

Experimental findings for a cross-modal account of dynamic binding in gesture-speech interaction

Kurt Erbach, Magnus Poppe, & Cornelia Ebert*

Abstract. We report results of experiments on pronoun and presupposition binding across modalities. We show that ordinary pronouns (in the spoken/written domain) can be dynamically bound to gesturally introduced discourse referents and that presuppositions induced by spoken/written triggers (via e.g. ‘again’ or ‘too’) can be bound likewise. These experiments support research that has proposed the existence of cross-modal binding to motivate a formal framework that can account for interaction of various input of linguistic content from different dimensions and modalities.

Keywords. multimodal; gesture; anaphora; pronouns; presupposition

1. Introduction. Ebert et al. (2020) (based on Ebert & Ebert 2014) suggest a formal framework for gesture semantics where certain iconic and pointing co-speech gestures introduce discourse referents that can serve as antecedents in anaphoric reference. Crucially, this necessitates a uni-dimensional dynamic system that allows for binding effects across dimensions and, in this case, modalities. Ebert et al. (2020) argue that co-speech gestures contribute propositional non-at-issue information by default. The further argue that this information arises as ‘constructional’ meaning due to the co-occurrence of gesture and speech. In their dynamic semantic framework, gestures introduce discourse referents for rigid designators as their core ‘lexical’ meaning: when a pointing gesture is performed this triggers the introduction of a discourse referent that is identified with the rigid concept of the gesture referent, and can be anaphorically picked up across dimensions—e.g. by a pronoun.

While the introduction of discourse referents by gesture has been claimed and implemented in the formal system of Ebert et al. (2020), this has not been experimentally demonstrated. Here we show that dynamic binding across dimensions can be made with respect to both pronouns and presupposition triggers like *again*. In the constructed example (1-a), where both cake and cookies are contextually salient, the pointing co-speech gesture, extending an index-finger towards the cake, is assumed to introduce a discourse referent for the gesture concept, resulting in an interpretation somewhat like ‘Have you eaten (some) cake?’. The discourse referent for the gesture concept is assumed to be able to bind to pronouns like *it* in the hypothetical response (1-b). On the other hand, if (1-a) had included a hand-over-stomach gesture to indicate being full (1-c) and crucially not introducing some food as a discourse referent, then presumably responding up with (1-b) would be infelicitous as it is unclear what *it* is supposed to be bound to. Moreover, if the question with the pointed-to referent (1-a) were responded to in such a way that answers the question albeit ignoring the referent (1-d), then this too was predicted to be marked under the assumption that the gesture referent is introduced, but not picked up. Yet, one would expect this case to be not as bad as an utterance with a pronoun that cannot be resolved to any referent.

*We would like to thank the organizers, reviewers, and audiences of ELM3, Linguistic Evidence 2024, and XPRAGit.2024 for their feedback on this project and the opportunity to present our work. Authors: Kurt Erbach, Goethe-University Frankfurt, Saarland University (erbach@hhu.de); Magnus Poppe, Goethe-University Frankfurt; Cornelia Ebert, Goethe-University Frankfurt (ebert@lingua.uni-frankfurt.de).

- (1) a. Have you eaten_{pointing to cake}?
 b. It was too sweet for me.
 c. Have you eaten_{hand over stomach}?
 d. Yeah, a few too many cookies.

Similarly, the jogging gesture in (2-a) is assumed to add the propositional content that Paul was jogging (when the speaker met him). If (2-b) is said in response to (2-a), it is assumed that the presupposition that is triggered by *again* in (2-b), namely that Paul went jogging before, is satisfied by the the propositional information about jogging given in the visual modality in (2-a). If (2-b) was a response to (2-c), the absence of a jogging gesture would mean the presupposition triggered by *again* would not be satisfied, at least under the assumption that people don't commonly meet while jogging and hence such a proposition cannot be accommodated. Conversely, a follow up like (2-d) is presumably odd following a jogging gesture (2-a) under the assumption that people do not jog in cafés, but following (2-c) ought to be fine assuming people often meet in cafés.

- (2) a. Yesterday I met Paul_{jogging gesture}?
 b. He went jogging again today.
 c. Yesterday I met Paul_{pointing behind self}?
 d. Was it in the café again?

