

UCLA

Experiments in Linguistic Meaning

Title

Experientiality markers in memory reports: A semantics-pragmatics puzzle

Permalink

<https://escholarship.org/uc/item/6wv0566j>

Journal

Experiments in Linguistic Meaning, 3(1)

Authors

Rosina, Emil Eva

Liefke, Kristina

Publication Date

2025-01-24

DOI

10.3765/elm.3.5815

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Experientiality markers in memory reports: A semantics-pragmatics puzzle

Emil Eva Rosina & Kristina Liefke*

Abstract. Some recent work in semantics and the philosophy of language suggests that the way we report events reflects whether we have personally experienced or witnessed these events (i.e. through linguistic elements dubbed ‘experientiality markers’). This paper provides experimental support for one such marker: German non-manner uses of *wie* [‘how’]. We argue that when they are embedded under the memory predicates *noch wissen* [‘still know’] and *sich erinnern* [‘REFL-remind’], free relative *wie*-complements mark the remembering of a personally experienced event. We support this claim through a series of online studies based on scale judgements. The results of our main study raise questions about the semantics-pragmatics interface of the experientiality marking property of *wie*, and about the robustness of experientiality markers in general. A series of complementary studies address these questions.

Keywords. Experiential remembering; memory predicates; attitude reports; knowledge; evidentiality; study formats; propositional attitudes; pragmatic competitions

1. Introduction. German memory predicates can combine with a declarative *dass*-[‘that’]-clause (1-b) and with an eventive-*wie* [‘how’] free relative (1-a).¹ This holds both for the reflexive predicate *sich erinnern* (lit. ‘oneself remind’) and for the complex *noch wissen* (lit. ‘still know’).

- (1) a. Ich {erinnere mich/ weiß noch}, wie Oma im Meer geschwommen ist.
I remind myself/ know still how Grandma in-the sea swim is
‘I remember Grandma swimming in the sea.’
b. Ich {erinnere mich/ weiß noch}, dass Oma im Meer geschwommen ist.
I remind myself/ know still that Grandma in-the sea swim is
‘I remember that Grandma was swimming in the sea.’ (German)

Interestingly, *dass*-clauses and *wie*-free-relatives can be coordinated under either of the above predicates. Following familiar ambiguity tests (cf. Sadock & Zwicky 1975), we thus assume a uniform semantics for *noch wissen* in (1-a) and (1-b) and another uniform semantics for *sich erinnern* in (1-a) and (1-b), such that these sentences use the same semantic entry for the matrix verb and hence form minimal pairs in our studies. In Rosina & Liefke (2024a), we give such a unified (and fully compositional) semantics of *noch wissen* as retained knowledge – of an informationally rich

*We thank Yonca Klisch, Jonas Koopmann and Robert Kurth for their help with the study design. We thank Markus Werning, Nina Haslinger, Justin D’Ambrosio, Alex Wiegmann, Sebastian Walter, Deniz Özyıldız, and Jan Köpping, and the audiences and reviewers of ELM 3, SinFonJA, WCCFL, DGfS, the Frankfurt Semantics Colloquium, and the MECORE closing workshop for their helpful comments. Authors: Emil Eva Rosina, Ruhr University Bochum (emil.rosina@ruhr-uni-bochum.de) & Kristina Liefke, Ruhr University Bochum (kristina.liefke@ruhr-uni-bochum.de). The research for this paper is supported by the German Research Foundation DFG as part of the research unit FOR 2812 *Constructing Scenarios of the Past* (grant no. 397530566).

¹German only admits two of the three uses of *how* that have been attested for English: next to the familiar manner-*wie* (‘the way in which’), it only allows for eventive uses of non-manner *wie* (the sole focus of the present paper; cf. Umbach et al. 2022, Liefke 2023). Factive uses of *how* (‘They told me how the tooth-fairy exists’, cf. Legate 2010) are not attested for German.

(*wie*) or poor (*dass*) proposition. Our semantics suggests (but does not explicitly claim) an extension to (stative uses of)² *sich erinnern*, English *remember* and other memory predicates, such that STILL+KNOW is the core of remembering and *noch wissen* is just the most transparent spell-out.

The present paper presents a series of online behavioural studies. Three of these studies (the ones presented in Sect. 2) vary with respect to the (German and English) memory predicates used in the test sentences. Their very similar results suggest that these predicates may indeed share a semantic core that interacts with different complements in a uniform way across the concrete spell-out of these predicates. Concerning the truth- and utterance-conditions of sentences like (1-a) and (1-b), a satisfactory semantic account has to position itself relative to philosophical, psychological, and neuroscientific work on memory. This work (going back to Tulving 1972) commonly distinguishes experiential (‘episodic’) remembering (i.e. recall of a personally experienced event) from fact-only (‘semantic’) remembering, i.e. recall of general facts, often based on indirect evidence or testimony. In our experiments, we introduce the siblings Red and Blue to personify these kinds of experience. In particular, we tell our participants the following about them in the case of the swimming scene that relates to the sentences in (1):

- (2) a. Red spent the summer two years ago with Grandma and saw her swimming in the sea.
- b. Blue spent that summer abroad and was told about Grandma’s swimming much later.

