languages

MDPI

Article

Conceptual Number in Bilingual Agreement Computation: Evidence from German Pseudo-Partitives

Jana Reifegerste [1,2,*], Ayse Garibagaoglu [2] and Claudia Felser [2]

[1] Department of Neurology, Georgetown University, Washington, DC 20007, USA
[2] Potsdam Research Institute for Multilingualism, University of Potsdam, 14476 Potsdam, Germany
* Correspondence: jr1754@georgetown.edu

Abstract: During subject–verb agreement (SVA) computation, the conceptual or notional number of the subject can affect whether speakers choose a singular or a plural verb, potentially overriding the grammatical number of the subject's head. The influence of notional number has hardly been investigated in bilinguals, however. Most previous research on bilingual agreement computation has focused on agreement errors, and less is known about agreement computation in cases where multiple licit options exist. One such phenomenon is pseudo-partitives (German: *eine Tüte Nüsse* 'one bag of nuts'), for which a verb may agree with either the first or the second noun phrase. We present data from 150 L1 speakers of German and Turkish–German early bilinguals who performed a sentence-completion task. While both groups showed awareness of the optionality in agreement, both preferred the first noun phrase as the agreement controller. Interestingly, notional plurality affected bilinguals' verb choices more than those of L1 speakers, whose responses were influenced by notional plurality only in the most challenging number conflict condition. We suggest that increased cognitive demands during bilingual SVA computation may render bilinguals more susceptible to conceptual effects.

Keywords: sentence processing; subject–verb agreement; conceptual number; bilingualism; pseudo-partitives; German

check for updates

Citation: Reifegerste, Jana, Ayse Garibagaoglu, and Claudia Felser. 2023. Conceptual Number in Bilingual Agreement Computation: Evidence from German Pseudo-Partitives. *Languages* 8: 147. https://doi.org/10.3390/languages8020147

Academic Editor: Jeanine Treffers-Daller

Received: 1 December 2022
Revised: 21 May 2023
Accepted: 26 May 2023
Published: 5 June 2023

1. Introduction

A large body of research on bilingual acquisition and processing has examined subject–verb agreement (SVA) in both sentence production and sentence comprehension. Most previous studies on SVA computation have focused on agreement errors that come about in linguistic contexts where a clear distinction can be made between a correct versus an incorrect option. Such studies usually find that the computation of SVA is mostly error-free for adult first-language (L1) speakers and early bilinguals (Alarcón 2021; Foote 2011; Hacohen and Schaeffer 2007), though late bilinguals may struggle with it (Chen et al. 2007; Hopp 2010; Lardiere 1998; Reifegerste et al. 2020; Sato and Felser 2010; Shibuya and Wakabayashi 2008; VanPatten et al. 2012). However, not in all instances of SVA is there a categorical distinction between one choice being correct and any other choice being incorrect; instead, a language may allow for several options to be licit. This kind of optionality may occur, for example, in situations in which the grammatical number of the subject of a sentence is not identical to the conceptual or 'notional' number of the subject represented in the speaker's mind, rendering both a singular or a plural choice for the verb potentially correct (e.g., *The family was/were arguing*). Such contexts, despite being reasonably common, have received relatively little attention within the field of bilingual language research.

The present study examines SVA in (functionally monolingual) L1 speakers of German and Turkish–German early bilinguals. The grammatical phenomenon under investigation is German pseudo-partitives, which are complex noun phrases (NPs) in which the first noun phrase (NP1) denotes a proportion of the second one (NP2) (Selkirk 1977). Pseudo-partitives such as *eine Tüte Nüsse* ('one bag of nuts') in (1) allow for the verb to agree in number with either NP1 (*Tüte* 'bag') or NP2 (*Nüsse* 'nuts').[1]

(1)	*Eine*	*Tüte*	*Nüsse*	*kostet/kosten*	*zwei*	*Euro.*
	one	bag	nuts	costs/cost	two	Euros

'One bag of nuts costs/cost two Euros.'

Speakers' agreement choices for pseudo-partitive subjects may be influenced by the grammatical number of NP1 and NP2, by the conceptual properties of the individual NPs, by their semantic relationship, or by the overall notional number of the complex subject phrase. Of particular interest for the current study is the subject phrase's notional number; that is, whether it is represented as one thing or several things in the speaker's mind.