The paradigms in (1) and (2) exemplify data that are used in our experiments by which we seek to support the idea that gesture can introduce discourse referents that can be picked up by anaphora. In this paper, we describe a series of experiments that test this idea via these paradigms and that ultimately show the sorts of acceptability contrasts that are predicted. In Section 2, we describe the design of our experiments, in Section 3 we report the results, and in Section 4 we describe a follow-up experiment and its results showing that the acceptability of pronoun binding with pointed-to referents depends on grammatical gender agreement. In Section 5, we discuss the results of the experiments, namely that these experiments can be taken as evidence of dynamic binding, wherein verbal anaphora can be bound to content introduced by gestures.

2. Experiment design. Two experiments were designed in German to test the contrasts demonstrated in (1) and (2). The complementary designs allowed each to be used as filler for the other. Both experiments had two factors each with two levels—GESTURE (pointing/non-iconic or iconic/non-pointing) and PRONOUN or PRESUPPOSITION (present or absent)—yielding four treatments, each of which was predicted to be judged as either felicitous (accept) or infelicitous (reject). The four conditions for the respective pronoun and presupposition binding experiments are listed in Tables 1 and 2.

GESTURE	PRONOUN	Prediction	Examples
pointing	present	felicitous	(1-a)+(1-b)
pointing	absent	infelicitous	(1-a)+(1-d)
non-pointing	present	infelicitous	(1-c)+(1-b)
non-pointing	absent	felicitous	(1-c)+(1-d)

Table 1: Conditions for pronoun-binding experiment

10 minimal pairs resembling (1) and (2) were distributed across lists for a 2x2 Latin square

GESTURE	PRESUPPOSITION	Factor level	Examples
iconic	present	felicitous	(2-a)+(2-b)
iconic	absent	infelicitous	(2-a)+(2-d)
non-iconic	present	infelicitous	(2-c)+(2-b)
non-iconic	absent	felicitous	(2-c)+(2-d)

Table 2: Conditions for presupposition-binding experiment

design (i.e. 10 items per list), and we recruited 80 native German speaking participants via Pro-lific (20 per list), so there were 200 responses per condition¹. In a variation of the covered-box task (cf. Fanselow et al. 2022), the sentence pairs were presented with the context, e.g. (1-c) in video form, and the follow-up, e.g. (1-b), being in written form as one choice in a pair of alternatives, the other being ‘covered’ (lit. “[geschwärzt]” (‘redacted’)). Participants were instructed that only one of the alternatives was a reasonable follow-up to the context, and they should select the reasonable response. All experimental materials can be found in the repository at <https://osf.io/3hvyt/>, along with the results and R code used for analysis.

Prior to performing the actual experiment, each participant performed four practice trials to get familiar with the design. The first two practice trials included instructions but no requirement to select the answer as predicted. The latter two practice trials included no instructions. The practice items followed the same basic formula albeit not with respect to pronoun reference and with different gestures and sentences in each utterance.

To control for quality, four attention-check questions were added using the same videos as the practice trials. Underneath the video was a statement that this question is designed to check that the participant is not a bot—following Silber et al. (2022), who showed that such statements increase accuracy in attention check questions—and the question asked something about the appearance of the speaker in the video—e.g. ‘Is the speaker wearing glasses?’. Answers from the participants that got more than one attention-check question wrong were discarded. The experiments were preregistered on OSF at <https://osf.io/35jbw>.

3. Results. One participant each in groups 1 and 3 failed more than one attention check test as did two participants in group 4. These participants’ answers were all removed from the results as well as those of randomly selected participants in groups 1, 2, and 3 such that each group had the same number of participants remaining (n=18). Therefore, 180 responses were analyzed per condition (cf. the originally collected 200 responses per condition).

3.1. PRONOUN BINDING EXPERIMENT RESULTS. The results for the pronoun experiment are summarized in Figure 1. In this experiment, follow-up sentences with pronouns to be bound to the discourse referent introduced with a pointing gesture (1-a)+(1-b) were largely accepted (n=115), and, surprisingly, those without such a pronoun (1-a)+(1-d) were accepted nearly as much (n=105). With non-pointing gesture contexts, follow-ups with pronouns to bind (1-c)+(1-b) were not accepted (n=63) unlike those with non-pointing continuations ((1-c)+(1-d), n=133).

¹Because there were a total of 10 items per condition, there was an uneven distribution of conditions across lists—e.g. three pointing–present items in List 2, two pointing–present items in List 3, two non-pointing–present items in List 2, three non-pointing–present items in List 3, etc.