Based on our semantics in Rosina & Liefke (2024a) and in line with literature on non-manner uses of *how* (Liefke 2023, Umbach et al. 2022), we expect that (1-a) unambiguously reports experiential memory (Red, (2-a)) while (1-b) is expected to report both fact-only (Blue, (2-b)) and experiential memory (cf. Fig. 1). By confirming this, we provide the first empirical evidence for experientiality markers in memory reports. The idea that *how* we report events reflects whether we have personally experienced or witnessed these events (i.e. through particular linguistic elements dubbed ‘experientiality markers’) is found in Bernecker (2010) a.o. Beyond its contribution to semantics, the empirical identification of experientiality markers in memory reports might be taken to provide further evidence for the two philosophically distinct types of remembering (i.e. experiential and fact-only remembering, going back to Tulving 1972), given some bridging principles. Additionally, being able to pinpoint specific experientiality markers like German *wie* is particularly useful for analyzing production data from psychological memory studies that did not themselves create the remembered (and subsequently reported) experience.

An important feature of Rosina & Liefke (2024a) is that, in this account, experientiality is not directly encoded in the semantics. Instead, this account attributes low acceptance rates of Blue’s (2-b) ‘remember how’-equivalents to Blue’s lack of good *evidence* for the informationally rich proposition ‘how *p*’ (roughly: ‘that the things were such-and-such when *p* in @’; see Rosina & Liefke 2024a for the interplay of informational richness and evidence-based knowledge). The intuition that one must have personally experienced an event in order to truthfully self-attribute ‘remembering how’ is an indirect effect of our world knowledge that direct experience is usually the best kind of evidence. Low ratings for BLUE+HOW-sentences by themselves do not distinguish between this account, an account that views *direct* (as opposed to just good) evidence as a require-

²Our semantics cannot, in its current form, yet capture eventive ‘is remembering right now’-uses that some memory predicates have in addition.

ment, and an account that writes the requirement of personal *experience* directly into the semantics (Stephenson 2010, Liefke & Werning 2024) – because Blue lacks all of these. To our knowledge, we are the only ones choosing the first option (good evidence), while the second (direct evidence) is prominent in related literature on perception (see e.g. Davis & Landau 2021 on *see that* vs. *see -ing*), and the third (personal experience) in previous philosophical and semantic accounts of experiential memory (Stephenson 2010, Liefke & Werning 2024). Our three studies targeting the semantics-pragmatics interface of memory reports, including this issue, are presented in Sect. 3. For a more detailed discussion of possible pragmatic effects and the inter-disciplinary significance of memory reports, see Rosina (2024).

2. Attesting experientiality markers.

2.1. GERMAN MAIN STUDY. In this section, we present our biggest study as a case of our general experimental paradigm. In the later sections, we introduce a series of complementary studies that are all variations of this main study in different respects. If not indicated otherwise, the experimental design is as described for the main study. More specifically, all studies except the QUD study (see Sect. 3.2) share the same basic background story of a family gathering (introduced below). With the exception of the speaker-ID study (see Sect. 3.1), all studies share the rating format, and all except the intermediate evidence study (see Sect. 3.3) share the same set of characters.

Alongside our target characters Red and Blue (see ex. (2) in Sect. 1), we introduce their cousin Pinkie for controls and tell the participants that Pinkie does not have *any* evidence concerning the events depicted in our target vignettes. The three teenage cousins represent different types of experience/evidence with respect to different past events of some mishap involving their grandmother. Fig. 1 introduces the background story and the characters' viewpoints on one example event.

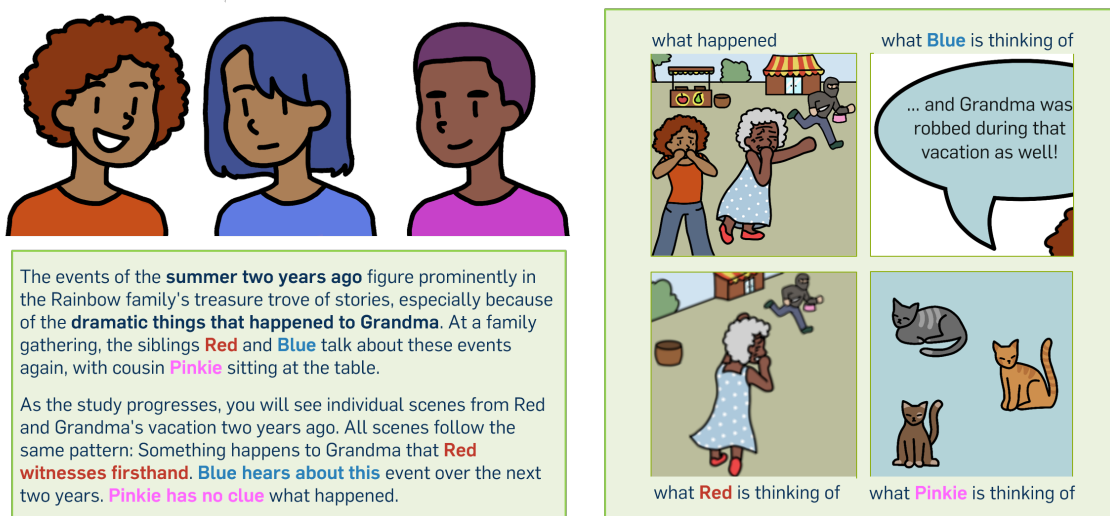


Figure 1: Composition of screenshots and text from the English version of the experiment

Against this background (presented in German), our main experiment targets the *dass/wie*-contrast in German memory reports with the verb *sich erinnern* ('REFL-remember'). Both the character uttering a sentence (RED, BLUE) and the complementizer (WIE, DASS) are manipulated variables. By combining the values of these variables, we obtain four target items for each scene:

- (3) a. Red sagt: Ich erinnere mich, dass Oma überfallen wurde.
 b. Blue sagt: Ich erinnere mich, dass Oma überfallen wurde.
 c. Red sagt: Ich erinnere mich, wie Oma überfallen wurde.
 d. Blue sagt: Ich erinnere mich, wie Oma überfallen wurde.
 Blue says: I remind REFL how Granny robbed was
 ‘{Red/Blue} says: I remember {that/how} Granny was robbed.’ (German)

Our main experiment is a Qualtrics online rating study that asks participants to judge sentences of this form against the background of a given scenario, consisting of ‘what happened’ and the speaker’s mnemonic perspective on it, as exemplified in Fig.1.³ Participants were asked to provide ratings on “Der grün hinterlegte Satz, von [SPEAKER] gesagt, beschreibt die Situation ...” – ‘The sentence marked in green, uttered by [SPEAKER], describes the situation ...’ – on a scale from 1 (*gar nicht richtig*, ‘not correctly at all’) to 7 (*völlig richtig*, ‘absolutely correctly’).⁴ Based on our background assumptions and literature-informed expectations (see Sect. 1), we formulated two hypotheses before preregistering our study (Rosina & Liefke 2024b) and then collecting results. Both hypotheses together would show that *wie* in ‘*sich erinnern*’-reports is an experientiality marker in the sense that it disambiguates for experiential memory in contrast to ‘*sich erinnern, dass*’.

- (4) a. **Hypothesis I:** Higher ratings for the RED+WIE than for the BLUE+WIE condition ✓***
 b. **Hypothesis II:** Higher ratings for BLUE+DASS than for BLUE+WIE ✓***

We recruited participants via Prolific and gathered data from 60 German mono-lingually raised native speakers aged 18–65, six of whom we excluded from analysis based on a control performance of $\leq 75\%$. The combination of our two manipulated variables (SPEAKER and COMPLEMENTIZER) resulted in four conditions, exemplified by (3). Alongside the robbery scenario, we used three more target scenarios: Grandma swimming and almost drowning, falling out of a canoe, and burning Red’s birthday cake. Hence, our study consisted of 16 target items, augmented with 16 control items.⁵ We tested within-subjects in order to facilitate a-posteriori reasoning. (E.g.: Are there two kinds of QUD-accommodators?) As a result, each participant provided four judgements per condition, resulting in 216 data points per condition. Hypothesis I was clearly confirmed with an extremely strong contrast (see Fig. 2; see Tab. 1 for interaction). Hypothesis II was also confirmed, but with a weaker contrast due to the lower-than-expected rating of BLUE+DASS.⁶

We consider these results evidence for *wie* as an experientiality-marker (in the semantics/

³A mock version of the German main study can be accessed via https://bochumpsyh.eu.qualtrics.com/jfe/form/SV_ezVoYefvfcY04Iu.

⁴We decided for these instructions and for this naming of the endpoints of our scale after comparing all similar experiments in the proceedings of ELM1 and ELM2. Terminology involving accuracy or ‘good fits’ tends to be more sensitive to pragmatic and socio-linguistic effects, while asking participants to grade truth has a non-trivial ontological flavor. Cf. Zhu & Ahn (2023) for the effect of instructions.

⁵False control sentences attribute Pinkie remembering of the events involving Grandma. True controls attribute Red perception of these events, or are unrelated true statements about Pinkie like ‘I am thinking about cats’ (as depicted by the Pinkie-picture, see Fig. 1). The target and control items are randomized within 8 blocks in order to prevent certain orders, and so is the order of the blocks.

⁶All analyses presented in this paper were done using Cumulative Link Mixed Effect Models fitted with the Laplace approximation, with participant and scene as random intercepts. (For motivation of the choice, see Liddell & Kruschke 2018); significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

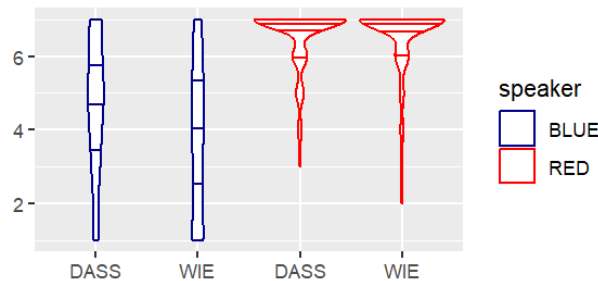


Figure 2: Ratings by condition and quartiles, main study

	Estimate	Std. Error	z-value	Pr(> z)
speaker:marker	0.9755	0.3014	3.237	0.00121 **
WIE:RED	4.2446	0.2557	16.597	<2e-16 ***
BLUE:WIE	-0.8313	0.1741	-4.775	1.8e-06 ***
RED:WIE	0.1442	0.2455	0.588	0.557

Table 1: Interactions of speaker and marker, main study

pragmatics neutral sense, see Sect. 1) in German ‘*sich erinnern*’-reports, in line with our semantics from (Rosina & Liefke 2024a), but not exclusively so. Importantly, the main experiment by itself does not provide any conclusive evidence on how lexically specific the effect is (either to the predicate *sich erinnern* or to *wie*-clauses). The experiments presented in Sect. 2.2 and 2.3 will suggest that evidentiality marking is indeed a cross-structural, cross-linguistic phenomenon.