From the perspective of bilingual acquisition, cases of optional agreement are particularly interesting because they are gradient phenomena that are not usually taught, so the constraints determining number agreement preferences need to be acquired through naturalistic exposure and use. Examining and comparing the constraints that govern L1 speakers' and bilinguals' agreement choices can elucidate what aspects of SVA computation are vulnerable to suboptimal input and acquisition conditions, and what strategies bilinguals may use to resolve number conflicts (Jessen et al. 2021).

1.1. Grammatical versus Conceptual Number in Subject–Verb Agreement

In many languages, finite verbs must agree with the subject in one or more morphosyntactic features (Mallinson and Blake 1981). Of interest for the present study is one of the most common features encoded in SVA: number—that is, whether the subject refers to a single entity (singular) or several entities (plural).

The number of a subject is most commonly thought of as its grammatical number, which can be denoted through various means, most notably via inflectional morphology. For example, in many languages an affix marks whether a noun is grammatically singular or plural (e.g., English: *car—cars*, German: *Auto—Autos*; Turkish: *araba—arabalar*). Additionally, subjects also have conceptual or notional number, which is determined by whether the NP is thought to refer to one or multiple entities.

Situations in which the grammatical and conceptual number of a word do not coincide (e.g., *scissors*, *pants*, or *team*, *family*) give rise to uncertainty in language users about whether the verb that is supposed to agree with the NP should be in singular or plural number. One well-studied example for such potential mismatch is collective nouns (e.g., *family*, *team*, *police*), which refer to a group of individuals. Although these nouns are grammatically singular, their grammatical number may be overridden by their notional number. The decision of whether the relevant verb should receive singular or plural marking may depend on a variety of factors, such as the language variety (e.g., American vs. British English; Levin 2006), the diachronic variety (Bauer 1994; Dekeyser 1975; Liedtke 1910; Siemund 1995; but see Levin 2006), the verb to be used (Juul 1975; Quirk et al. 1985), and the noun itself (Depraetere 2003; Levin 2001). Indeed, Haskell et al. (2010) presented evidence from structural priming that speakers' choices for a singular versus a plural verb form are highly variable and can be influenced by previously encountered collective phrases.

Uncertainty about number agreement can also arise when the subject NP is syntactically complex, as in examples (2a–c).

(2) a. *the key to the cabinets*
  b. *a bunch of flowers*
  c. *neither Paul nor Ringo*

Noun phrases such as (2a) are well known for eliciting attraction errors, with speakers erroneously choosing a plural verb in the presence of a local plural noun despite the grammatical head of the complex NP being singular (e.g., Bock and Miller 1991). Several studies have pointed to a role for notional number in SVA computation in L1 speakers, mostly drawing on attraction errors. These studies usually investigate distributivity effects. For example, while the NP in (3a) likely refers to one key that fits into several cabinets, the NP in (3b) refers to several identical labels on several bottles (rather than a single label shared by multiple bottles).

(3) a. *the key to the cabinets* [non-distributive reading, notionally singular]
b. *the label on the bottles* [distributive reading, notionally plural]

Attraction errors are far more prevalent for preambles with a distributive/multiple-referents reading (3b) than for those with a non-distributive/single-referent reading (3a) (Vigliocco et al. 1995, 1996a, 1996b).

For pseudo-partitives such as (2b) and conjoined or disjoined NPs as in (2c), on the other hand, both singular and plural agreement may be deemed acceptable. While there is extensive psycholinguistic research on the role of conceptual properties of complex NPs in facilitating agreement errors (e.g., Brehm and Bock 2013; Gillespie and Pearlmutter 2011; Solomon and Pearlmutter 2004; Veenstra et al. 2014a; Vigliocco et al. 1995, 1996a, 1996b), few studies have examined notional number in situations of optional agreement. Some evidence for the role of notional number for variable agreement comes from Lorimor et al. (2016), who compared agreement choices for conjoined NPs that consisted of either two animate or count nouns (argued to be more notionally plural; e.g., *das Huhn und das Küken* 'the chicken and the chick') or two mass or deverbal nouns (argued to be more notionally singular; e.g., *das Heu und das Stroh* 'the hay and the straw') in Dutch and German. Participants' verb choices were affected by notional plurality, with conjoined NPs consisting of two animate/count nouns eliciting more plural verb choices than those made up of two mass/deverbal nouns. See also Lorimor (2007) for evidence from English.