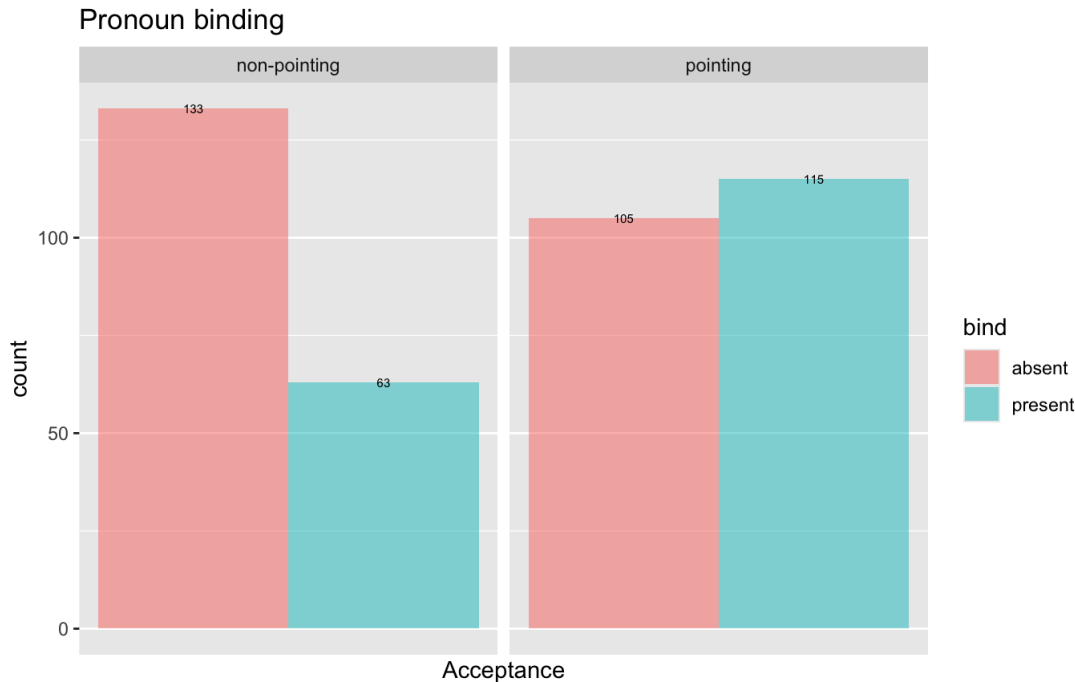


Figure 1: Number of follow-up acceptances in pronoun experiment; max-n = 180

To understand whether or not there is an effect of pointing in the resolution of pronoun binding with otherwise unspecified discourse referents—i.e. to understand if there is a difference between (1-a)+(1-b) and (1-c)+(1-b)—the results were analyzed with a mixed-effects model with binomial error distribution constructed in R (R Core Team 2015) using the lmerTest package (Kuznetsova et al. 2019). The maximal random effects structure justified by our experimental design should include a by-participant random effect, but this model did not converge, so we simplified the random effects structure by not allowing correlation between the random slope and random intercept for subjects (Barr et al. 2013). The full model translates to: $\text{response} \sim \text{bind} * \text{gesture} + (1|\text{questionID})$. The summary of the model is in Table 3; the main effect of bind–gesture interaction is statistically significant ($p = 0.0026$). To summarize, there is a very low likelihood that we would see these

Fixed effects	Estimate	SE	z	p
Intercept	1.3091	0.3988	3.283	0.00103
bind	-2.0635	0.5567	-3.707	0.00021
gesture	-0.8904	0.5516	-1.614	0.10646
bind:gesture	2.3420	0.7778	3.011	0.00260
Random effects	Name	Variance	Std.Dev.	
questionID	(Intercept)	1.172	1.083	

Table 3: Summary of glmer model for pronoun experiment

results if there was no interaction between binding pronouns and gesture. We take this as support of the proposal that discourse referents can be introduced with pointing gestures, and that pronouns

can then be bound to these discourse referents. And crucially, if no pointing occurred during an utterance in a context where possible discourse referents are not otherwise clear, then a follow-up with a to-be-bound pronoun is likely to be rejected.