That BLUE+DASS scored much lower than RED+WIE in our main study is a surprise: Since BLUE+WIE has even lower ratings than BLUE+DASS (so there is a huge main effect of BLUE), there are participants who do not grant Blue *any* kind of remembering even though she has reliable indirect evidence. A look at the individual participants’ widely distributed ratings of BLUE+DASS in Fig. 2 suggests a divide: While one group of participants is in line with our semantic-pragmatic explanation above, there is a second group whose members seem to have stricter conditions on memory. For members of this group, only experiential remembering is ‘real’ remembering in general – in the rating format of our main experiment, that is – see Sect. 3.1 for further discussion of this possible explanation and of pragmatic effects that might be intervening here. For now, note that we do not know yet what it is about Red that makes them a good speaker of WIE-sentences. It could be direct witnessing or particularly good evidence (as predicted by our semantics in Rosina & Liefke 2024a) at this point. We will come back to this issue in Sect. 3.3. Finally, note that the results leave room, in principle, to reason that the low BLUE+WIE ratings could instead be due to pragmatic competition with ‘*sich erinnern, dass*’ which could be preferred for independent reasons for BLUE. However, the most straightforward spell-out of such a pragmatic account would have to view *dass* as an indirectness marker. The high ratings of RED+DASS speak against such an account. Our speaker-ID study (see Sect. 3.1) will shed some more light on these considerations.

2.2. ENGLISH STUDY. Our study of experientiality marking through eventive uses of *how* has, until now, focused exclusively on German. To see whether this marking is cross-linguistically more robust (at least in a minimal sense), we have conducted a follow-up study that replicates the main

study for (American) English. The resulting English rating study with 27 participants (after exclusions) is a first hint that it might be quite robust.⁷ We did not preregister this study. It is mostly an English translation of half of the main experiment, but with the memory predicate *remember* and the hypothesized marker gerundive *-ing* small clauses ('GSC') instead of *wie/how*-clauses.⁸ These gerundive *-ing*-constructions are generally considered the prototypical way of reporting experiential remembering in English (see Stephenson 2010, Bernecker 2010). Bernecker (2010)'s claim that experientiality of rememberings is grammatically encoded refers to English GSC specifically. Importantly, Bernecker (2010) also holds the inverse of this claim, viz. that *that*-clauses are *non-experientiality/indirectness* markers. Our English study (i) provides evidence against this latter claim, and (ii) will show English GSC-constructions to have more-or-less the same effect as eventive *wie*-['how'] constructions in German. More complex than in the case of German *wie/dass*-clauses, *that*-clauses can differ from GSC in two ways besides the presence/absence of *that*: the *that*-clause can feature a past progressive or past simple verb form. The past progressive makes use of the *-ing* form like the GSC does, so it seems the better candidate for a minimal pair. On the other hand, some events may not be naturally reported with progressive aspect (leading to pragmatic disturbance),⁹ and the *-ing* form itself may turn out to be an experientiality marker, as opposed to the small clause character of GSC. For these reasons, we decided to include two versions of *that*-clauses for each GSC, leading to 'minimal triplets' (5) and six conditions in total.¹⁰

- (5) a. Red says: I remember that Grandma got robbed.
 b. Blue says: I remember that Grandma got robbed.
 c. Red says: I remember Grandma getting robbed.
 d. Blue says: I remember Grandma getting robbed.
 e. Red says: I remember that Grandma was getting robbed.
 f. Blue says: I remember that Grandma was getting robbed.

The participants judged these sentences against only two of the target scenes from the main study (robbery and swimming), such that they answered 12 target items and 12 controls, resulting in 54 analysed data points per condition. (This explains the lower significance of the effects relative to the main study.) The English instructions read as follows: "The sentence highlighted in green, said by [SPEAKER], describes the situation... 1 (not correctly at all) ... 7 (completely correctly)". We analyzed the pairings GSC/THAT-SIMPLE and GSC/THAT-ING separately and confirmed Hypotheses Ib and IIb in (6) when the *that*-clause uses past simple as in (5-a-b).

- (6) a. **Hypothesis Ib:** Higher ratings for RED+GSC than for BLUE+GSC

✓***

⁷Since the kind of non-manner uses of *how* that are relevant here seem more common in American English (see Liefke 2023), we conducted this study with monolingual-native English speakers currently living in the US. The only other screener on Prolific was age (18–65).

⁸A mock version of the English study can be accessed via https://bochumpsyh.eu.qualtrics.com/jfe/form/SV_afNNKtpBb9wVmvA.

⁹We thank Justin D'Ambrosio, p.c., for pointing us to this.

¹⁰We are aware of the in-principle syntactic ambiguity of the *-ing*-constructions in (5-c-d) between a GSC-constituent and a DP+adjunct (equivalent to 'I remember Grandma, as she was getting robbed'). The syntax of the construction we dub 'GSC' is irrelevant for its status as an experientiality marker for now, but will have to be considered for a future compositional account.

b. **Hypothesis IIb:** Higher ratings for BLUE+THAT-SIMPLE than for BLUE+GSC ✓*

The results of this second study show a significant interaction between MARKER and SPEAKER. The effects are not quite significant for ‘... that Grandma was getting robbed’ (5-e-f), suggesting that *-ing* itself contributes to experientiality/evidentiality. Comparing the results in Fig. 3 with the results of the main study in Fig. 2, the close parallel is evident.