Notional number effects can be captured by the 'Marking and Morphing' account of agreement (Bock et al. 2001; Eberhard et al. 2005). According to this model, agreement computation has two components, and effects of conceptual number on speakers' agreement choices are attributed to the first of these. During *number marking*, a subject NP is marked and mentally represented as singular or plural on the basis of conceptual or message-level features. *Number morphing*, in contrast, is a grammatically-driven process and occurs during constituent-structure assembly. It is during this process that number specifications are morphologically instantiated and agreement operations are implemented. When the conceptual and grammatical number of a subject are in conflict, the resolution of this conflict may result in conceptual number features overriding grammatical ones, potentially facilitating attraction errors. Hartsuiker and Barkhuysen (2006) proposed that the first ('marking') stage—the mapping of conceptual features to lexical representations—is not vulnerable to cognitive resource limitations, whereas the second ('morphing') stage is. This means that cognitive pressure or resource limitations may result in the second stage not being completed, leading to a greater likelihood of conceptual number features overriding grammatical number features.

Here, we investigate how conceptual number affects speakers' choices between different licit options during SVA computation with pseudo-partitives. Pseudo-partitives such as *a bunch of flowers* are characterized semantically as comprising a quantifier (NP1) and a restrictor (NP2). According to Smith et al. (2018), we can distinguish (at least) three types of pseudo-partitive, depending on the semantic nature of NP1: container words, collection words, and measure words. Container words refer to an object that contains its content within a defined space (e.g., *a glass of water*, *a bottle of pills*). They can be referred to on their own (i.e., without the NP2; e.g., *a glass*), and usually cannot take an abstract NP2 (e.g., #*a glass of concepts*). Collection words also denote some kind of spatial configuration, albeit more loosely than container words (e.g., *a stack of sandwiches*, *a pile of shirts*). Unlike containers, they usually cannot stand alone without an (implicit or explicit) referent. Lastly, measure words refer to a quantity of something without an implied spatial configuration (e.g., *a bevy of girls*, *an ounce of butter*). Despite the implications of the term, the NP1 need not be a standardized unit of measurement, but can also be more abstract (e.g., *a lot of*, *a bunch of*).

Smith et al. (2018) presented participants with singular–plural pseudo-partitives consisting of container NPs, collection NPs, or measure NPs and asked them rate these as "one thing" versus "more than one thing." Container NPs were rated significantly more likely to refer to "one thing" than collection NPs, and collection NPs were rated marginally more likely to refer to "one thing" than measure NPs, indicating a graded difference in notional plurality between the different types of pseudo-partitive. The authors also present results from a speeded sentence-completion task, in which participants read preambles containing pseudo-partitives with a singular–plural structure (e.g., *a tube of balls*) with different types of NP1 (containers, collections, measures), and were asked to pick either a singular or a plural verb form as the best-fitting verb. The descriptive results indicate a step-wise increase in plural responses, dovetailing with the order of notional plurality observed in the rating experiment (container words < collection words < measure words).[2] Thus, the results suggest that singular–plural pseudo-partitives with container nouns are more likely to be represented as single entities in a speaker's mind, while singular–plural pseudo-partitives formed with measure NP1s are more likely to be represented as multiple entities. One likely reason for this difference is that measures need not have an equivalent real-world referent—for example, *one bag of nuts* refers to exactly one object, while *one pound of nuts* might not—, making pseudo-partitives with measure NP1s overall harder to imagine and more abstract (Scontras 2014).

We will focus on the two types of pseudo-partitive that according to Smith et al. (2018) constitute the end points of the semantic plurality continuum: Pseudo-partitives with container nouns (which tend to be interpreted as individual objects) and measure nouns (which lean towards being notionally plural). Specifically, we examine how differences in their perceived plurality may affect SVA of German pseudo-partitives in German L1 speakers and Turkish–German bilinguals.

