participants via Prolific and offered them £0.7 for taking the survey. Prolific users who participated in Study 2a were not able to take part in Study 2b. 200 respondents

completed the survey, of which twenty four were excluded, due to incorrectly answering at least one of our mathematical tasks or having a degree in philosophy. Thus, the final sample size was $N = 178$. The average age of the participants was 41.2 years ($SD = 14,36$); 51% self-identified as female and 49% as male.

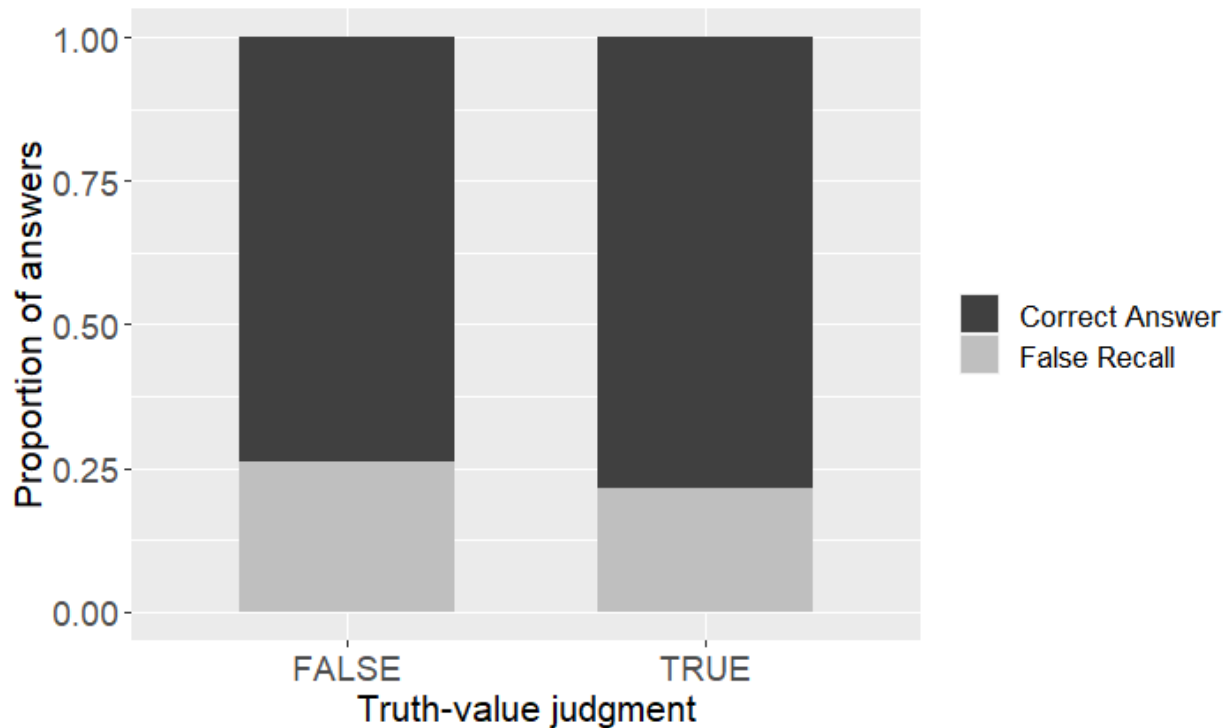
5. 4 Results

The data regarding participants' intuitions about the truth-value of the crucial utterance were, unsurprisingly, in line with the findings of our previous two studies. Laypersons were more likely to judge that the utterance "These are green" in the Rejection context is false (57.9%) than true (42.1%); $\chi^2(N = 178) = 4.4$; $p = 0.035$.

When it comes to our control memory tasks, only one respondent (0.6%) falsely recalled the sentence "Once the paint dries, Pia collects the leaves and puts them in a metal bucket" and only five (2.8%) failed the recognition task about the motivations of Pia's friend request; thus, we can say that our participants were very effective at memorizing the relevant features of the scenario. On the other hand, when it comes to the critical recognition question, overall, a noticeable number of participants (24.2%) falsely recalled the sentence "These leaves are green" and the remaining 75.8% correctly chose the option "These are green". Nevertheless, we did not find the predicted relationship between negative judgments in the truth-value question and false recalls in the critical recognition task. Regardless of the answer provided in the question concerning the truth-value of the crucial utterance, participants were likely to falsely recall the sentence "These leaves are green" to a similar degree: among those who gave a positive judgment, there were 21.3% of false recalls, and among those who judged the utterance false, there were 26.2% of false recalls;

$z = 0.75$; $p = 0.226$ (one-tailed). Therefore, we cannot conclude that Study 2b yielded data in support of CS. The main result is illustrated in the figure below.

Figure 3. Percentages of false recalls of the sentence “These leaves are green” and correct recalls (“These are green”) depending on the truth-value judgment (“True” or “False”) in Study 2b.