3.2. PRESUPPOSITION BINDING EXPERIMENT RESULTS. The results for the presupposition experiment are summarized in Figure 2. In this experiment, follow-up sentences with presupposi-

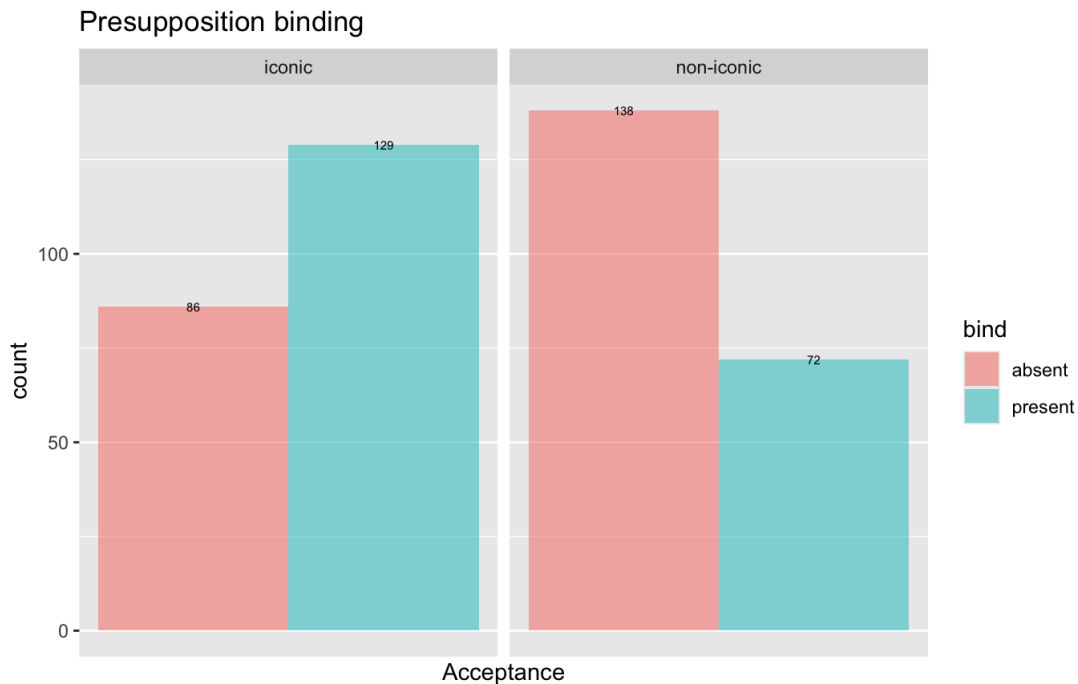


Figure 2: Number of follow-up acceptances in presupposition experiment; max-n = 180

tion triggers like *again* to be bound to the propositional referent introduced with an iconic gesture introducing some propositional content (2-a)+(2-b) were largely accepted (n=129) while the same follow-ups were rejected (n=72) if the gesture did not introduce such a propositional referent (2-c)+(2-b). Our control items were also judged as predicted: when iconic gesture contexts were followed with a statement that did not refer to the gesture, rather it contained an item that clashed with the gesture, (2-a)+(2-d), these follow-ups were rejected (n=86), but when a context gesture did not introduce a discourse referent (2-c)+(2-d), the same follow-ups were accepted (n=138).

To understand whether or not there is an effect of iconic gestures in the resolution of presupposition trigger binding with otherwise unspecified discourse referents—i.e. to understand if there is a difference between (2-a)+(2-b) and (2-c)+(2-b)—the results were analyzed with a mixed-effects model with binomial error distribution constructed in R (R Core Team 2015) using the lmerTest package (Kuznetsova et al. 2019). The maximal random effects structure justified by our experimental design translates to: $\text{response} \sim \text{bind} * \text{gesture} + (1|\text{questionID}) + (1|\text{participant})$. The summary of the model is in Table 4; the main effect of bind–gesture interaction is statistically significant ($p = 0.00016$). To summarize, there is a very low likelihood that we would see these results if there was no interaction between binding presupposition triggers and to propositional referents introduced with an iconic gesture. We take this to support the claim that propositional

Fixed effects	Estimate	SE	z	p
Intercept	-0.3056	0.4885	-0.626	0.53152
bind	1.4737	0.6809	2.164	0.03045
gesture	1.8712	0.6891	2.715	0.00662
bind:gesture	-3.6653	0.9710	-3.775	0.00016
Random effects	Name	Variance	Std.Dev.	
participant	(Intercept)	0.353	0.5941	
questionID	(Intercept)	1.912	1.3829	

Table 4: Summary of glmer model for presupposition experiment

discourse referents can be introduced with iconic gestures, and that presupposition triggers can then be bound to these discourse referents. Moreover, if no iconic gesture occurred during an utterance in a context where possible discourse referents are not otherwise clear, then a follow-up with a to-be-bound presupposition trigger is likely to be rejected.