1.2. Agreement with Pseudo-Partitives in German

The choice of agreement controller is not always obvious for pseudo-partitive subjects, making them an interesting linguistic phenomenon to study (compare, e.g., Grestenberger 2015; Jessen et al. 2021; Scontras 2014; Selkirk 1977; Smith et al. 2018; Stickney 2009). German pseudo-partitives are formed by combining NP1 and NP2 without a preposition in between; see examples (4–6).

(4)	a.	*eine*	*Tüte*	*Nüsse*	[singular–plural pseudo-partitive, container NP1]
		one	bag	nuts	
		'one bag of nuts'			
	b.	*zwei*	*Tüten*	*Nüsse*	[plural-plural pseudo-partitive, container NP1]
		two	bags	nuts	
		'two bags of nuts'			
(5)		*zwei*	*Pfund*	*Nüsse*	[plural-plural pseudo-partitive, measure NP1]
		two	pound	nuts	
		'two pounds of nuts'			
(6)		*zwei*	*Pfund*	*Mehl*	[plural-singular pseudo-partitive, measure NP1]
		two	pound	flour	
		'two pounds of flour'			

The number marking on the two constituent NPs shows a certain degree of complexity. Container NP1s as in (4a,b) are generally marked for number on the determiner and/or the noun (depending on the noun's gender and whether it takes an overt plural affix). While marking measure NP1s for number is not considered ungrammatical, they usually have no overt number marking; see examples (5) and (6) (Hennig et al. 2016). Inflectional marking on the NP2 depends on whether the NP2 refers to a countable or an uncountable entity. Countable NP2s can carry plural marking (examples (4a,b)), while uncountable NP2s are unmarked and grammatically singular (example (6)).

SVA with German pseudo-partitive subjects has been the subject of some debate. According to the German reference grammar *Duden*, the prescriptive norm for pseudo-partitives is agreement of the verb with NP1. However, the *Duden* acknowledges that agreement with NP2 is also possible, and lists pseudo-partitives in its list of "cases of doubt" (Hennig et al. 2016). Thus, SVA in German pseudo-partitives is not determined by categorical rules of the grammar (see also Jaeger 1992; Wegerer 2012). From a formal syntactic perspective, variability in the choice of agreement controller reflects the possibility of analyzing either NP1 or NP2 as the syntactic head of the pseudo-partitive NP (Selkirk 1977).

SVA in German pseudo-partitives has scarcely been studied experimentally. We are aware of one published study examining number agreement of container nouns (Jessen et al. 2021). Similar to the present study, the participants in Jessen et al. (2021) were also German L1 speakers and Turkish–German bilinguals, whose age-of-acquisition (AoA) of German ranged from two to 27 years. Given the relevance of Jessen et al.'s (2021) study for the present one, we will discuss its methods and findings in some detail here.

Jessen et al. (2021) examined how grammatical number marking within the subject phrase affected speakers' agreement preferences. Experiment 1 asked participants to rate the acceptability of sentences containing pseudo-partitives such as (4–6) (manipulating NP1 number, NP2 number, and verb number) on a five-point Likert scale. Experiment 2 was a speeded forced-choice sentence-completion task in which participants were presented with preambles containing pseudo-partitive NPs (manipulating NP1 number and NP2 number). Across both experiments, both the L1 speakers and the bilingual group preferred NP1 as the agreement controller. NP2 also affected participants' preference patterns, though in an asymmetrical way: plural NP2s were more likely to trigger plural agreement than singular NP2s were to trigger singular agreement. This pattern is reminiscent of the singular–plural asymmetry found in agreement attraction studies with L1 speakers (e.g., Bock and Miller 1991) and (perhaps less robustly) with bilingual speakers (Reifegerste et al. 2020). Interactions indicated stronger effects of NP1 number on L1 speakers' (vs. bilinguals') responses and vice versa for NP2 number. The authors combined the agreement preferences resulting from the acceptability task (Experiment 1) with a Gradient Symbolic Computation model (Smolensky et al. 2014), and then compared the model's predictions against the findings from the production task (Experiment 2). The model computed the same relative constraint rankings for both participant groups on the basis of their acceptability ratings, and this constraint hierarchy also predicted participants' performance in the production task rather well.