5. 5 Discussion

As pointed out in Section 4.5, we worried that the methods we employed in Study 2a were not sensitive enough to detect evidence in favor of the Contextual Supplementation hypothesis (CS). Our main doubt was that the critical recognition task might have been too difficult and that, for this reason, it was not possible to observe the relationship between verdicts concerning the truth-value of the target utterance of the protagonist in the Rejection, noun-absent variant of the *Leaves* scenario (“These are green”) and false recognitions of the noun “leaves.” Therefore, we decided to conduct Study 2b, in which the critical recognition task was designed to be less

difficult. In fact, it turned out to be so, while nevertheless eliciting a noticeable number of false recalls of the sentence “These leaves are green.” However, we discovered no relationship between the truth-value judgments and the frequency of false recalls. Therefore, we found no evidence that the negative truth-value judgments in the Rejection, noun-absent version of the Leaves scenario are driven by the mechanism of contextual supplementation of the noun “leaves” in the target utterance “These are green”.

6. Conclusions

The fact that for some context-shifting experiments, contextual shift has the predicted impact on folk semantic intuitions, seems to be a well-documented empirical phenomenon. Although not every context-shifting scenario designed by philosophers affects folk judgments in the same way – as reported by Ziółkowski (2021), for some scenarios the difference in truth-evaluations between the Acceptance and Rejection context is very small or there is no difference at all – the contextual effect in the *Leaves* scenario seems to be robust, as documented in previous research by Hansen and Chemla (2013), Ziółkowski (2021), and our Study 1.

We believe that further empirical investigation concerning context-shifting experiments should focus on the sources of differences between various scenarios and, in cases where robust contextual effects are found, discovering the mechanisms responsible for generating these patterns of folk intuitions. Our two studies presented in this chapter are an invitation to conducting more experiments that would explore these issues. Here, we focused exclusively on the problem of context-sensitivity of color ascriptions and the theoretical explanatory framework

offered by Vicente (2015) that attributes such contextual effects to the polysemy of nouns (rather than context-sensitivity of color adjectives, as it is standardly assumed).

According to Vicente's theory, it is the noun "leaves" that causes the truth-value variation across the Acceptance and the Rejection context. We pointed out that this can be understood as a claim about the surface form (NT-SF), or about some underlying level of representation (NT-UR). Our Study 1, which was a standard χ -phi adaptation of context-shifting experiments, provided strong evidence against NT-SF. Our studies 2a and 2b were aimed at finding possible evidence in favor of NT-UR. In this case, we had to go beyond the standard survey methods that have been used in research on context-sensitivity thus far. We offered a new approach that was inspired by the semantic integration paradigm and involved indirect reasoning from the results of a recognition task. We failed to detect a significant difference in false recall rates (an indirect measure of the activation of the concept "leaves") between participants who answered the truth-value question consistently with Vicente's theory and those who did not. This provides some reason to be skeptical about NT-UR. All in all, our studies provide preliminary empirical evidence against Vicente's account of Travis-style truth-conditional variation. Nonetheless, this conclusion has to be taken with caution, given the exploratory nature of studies 2a and 2b.

We hope that our project will encourage other researchers to develop new methods that will foster further exploration of the mechanisms responsible for the sensitivity of folk truth-evaluations to contextual variation.