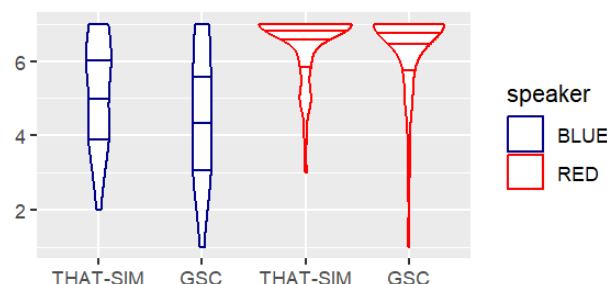


Figure 3: Ratings by condition and quartiles, English study

First, note that RED+THAT-SIMPLE has very high ratings, and there is no significant contrast with RED+GSC. (This part also holds for RED+THAT-ING.) This is the first clear empirical evidence against Bernecker (2010)’s claim that *that*-clauses mark indirectness of evidence/experience. The absence of any significant preference for RED+GSC/WIE over RED+THAT/DASS in the two studies is also interesting in the light of possible pragmatic competition: If both options are open for the direct experiencer Red, semantically, one might expect pragmatically decreased ratings for RED+THAT/DASS due to ‘maximize precision’, because Blue can use THAT/DASS as well, but not GSC/WIE (cf. Grice 1975). We observe no such effect in any of the rating studies,¹¹ and will return to this point when discussing the speaker-ID study in Sect. 3.1. A link with the BLUE+THAT/DASS puzzle suggests itself: If for whatever reason Blue cannot acceptably utter *any* memory report, the choice of the complement does not in fact maximize precision from Red’s perspective.

Coming back to the general picture, the results of the English study – including the puzzle on the lower-than-expected ratings for BLUE+DASS (now BLUE+THAT-SIMPLE) – closely resemble the German results from the main study. This suggests an at least minimal robustness of experientiality marking across languages (German and English), memory predicates (*sich erinnern* and *remember*), and complement structures marking experientiality (*wie/how*-clauses and GSC).

2.3. GERMAN ‘NOCH WISSEN’ STUDY. This case becomes even stronger when we consider the results of a third study that only differs from the German main study in the following respects:

(i) Most importantly, the matrix predicate is *noch wissen* [‘still know’] instead of *sich erinnern* [‘REFL-remember’]. Investigating the parallel or different distribution of these two predicates is motivated by the considerations discussed in Sect. 1 and our Rosina & Liefke (2024a), suggesting that NOCH+WISSEN could be the core of memory predicates in general. If *noch wissen* shows the same empirical pattern as *sich erinnern*, this motivates a closely related semantics, minimally concerning the selectional flexibility and the relation to *wie*-complements and experientiality.

(ii) This was a smaller-scale study, with 37 participants after exclusions and no preregistration.

¹¹This can also be attributed to our instructive formulations which were – it seems, successfully – aimed at truth-conditional effects as far as possible in an acceptability experiment, see fn. 4.

(iii) Like for the English study, we used only half of the scenes from the main study to minimize participation time for participants. This led to 8 target items (2 per condition) plus 8 controls and 74 data points per condition.

Our results confirm both hypotheses in (7) and significant interaction of MARKER and SPEAKER.

- (7) a. **Hypothesis Ic:** Higher ratings for RED+WIE than for BLUE+WIE ✓***
 b. **Hypothesis IIc:** Higher ratings for BLUE+DASS than for BLUE+WIE ✓***

Besides providing tentative support for our idea that our semantics for *noch wissen* in Rosina & Liefke (2024a) may be generalized, the combined results of the three experiments presented so far target the relation between language, evidence, and cognition at its core. A first, modest conclusion is that an account of ambiguity or polysemy of individual memory verbs has become extremely unattractive in the light of the (weakly, so far) cross-linguistic and cross-predicate (also within German) picture, and we should aim at finding a core, unified semantics of memory reports. The idea would be to give *remember*-equivalents a unified semantics that gives a different output in terms of use-conditions for *wie*- and GSC-complements on the one hand and *dass/that*-complements on the other hand. Our Rosina & Liefke (2024a) achieves this in a compositional manner for the case of *wie/dass*. To extend our coverage to the English results, GSC would have to be analyzed as informationally rich propositions like *wie*-complements.

The above considerations suggest a more far-reaching – but also more speculative – parallel with other attitude predicates and with perception predicates. The first part of this parallel lies in the presence of a clear relation between *noch wissen* and *wissen* [‘know’], such that (even experiential) remembering is a kind of knowledge (cf. Hörl 2022). The effect of informational richness on evidence is present also in the case of present-time imaginative knowledge. The speaker must be constructing an informationally rich scenario in their mind, *and* have very good (most likely, experiential) evidence for ‘the way things are’ in the neighbors’ house at the time of uttering (8-b).

- (8) a. Having watched the neighbors snap at each other and storm into their house:
 b. Ich weiß (genau/schon/ja/#noch), wie die jetzt (wieder) aufeinander losgehen.
 I know (exactly/PART/PART/still) how these now (again) at-each-other attack
 ‘I know (exactly) how these two are fighting (again) right now.’ R&L (2024a) ex. (14)

Davis & Landau (2021) investigate a similar effect for GSC under perception verbs. Perhaps this general interaction of informational richness, experience, and evidence can be described in a way uniform across these very different predicates. For an account in the spirit of our Rosina & Liefke (2024a), they would all have to share an evidential core which interacts pragmatically with informational richness (which is, again, encoded by different structures) in a way that increases the required quality of evidence the more informationally rich the object of this evidence is. This constitutes further support that cognitive concepts of evidentiality ‘precede’ its natural language realizations (Ünal & Papafragou 2018), but in a way that does not require any clear categorisation of direct vs. indirect evidence, since it is only about *good* vs. *bad* evidence. We will return to the issue of the gradient nature of evidence in Sect. 3.3, but leave the general discussion of how broad the phenomenon is open for future work.