3.3. INTERIM DISCUSSION. In general, the results can be interpreted as showing that participants accept follow-ups with pronouns and presupposition triggers that can be bound to discourse items introduced with pointing and iconic gestures respectively more readily than when there are no possible binders for these pronouns and presuppositions. In both experiments these items were accepted more than 50% ($n=90$) of the time: pronouns ($n=115$); presupposition triggers ($n=129$). It may be worth noting that these rates of acceptance are closer to random acceptance than complete acceptance (100%, $n=180$), which may be due to a number of reasons: First, the entire experimental paradigm is a bit unnatural in that the contexts were presented in video format and the follow-ups were presented not only in a written format, but also in a covered box task, both of which may have led to fewer than expected acceptances of follow-up items. This possibility might be supported by the fact that even our “acceptable” contrast items (i.e. non-pointing (1-c) + item-to-bind-absent (1-d) for the pronoun experiment, and non-iconic (2-c) + item-to-bind-absent (2-d) for the presupposition trigger experiment) had acceptance rates closer to random choice than complete acceptance: $n=133$ in the pronoun experiment and $n=135$ in the presupposition trigger experiment. Another reason for low acceptance might be the covered-box paradigm, where participants were invited to imagine that one of the two follow-ups was definitely acceptable and the other definitely unacceptable, and only one of the two was shown. This may have contributed to the given acceptance rates because participants may have imagined even more acceptable responses than the ones given in certain cases.

Another notable result is the fact that, in the pronoun experiment, follow-up items that ignored the discourse referent introduced by pointing, (1-a)+(1-d), were accepted more than half the time ($n=105$, 58.33%). While this was the closest count to random acceptance (50%), two of the other three conditions predicted to be marked were similarly close to random acceptance: for GESTURE-iconic + PRESUPPOSITION-absent, (2-a) + (2-d), follow-ups were rejected 47.77% of the time ($n=86$), and for the condition GESTURE-non-iconic+PRESUPPOSITION-present, (2-c)+(2-b), follow-ups were rejected 40% of the time ($n=72$). This too could speak to the issues with the experimental design discussed above. A final alternative may be that, as mentioned above, it simply

might be natural to ignore a discourse referent introduced with a co-speech pointing gesture in the scope of a question, and for sure not as bad as having to interpret a to-be bound pronoun that has no antecedent.

4. Pronoun binding follow-up experiment. We conducted a follow-up to the pronoun experiment to test for an effect of pronoun agreement when discourse pronouns in follow-up sentences are meant to bind to discourse referents that are not lexically specified but possibly pointed to with a pointing gesture. The data from the previous pronoun binding study consisted of 10 items testing pronoun resolution and gesture, 8 of which contained pronouns that agreed only with the unspoken noun that could refer to the object pointed to as opposed to also agreeing with the other item visible in the context. For example, with the question “Hast du schon gegessen?” (‘Have you already eaten’) a piece of cake was pointed to during ‘already eaten’; note a package of cookies was also visible in this context. The participants were asked to choose between a pair of responses, one visible, one not, and in the PRONOUN-present condition, the visible response was “Er war zu süß für mich.” (‘It.M was too sweet for me’). Crucially, the German noun *Kuchen* (‘cake.M’) has masculine gender which could be used to bind the pronoun in the follow-up question, while the alternative object, which was not pointed to, a package of cookies (*Kekse* ‘cookies’) did not agree in gender with the pronoun in the visible response. Because such a bias existed in 80% of the target items, if pronoun agreement with an unspoken noun was sufficient without pointing, then we might have expected that the visible responses were selected for that 80% of control items. Instead, what we saw is that visible responses were selected for only 35% (n=63) of the control items where a pronoun agreed in gender with one of the visible items that were not pointed to. In contrast, when pointing was included in the context, the visible responses were selected 63.89% (n=115) of the time.