References

- Abney, Steven P. 1987. "The English Noun Phrase in Its Sentential Aspect." Massachusetts Institute of Technology.
- Bransford, John D., and Jeffery J. Franks. 1971. "The Abstraction of Linguistic Ideas." *Cognitive Psychology* 2 (4): 331–50.
- Brocher, Andreas, Stephani Foraker, and Jean-Pierre Koenig. 2016. "Processing of Irregular Polysemes in Sentence Reading." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42 (11): 1798.
- Cappelen, Herman, and Ernest Lepore. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. John Wiley & Sons.
- Chao, Wynn. 1987. "On Ellipsis (Syntax)." University of Massachusetts Amherst.
- Clapp, Lenny. 2012. "Indexical Color Predicates: Truth Conditional Semantics vs. Truth Conditional Pragmatics." *Canadian Journal of Philosophy* 42 (2): 71–100.
- Cohan, Jocelyn. 1998. "Semantic Features of Determiners: Toward an Account for Complements of D." Edited by Alexis Dimitriadis, Hikyung Lee, Christine Moisset, and Alexander Williams. *University of Pennsylvania Working Papers in Linguistics* 5 (1): 37–47.
- Cruse, D. Alan. 2004. "Lexical Facets and Metonymy." *Ilha Do Desterro a Journal of English Language, Literatures in English And Cultural Studies*, no. 47: 073–096.
- Devitt, Michael, and Brian C. Porter. 2021. "Testing the Reference of Biological Kind Terms." *Cognitive Science* 45. <https://doi.org/10.1111/cogs.12979>.
- Fodor, Jerry A., and Ernie Lepore. 1998. "The Emptiness of the Lexicon: Reflections on James Pustejovsky's The Generative Lexicon." *Linguistic Inquiry* 29 (2): 269–88.
- Gentner, Dedre. 1981. "Integrating Verb Meanings into Context." *Discourse Processes* 4 (4): 349–75.
- Hansen, Nat. 2011. "Color Adjectives and Radical Contextualism." *Linguistics and Philosophy* 34 (3): 201–21.
- Hansen, Nat, and Emmanuel Chemla. 2013. "Experimenting on Contextualism." *Mind & Language* 28 (3): 286–321.
- Harmon, Ian, and Zachary Horne. 2016. "Evidence for Anti-Intellectualism about Know-How from a Sentence Recognition Task." *Synthese* 193 (9): 2929–47.
- Kennedy, Christopher, and Louise McNally. 2010. "Color, Context, and Compositionality." *Synthese* 174 (1): 79–98.
- Klepousniotou, Ekaterini, Debra Titone, and Carolina Romero. 2008. "Making Sense of Word Senses: The Comprehension of Polysemy Depends on Sense Overlap." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34 (6): 1534.
- Liu, Michelle. 2022. "Mental Imagery and Polysemy Processing." *Journal of Consciousness Studies* 29 (5–6): 176–89.
- Llombart-Huesca, Amàlia. 2002. "Anaphoric One and NP-ellipsis." *Studia Linguistica* 56 (1): 59–89.
- Lobeck, Anne C. 1995. *Ellipsis: Functional Heads, Licensing, and Identification*. New York & Oxford: Oxford University Press.
- Neale, Stephen. 2016. "Silent Reference." In *Meanings and Other Things: Themes from the Work of Stephen Schiffer.*, edited by Gary Ostertag, 229–344. New York: Oxford University Press.
- Nichols, Shaun, N. Ángel Pinillos, and Ron Mallon. 2016. "Ambiguous Reference." *Mind* 125 (497): 145–75.

- Paradis, Carita. 2004. "Where Does Metonymy Stop? Senses, Facets, and Active Zones." *Metaphor and Symbol* 19 (4): 245–64.
- Powell, Derek, Zachary Horne, Angel Pinillos, and Keith Holyoak. 2013. "Justified True Belief Triggers False Recall of 'Knowing'." In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 35.
- Powell, Derek, Zachary Horne, N. Ángel Pinillos, and Keith J. Holyoak. 2015. "A Bayesian Framework for Knowledge Attribution: Evidence from Semantic Integration." *Cognition* 139: 92–104.
- Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: The MIT Press.
- Rothschild, Daniel, and Gabriel Segal. 2009. "Indexical Predicates." *Mind & Language* 24 (4): 467–93.
- Sainsbury, Richard M. 2001. "Two Ways to Smoke a Cigarette." *Ratio* 14 (4): 386–406.
- Sulin, Rebecca A., and D. James Dooling. 1974. "Intrusion of a Thematic Idea in Retention of Prose." *Journal of Experimental Psychology* 103 (2): 255.
- Szabó, Zoltán Gendler. 2001. "Adjectives in Context." *Pragmatics and Beyond New Series*, 119–46.
- Tobia, Kevin P., George E. Newman, and Joshua Knobe. 2020. "Water Is and Is Not H₂O." *Mind & Language* 35 (2): 183–208.
- Travis, Charles. 1994. "On Constraints of Generality." In *Proceedings of the Aristotelian Society*, 94:165–88.
- . 1997. "Pragmatics." In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright, 87–107. Oxford: Blackwell Publishers.
- Vicente, Agustín. 2012. "On Travis Cases." *Linguistics and Philosophy* 35 (1): 3–19.
- . 2015. "The Green Leaves and the Expert: Polysemy and Truth-Conditional Variability." *Lingua* 157: 54–65.
- . 2017. "What Words Mean and Express: Semantics and Pragmatics of Kind Terms and Verbs." *Journal of Pragmatics* 117: 231–44.
- . 2018. "Polysemy and Word Meaning: An Account of Lexical Meaning for Different Kinds of Content Words." *Philosophical Studies* 175 (4): 947–68.
- . 2021. "Chomskyan Arguments against Truth-Conditional Semantics Based on Variability and Co-Predication." *Erkenntnis* 86 (4): 919–40.
- Waris, Otto, Anna Soveri, Miikka Ahti, Russell C. Hoeffing, Daniel Ventus, Susanne M. Jaeggi, Aaron R. Seitz, and Matti Laine. 2017. "A Latent Factor Analysis of Working Memory Measures Using Large-Scale Data." *Frontiers in Psychology* 8: 1062.
- Waskan, Jonathan, Ian Harmon, Zachary Horne, Joseph Spino, and John Clevenger. 2014. "Explanatory Anti-Psychologism Overturned by Lay and Scientific Case Classifications." *Synthese* 191 (5): 1013–35.
- Ziółkowski, Adrian. 2021. "The Context-Sensitivity of Color Adjectives and Folk Intuitions." *Filozofia Nauki* 29 (2): 157–88.