3. Experiments on pragmatic effects. We now return to another route of open issues from Sect. 1 and the main experiment discussed in Sect. 2.1 that have only been strengthened by the English study and the *noch wissen* study. Remember that we could not conclusively distinguish between truth-conditional and use-conditional effects. In particular, RED+THAT/DASS was rated significantly higher than BLUE+THAT/DASS, due to an extremely significant main effect of SPEAKER. For some participants, the indirect experiencer Blue seems to be granted no remembering at all. Three complementary studies address these issues.

3.1. SPEAKER-ID STUDY. The set-up and the phrasing of the five experiments were aimed at truth-conditional semantics (for the influence of instructive formulations on results, see Zhu & Ahn 2023). In an attempt to control for pragmatic competition, we ran a smaller, non-preregistered experiment in another format with 29 German-native participants after exclusions and four target plus four control items.¹² Since this format only has two conditions, this amounts to 58 data points per condition. In this speaker-identification format (inspired by Davis & Landau 2021), participants choose “Wer sagt [SENTENCE]? – Red - Blue - Pinkie” (‘Who says [SENTENCE]? – Red, Blue or the control character Pinkie’)¹³ for each of the sentences in (9) presented in a given scene (cf. Fig.1).

- (9) a. Ich weiß noch, dass Oma überfallen wurde.
 b. Ich weiß noch, wie Oma überfallen wurde.
 I know still how Granny robbed was
 ‘I remember {that/how} Granny was robbed.’
 c. Ich weiß noch, dass Oma im Meer geschwommen ist.
 d. Ich weiß noch, wie Oma im Meer geschwommen ist.
 I know still how Granny in-the sea swim is
 ‘I remember {that/how} Granny was swimming in the sea.’ (German)

Note that only the complementizer varies between (9-a/c) and (9-b/d), since the character is now selected instead of given. This leads to a 2x2 setup with the speaker as the dependent variable and the complementizer the only manipulated variable. We forced participants to decide for exactly one character, instructing them to choose the one who is more likely to have uttered the sentence, if more than one or none of them could have said it. The results confirm both hypotheses in (10):

- (10) a. **Hypothesis i:** RED is selected more often given the WIE-condition ✓ 84%
 b. **Hypothesis ii:** BLUE is selected more often given the DASS-condition ✓ 64%

Like the results of the rating studies, these provide evidence for some version of the claim that German *wie*-complements mark experientiality. Two more things are noteworthy about the speaker-ID results: First, the effect of BLUE>RED for the DASS-condition is weaker than the effect of RED>BLUE for the WIE-condition. Since truth-conditions show stronger effects in experiments than competitions do, this supports the status of *wie* as a *semantic* marker (of direct experience or good evidence, see Sect. 3.3) in opposition to *dass*, which is not an indirectness marker, semanti-

¹²A mock version of the German speaker-ID study can be accessed via https://bochumpsych.eu.qualtrics.com/jfe/form/SV_5j8cj1JehaAk5M2.

¹³No participant selected Pinkie for any of our target items, which facilitates the 2x2-analysis.

cally, but is assigned more often to Blue because of pragmatic competition: Blue does not have any other way to express her remembering, because the *wie*-sentence is excluded qua truth-conditions; Red could have used the *wie*-sentence instead to maximize precision.

While this explanation is intuitively plausible when considered in isolation, it is in tension with what we observed for the rating studies. Even the confirmation of Hypothesis ii is unexpected relative to the results of all five rating studies: The preference for BLUE > RED in the DASS-condition clashes with significantly higher ratings for RED + DASS than for BLUE + DASS. If participants are presented with the German equivalent of the sentence ‘I remember that Grandma was swimming in the sea’, they choose Blue as a speaker. However, if we ask participants about this very sentence uttered by Blue, they rate it lower than a version uttered by the direct experiencer Red. Remember that after the main experiment, we hypothesised that many participants have very strict conditions of remembering, which exclude all cases but experiential remembering, regardless of the complementizer. However, if experiential remembering was just always ‘the real’ memory, we wouldn’t expect any preference for BLUE in the DASS-condition in the speaker-ID Experiment, but rather a similarly strong preference for RED in both conditions.

Directly after analyzing the speaker-ID results, we suspected that the forced choice design gives rise to the pragmatic competition we intended while our judgement scale design is more sensitive to the accommodation of different Questions Under Discussion (QUD; simplifying: the purpose of the conversation). This could have been a weakness of the general design of our experimental paradigm: That the grandchildren are said to exchange stories of the old time (see Fig. 1) might lead some people to accommodate a QUD like ‘Who was there when that happened?’. These considerations motivated the study presented in the next section.

3.2. QUD-MANIPULATED STUDY. In attempt to test for a possible effect of the QUD on the results of the rating studies, we contrasted the main experiment with a version that introduces a fact-based QUD like ‘Who knows the most facts about Grandma?’ (preregistered: Rosina & Liefke 2024d). In order to enable this comparison, the target items themselves were exactly the same as in the main experiment, with *sich erinnern* as the matrix predicate and all 16 target and 16 control items. We made only the following changes:

(i) We changed the background story such that it keeps the characters, their experiences, and the family constellation constant, but now introduces a quiz context instead of an informal conversation at a family gathering. The goal of the quiz game is to utter as many true facts about Grandma as possible. The reasoning behind this was that the original background story could have made the QUD ‘Who had which direct experiences with Grandma?’ salient, leading to the speaker main effect and the impression that BLUE does not have good memory at all, compared to RED, even in the fact-only DASS case. While the target items remained unchanged in every other respect, a reminder that the teenagers are participating in this kind of quiz was displayed with all items.