While the inferential statistics for the previous study found that no interaction between GESTURE and PRONOUN-to-bind is unlikely ($p = 0.0026$), strictly speaking, the interaction of gesture and pronoun agreement has not been tested. In other words, while it seems that the presence of an object that, when named, would agree with the pronoun in question is not sufficient to accept statements that contain pronouns meant to bind with such objects, it is unclear if it is the pointing gesture alone or gesture+agreement interaction that drives the acceptance of follow-ups in the gesture+pronoun-to-bind condition. By looking at pronoun agreement and gesture interaction, the present study provides the opportunity to see how participants behave when the visible responses contain a pronoun that is unlikely to be bound to an unspoken noun referring to one of the objects. In other words, one of the conditions in the covered box paradigm will contain visible responses that do not agree in gender with the visible objects in the context that can be referred to with nouns that could serve as an antecedent for said pronoun.

By minimally modifying the design of the previous study, the following paradigm (3) and conditions (Table 5) were developed to test the interaction of gesture and agreement.

- (3) a. Hast du schon gegessen? (‘Have you eaten_{pointing to cake.M}?’)
 b. Ja. Der war aber zu süß für mich. (‘Yes, but it_{DPRO.M.SG} was too sweet for me’).
 c. Hast du schon gegessen? (‘Have you eaten_{hand over stomach}?’)
 d. Ja. Die war aber zu süß für mich. (‘Yes, but it_{DPRO.F.SG} was too sweet for me’).

The first condition in Table 5 is the same as the previous pronoun binding experiment though we switched to using D-pronouns because these have been shown to be preferable in spoken modality (Patil et al. 2023). The second condition tests whether a pronoun that does not agree with the pointed-to item is acceptable: in (3-d), the pronoun *die* has feminine grammatical gender and therefore does not agree with *Kuchen* (‘cake.M’), which is the unspoken noun denoting the object pointed to in (3-a). The third condition is again the same as in the previous experiment albeit with the D-pronoun, and the fourth condition completes the set by testing to see whether a non-agreeing pronoun is acceptable when no specific item has been mentioned. According to the results of our previous study, this should also be rejected as there is no item mentioned to agree with. Apart

GESTURE	AGREEMENT	Factor level	Examples
pointing	agree	felicitous	(3-a)+(3-b)
pointing	disagree	infelicitous	(3-a)+(3-d)
non-pointing	agree	infelicitous	(3-c)+(3-b)
non-pointing	disagree	infelicitous	(3-c)+(3-d)

Table 5: Conditions for pronoun-agreement experiment

from these modifications, the follow-up experiment followed the same design as procedure as the previous experiments.

4.1. RESULTS OF PRONOUN AGREEMENT EXPERIMENT. All 80 participants answered at least 75% of the control items correctly so no answers had to be thrown out. The results for the pronoun agreement experiment are summarized in Figure 3. In relative terms, the results of this experiment

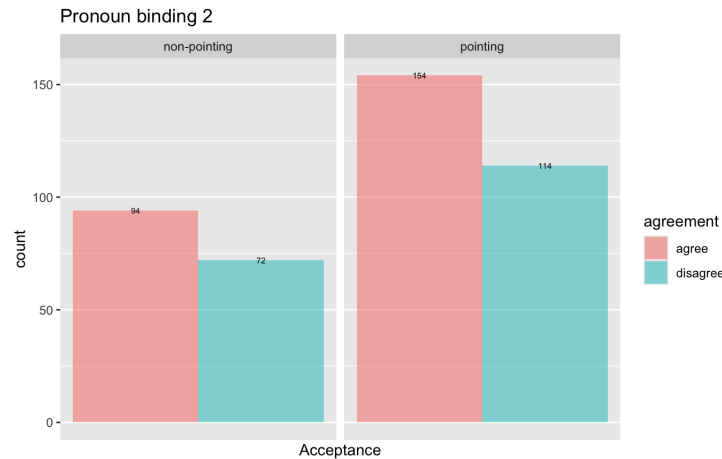


Figure 3: Number of follow-up acceptances in pronoun follow-up experiment; max-n = 200

followed the predictions in that the pointing–agree condition, (3-a)+(3-b), received the highest number of acceptances (n=154; 77%) making it clearly acceptable, as predicted, while all others were markedly less acceptable. In particular, the pointing–disagree condition, (3-a)+(3-d), received an above chance (50%) number of accepts (n=114; 57%), which we took to be acceptable in the previous experiments, and therefore deviates from our predictions in these terms. The non-pointing conditions both received a lower than chance number of accepts: non-pointing–agree n=94 (47%) and non-pointing–disagree n=72 (36%).