(ii) We replaced all of the original control items (see fn. 5 for the nature of the original ones), introducing completely new scenes and pictures. The new controls were aimed to support the fact-based QUD by having as little connection to the teenagers’ personal past as possible. Examples of this are sentences about Grandma’s tattoo, previous jobs, or the color of her motorcycle.

(iii) We recruited only 40 participants and excluded 2 of them, so the conditions QUD-FACT and QUD-ORIGINAL are unevenly distributed in the pooled analysis.

The original Hypotheses I and II were reproduced with the new background story (as Id and IId below) without any noticeable differences between the two studies. To our surprise, Hypothesis III, which had assumed an effect of the QUD-manipulation, was falsified.

- (11) a. **Hypothesis Id:** Higher ratings for RED+WIE than for BLUE+WIE ✓***
- b. **Hypothesis IId:** Higher ratings for BLUE+DASS than for BLUE+WIE ✓***
- c. **Hypothesis III:** Significant reduction of the main effect RED>BLUE in the QUD-FACT condition of a pooled analysis with the results of the main study ✗

The fact that the QUD does not effect any difference between the formats has two consequences for our theory formation: First, having excluded the first candidate, we have to keep looking for the reason for the difference between the scale judgement format and the speaker-ID format with respect to the case of BLUE+DASS. We suspect that the two formats support/block different pragmatic competitions, an observation that may have far-reaching consequences for methodology at the semantics-pragmatics interface in general, and that is discussed a bit more in Rosina (2024). Second, we are supported in our original experimental design and conclude that the background story was not a disturbing factor in any sense in the other experiments.

3.3. INTERMEDIATE EVIDENCE STUDY. Concluding the series of experiments, we designed a version of the main study (preregistered: Rosina & Liefke 2024c) with 36 German participants after exclusions. This ‘intermediate evidence’-study aims to distinguish between accounts of experientiality markers that encode direct evidence or experience directly in the semantics (Stephenson 2010, Liefke & Werning 2024) and our Rosina & Liefke (2024a) account that locates experientiality at the level of pragmatics and claims that eventive *wie* [‘how’] in isolation is – semantically speaking – only a marker of informational richness. In Rosina & Liefke (2024a), we show that when eventive *wie* is embedded under predicates of knowledge and remembering, the combined semantics marks *good enough* evidence (regardless of the kind of evidence).

Like in the main study, the memory predicate in this study is *sich erinnern* [‘REFL-remember’], but we used only two target scenes (here: Grandma swimming in the sea and burning a cake), like in the ‘*noch wissen*’ study. The most important change we made for this study is that we extended our cast of characters (see ex. (2) and Fig. 1) by one additional character, Red and Blue’s cousin Goldie. The idea behind Goldie is that she has evidence that lies between Red’s and Blue’s in terms of ‘quality’/reliability, but no direct experience. Specifically, to create this kind of evidence we told participants that Goldie always missed the things that happened to Grandma by a few minutes, but saw the immediate result (e.g. Grandma’s broken leg or smoke in the kitchen). Adding GOLDIE as a value of SPEAKER, and maintaining the values WIE and DASS for the MARKER variable as in all German rating studies results in six conditions with 72 data points each. Based on our results from the previous studies, we reasoned that ratings between Red’s and Blue’s for Goldie in the WIE-condition would confirm the gradability of the concept licensing experientiality/evidence marking. This is exactly what we found. There were highly significant main effects of RED>GOLDIE>BLUE, and the DASS/WIE-contrast shrinks ‘in the RED direction’, see our hypotheses in (12) and Fig. 4.

- (12) a. **Hypothesis Ie:** Higher ratings for RED+WIE than for BLUE+WIE ✓***
- b. **Hypothesis If:** Higher ratings for RED+WIE than for GOLDIE+WIE ✓***

- c. **Hypothesis Ig:** Higher ratings for GOLDIE+WIE than for BLUE+WIE ✓***
- d. **Hypothesis Iie:** Higher ratings for BLUE+DASS than for BLUE+WIE ✓***
- e. **Hypothesis IIf:** Higher ratings for GOLDIE+DASS than for GOLDIE+WIE ✗
 $Pr(> |z|) = 0.114$
- f. **Hypothesis IIg:** The effect confirming Iie is bigger than the effect confirming IIf ✓

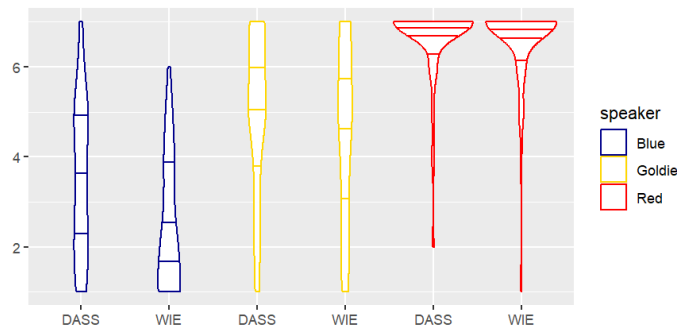


Figure 4: Ratings by condition and quartiles, ‘intermediate evidence’-study

These results provide support for our semantics in Rosina & Liefke (2024a) and against accounts that rely on direct experience or a specific *kind* of evidence to license experientiality markers in memory reports. However, proponents of such accounts could of course re-conceptualize their core concepts as gradient (e.g. based on a scale of directness of experience). We view our Rosina & Liefke (2024a) account as more straightforward, because it requires no mapping from kinds of evidence to a scale in the semantics. The relationship between evidence and informational richness on our account is built on one principle: the more informationally rich, the harder to be evidentially supported. (For more discussion of this, see Rosina 2024.)

4. Conclusion. Our studies confirm the common assumption that the way we report events reflects whether we have personally experienced or witnessed these events. We provide experimental support for two such markers of experientiality: German eventive *wie* [‘how’] and English gerundive *-ing* small clauses. Our studies further show that the effect of experientiality marking is not specific to any particular memory predicate, since the results for such marking under *sich erinnern*, *noch wissen* and *remember* are very similar. Relating these results to compositional accounts of memory reports, our ‘*noch wissen*’-study provides tentative evidence for STILL+KNOW as the core of memory predicates, as suggested in Rosina & Liefke (2024a). The results of our ‘intermediate evidence’-study support our idea that only informational richness and (quality of) evidence are semantically encoded, and the common requirement of direct experience is only a pragmatic effect.

Interestingly, even *dass/that*-sentences uttered by our indirect evidence character Blue receive relatively low ratings, compared to the same sentence uttered by our direct experiencer Red. Far from marking *non*-experientiality (Bernecker 2010), our rating studies suggest that the status of ‘remembering that’ as reporting (also) fact-only remembering is questionable. This is in tension with the results of our speaker-ID study, where participants chose the indirect experiencer to be more likely to have uttered the ‘*remember that*’-equivalent. We suspect that the two formats give rise to different kinds of pragmatic competition and leave the formalisation of this effect to future work.

References

- Bernecker, Sven. 2010. *Memory: A philosophical study*. Oxford: Oxford University Press.
- Davis, E. Emory & Barbara Landau. 2021. Seeing vs. seeing that: Children's understanding of direct perception and inference reports. In *Proceedings of ELM 1*, 125–135.
- Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry Morgan (eds.), *Speech acts [syntax and semantics 3]*, 41–58. Academic Press.
- Hörl, Christoph. 2022. A knowledge-first approach to episodic memory. *Synthese* 200(5).
- Legate, Julie Anne. 2010. On how 'how' is used instead of 'that'. *Natural Language and Linguistic Theory* 28(1). 121–134. 10.1007/s11049-010-9088-y. <https://doi.org/10.1007/s11049-010-9088-y>.
- Liddell, Torrin & John Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 328–348.
- Liefke, Kristina. 2023. Two kinds of English non-interrogative, non-manner *how*-complements. In Lukaszukasz Jedrzejowski & Carla Umbach (eds.), *Non-interrogative subordinate wh-clauses* Oxford Studies in Theoretical Linguistics, 24–62. Oxford UP.
- Liefke, Kristina & Markus Werning. 2024. Diachronicity matters! how semantics supports discontinuism about remembering and imagining. *Topoi* 1–23. 10.1007/s11245-024-10068-1.
- Rosina, Emil E. 2024. Semantic experiments on 'remember' as Gettier cases of memory. Draft. <https://lingbuzz.net/lingbuzz/008328/>.
- Rosina, Emil E. & Kristina Liefke. 2024a. German 'noch genau wissen': Uniform semantics, distinct effects. *accepted for the Proceedings of WCCFL*.
- Rosina, Emil E. & Kristina Liefke. 2024b. German 'wie' in memory reports. 10.17605/OSF.IO/V5PEH. OSF preregistration.
- Rosina, Emil E. & Kristina Liefke. 2024c. German 'wie' in memory reports - intermediate evidence follow-up. 10.17605/OSF.IO/BM38W. OSF preregistration.
- Rosina, Emil E. & Kristina Liefke. 2024d. German 'wie' in memory reports - QUD manipulation. 10.17605/OSF.IO/9BV3K. OSF preregistration.
- Sadock, Jerrold M. & Arnold M. Zwicky. 1975. Ambiguity tests and how to fail them. In J. Kimball (ed.), *Syntax and semantics*, vol. 4, 1–36. New York.
- Stephenson, Tamina. 2010. Vivid attitudes: Centered situations in the semantics of *remember* and *imagine*. *Semantics and Linguistic Theory (SALT)* 20. 147–160. 10.3765/salt.v0i20.2582.
- Tulving, Endel. 1972. Episodic and semantic memory. In Endel Tulving & W. Donaldson (eds.), *Organization of memory*, 381–402. New York: Academic Press.
- Umbach, Carla, Stefan Hinterwimmer & Helmar Gust. 2022. German *wie*-complements: Manners, methods and events in progress. *Natural Language and Linguistic Theory* 40. 307–343. 10.1007/s11049-021-09508-z.
- Zhu, Ziling & Dorothy Ahn. 2023. Effects of instruction on semantic and pragmatic judgment tasks. In *Proceedings of ELM 2*, 322–330.
- Ünal, Ercenur & Anna Papafragou. 2018. Relations between language and cognition: Evidentiality and sources of knowledge. *Topics in Cognitive Science* 12. 10.1111/tops.12355.