To analyze the effects of gesture and agreement on the acceptability of the follow-up items, the results were analyzed with a mixed-effects model as before. The maximal random effects structure justified by our experimental design translates to the following: $\text{response} \sim \text{agreement} * \text{gesture} + (1|\text{participant}) + (1|\text{questionID})$. The results of the model are summarized in Table 6. As seen

Fixed effects	Estimate	SE	z	p
(Intercept)	-0.7860	0.3769	-2.085	0.037
agreement	0.6206	0.5160	1.203	0.229
gesture	1.1294	0.5139	2.198	0.028
agreement:gesture	0.5603	0.7297	0.768	0.443
Random effects	Name	Variance	Std.Dev.	
participant	(Intercept)	0.5369	0.7327	
questionID	(Intercept)	1.0314	1.0156	

Table 6: Summary of glmer model for pronoun follow-up experiment

in Table 6, there is a main effect of gesture ($p=0.028$) but not of agreement ($p=0.229$) or gesture–agreement interaction ($p=0.443$). An ANOVA between this model against one without interaction between agreement and gesture does not show a significant difference ($p=0.4454$), however an ANOVA between the model without interaction between agreement and gesture and one only with gesture as a fixed effect—i.e. without agreement as a fixed effect at all—does show an effect of agreement ($p=0.0182$).

4.2. DISCUSSION OF AGREEMENT EXPERIMENT. We take the results of the ANOVA between linear models with and without agreement as evidence that there is an effect of agreement on the acceptance of follow-ups in the pronoun agreement experiment. In other words, while participants did accept pronouns at a rate above chance when they do not agree in gender with the unspoken names of objects visible in the context, such pronouns are considered marked to those that do agree.

Given that the linear model shows no effect of agreement (in contrast to the ANOVA) despite showing an effect of gesture, one might also be inclined to assume that the pointing gesture drives the agreement in these. This might be corroborated by the above-chance rate of acceptance in the pointing–disagree condition. Looking back at our stimuli, they generally also did not agree with the unspoken noun that would denote the object not pointed to, though it could have for a few of the items depending on exactly how the not-pointed-to object was interpreted. It also could have been the case that interpretations we did not anticipate were thought of by the participant—e.g. “die Torte” (‘the_F layer-cake’) instead of “der Kuchen” (‘the_M cake’). In short, though it is clear that the participants strongly preferred pronouns that agree with unspoken nouns denoting pointed-to objects to those that do not agree, further work is necessary to understand the acceptance rate of pronouns in the pointing–disagree condition.

5. Discussion. The results of the three experiments support the proposals by Ebert & Ebert (2014), Ebert et al. (2020), namely that gestures can introduce discourse referents that pronouns and pre-supposition triggers can bind to. In the first experiment, we showed that participants largely accepted the idea of using pronouns to refer to discourse referents introduced with pointing gestures,

in contrast to instances when pronouns were used to refer to objects visible in the context but were not pointed to. In the second experiment, we showed that participants largely accepted the idea of using presupposition triggers like *again* to refer to discourse referents introduced with iconic gestures, in contrast to instances when these referents were ignored, or when presupposition triggers did not have a clear referent to bind to. Finally, in the third experiment we showed that participants were more likely to accept pronouns that agree in gender with the unspoken noun of a pointed-to object than pronouns that disagree. In other words, we have seen the acceptability of anaphora binding to discourse referents introduced with gender in three distinct experiments.

Within and between the two pronoun experiments, we also saw a contrast between instances where pronouns are used without clear discourse referents in the preceding context and instances where they are. In short, the question is, why does pointing play such a key role when pronoun gender can otherwise disambiguate what is being referred to? Recall that, following pointing gestures, the referring to the pointed-to-but-unnamed object with a pronoun that agrees in gender with potential noun is accepted at a relatively high rate (63% in the first experiment and 77% in the follow-up). Meanwhile, the same sentence is generally rejected if the object is not pointed to in the preceding context (35% acceptance in the first experiment and 47% in the follow-up). Given the effect of agreement in the follow-up, why do participants reject pronouns at these rates? And at the same time, why do they accept pronouns that do not agree with nouns denoting discourse referents introduced by pointing at a relatively high rate (57%)?

We argue that the account of the semantics of gesture argued for in Ebert et al. (2020) (based on Ebert & Ebert 2014) can straightforwardly account for the acceptance rates of pronouns seen in these experiments. In particular, what is argued is that co-speech gestures (which were used in the above experiments) are interpreted as follows (4): the spoken lexical items introduce a predicate $N_p(x)$, where items subscripted by p are at-issue content and those subscripted by p^* are not-at-issue content (see AnderBois et al. 2015); the gesture introduces a discourse referent, z , which is equivalent to a rigid designator, I_g , which, applied to all worlds, yields the same designated gesture referent g ; and the temporal alignment between the lexical items and the gesture introduces a similarity predicate, $SIM_{p^*}(x, z)$, that applies to the argument of the lexical predicate, i.e. the concept introduced in speech, and the rigid gesture concept, and requires the two—what is introduced in speech and what is introduced via gesture—to be similar in certain contextually relevant respects.

$$(4) \quad [x] \wedge N_p(x) \wedge [z] \wedge z = I_g \wedge SIM_{p^*}(x, z)$$

What this means for the results of our study is that, while pointing introduces an object z that can be picked up in later discourse, exactly how this object is named in speech still remains unclear. The somewhat low rate of acceptance of follow-ups with pronouns agreeing with unspoken nouns denoting the pointed-to objects ($n=154$; best possible 200) fits with this analysis given some bridging must be assumed: since the pronoun is assumed to bind to the referent z that has been introduced by the gesture, the participants still have to find a spoken language expression for this referent. But there are still some uncertainties involved at this point—participants have to bridge the pronoun’s grammatical gender to some noun concept, which they have to provide themselves. In other words, the difference between accepting *der* (‘the.M’), versus *die* (‘the.F’), when the speaker in the video points to cake depends on whether or not participants decide to name and refer to the object pointed to as *Kuchen* (‘cake.M’).

This analysis also leaves room for pronouns that do not agree with the unspoken noun that denotes the object in question to nevertheless bind to these discourse referents. At the same time, while non-pointing gestures can be interpreted in the same way, the objects they would point to would differ from those referred to by the pronoun. For example, if the non-pointing gesture in (1-c) “points” to a state of being full or having eaten, then the pronoun in (1-b) could not bind to this object because a state of being full cannot be the argument of ‘It was too sweet for me.’ In other words, because the context did not provide any salient discourse referents for pronouns in items like (1-b), participants accepted these items less.

In summary, we have provided experimental evidence that supports the formal model in Ebert et al. (2020) in which gestures introduce discourse referents that pronouns and presupposition triggers like ‘again’ can bind to. We have also demonstrated the effect of gender on binding pronouns to discourse referents introduced by a pointing gesture. Lastly, we argued that the semantic analysis in Ebert et al. (2020) can account for the fact that pronouns that do not agree with nouns denoting pointed-to objects are nevertheless generally accepted.

References

- AnderBois, Scott, Adrian Brasoveanu & Robert Henderson. 2015. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32(1). 93–138.
- Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3). 255–278.
- Ebert, Christian, Cornelia Ebert & Robin Hörnig. 2020. Demonstratives as dimension shifters. In *Proceedings of sinn und bedeutung*, vol. 24 1, 161–178.
- Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. *Handout of a talk given at Semantics and Philosophy in Europe (SPE 7), Berlin* 28.
- Fanselow, Gisbert, Malte Zimmermann & Mareike Philipp. 2022. Accessing the availability of inverse scope in german in the covered box paradigm. *Glossa: a journal of general linguistics* 7(1).
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2019. ‘lmerTest’ r package version 3.1-0. <https://cran.r-project.org/web/packages/lmerTest/index>.
- Patil, Umesh, Stefan Hinterwimmer & Petra B Schumacher. 2023. Effect of evaluative expressions on two types of demonstrative pronouns in german. *Glossa: a journal of general linguistics* 8(1).
- R Core Team. 2015. *R: A language and environment for statistical computing*. Foundation for statistical computing. <http://www.R-project.org/>.
- Silber, Henning, Joss Roßmann & Tobias Gummer. 2022. The issue of noncompliance in attention check questions: False positives in instructed response items. *Field Methods* 34(4). 346–360.