Thus, Jessen et al. (2021) established that bilinguals are aware of the two licit agreement options for German pseudo-partitives, even though this phenomenon is not explicitly taught and does not exist in Turkish. Turkish pseudo-partitives never give rise to a number agreement conflict as both NP1 and NP2 are unmarked for number, and plurality is indicated by a numeral preceding the complex NP. Perhaps even more strikingly, the bilingual groups' preference patterns approximated those of the L1 group, although the bilinguals showed greater reliance on surface-level cues (e.g., noun proximity) than the L1 group. The present study seeks to build on the findings by Jessen et al. (2021) by examining the extent to which speakers' choice between the different licit agreement options is affected by the semantic category of the quantifier (NP1) and by the notional plurality of the subject phrase.

1.3. Conceptual Number in Bilingual Agreement Computation

The extent to which notional number plays a role in bilingual speakers' agreement choices has not been widely investigated, and the majority of studies have focused on late bilinguals, who acquired their L2 after childhood. The findings of these studies are mixed. Nicol and Greth (2003), for example, found equivalent distributivity effects in both languages in a group of English–Spanish late bilinguals. In contrast, Hoshino et al. (2010) examined L1 and L2 agreement computation in English–Spanish (Experiment 1) and Spanish–English late bilinguals (Experiment 2). Experiment 1 found the English–Spanish group to show sensitivity to the distributivity manipulation only in their L1 English, not in their L2 Spanish. The authors suggest that the discrepancy with previous findings might be due to proficiency differences between their participants and the group tested by Nicol and Greth (2003); this idea was supported by Experiment 2, in which highly proficient (but not less proficient) L2 speakers showed the relevant effects of the critical distributivity manipulation in their L2.

Findings from a study by Wei et al. (2015), which compared Chinese–English and Uyghur–English late bilinguals, suggest that L1 properties may also affect bilingual SVA computation. L2 English speakers with Uyghur (which has obligatory SVA) as their L1 showed distributivity effects during the production of English sentences, while even highly proficient English L2 speakers with Chinese (which does not have obligatory SVA) as their L1 did not (Experiments 1 and 2). Only when sentence preambles were accompanied by pictures of the referents did the Chinese–English bilinguals show distributivity effects (Experiment 3). In contrast, Jackson et al. (2018) found that L2 English speakers with Swedish (which has robust noun morphology, but no number-marking on verbs) or Chinese as their L1s were affected to similar degrees by the distributivity encoded in the stimulus material, with no effects of proficiency.

We are aware of two studies investigating the role of notional number in SVA computation in early bilinguals. Nicol et al. (2001) examined early Spanish–English and late English–Spanish bilinguals using an agreement-production task (Bock and Miller 1991). Participants were presented with sentence preambles which they were asked to repeat and complete with a sensible sentence ending. Both groups showed distributivity effects similar to those found for monolingual speakers. Foote (2010) also found distributivity effects in English–Spanish and Spanish–English bilinguals; however, in this study, effect sizes were modulated by both AoA and proficiency (which were crossed in the experimental design), with the largest effects for early bilinguals with advanced proficiency and the smallest effects for late bilinguals with intermediate proficiency.

To summarize, the extent to which bilinguals are sensitive to notional number during SVA computation is not entirely clear. It appears that early bilinguals tend to show sensitivity to the conceptual number of nouns or subject phrases, while the findings are more mixed for late bilinguals. Existing studies that tested early bilinguals have exclusively examined Spanish and English as target languages. This calls not only for further research investigating other languages, but, more critically, research on language combinations in which the two languages in a bilingual speaker's mind may differ from one another with regard to how SVA is computed (e.g., Jackson et al. 2018; Wei et al. 2015). Second, and more importantly, the research cited above examines exclusively how SVA computation is affected by notional number in environments with a clear grammatically correct (vs. incorrect) response. The present study, in contrast, investigates how conceptual number affects L1 speakers' and bilinguals' choices between different licit agreement options.

1.4. The Present Study

We present data from an off-line sentence-completion task in which German native speakers and Turkish–German early bilinguals read sentence fragments containing German pseudo-partitive constructions and chose whether a sentence should best be completed with a singular or a plural verb. Recall that L1 Turkish speakers cannot draw on their first language when choosing between different verb forms in German since there is no number conflict between NP1 and NP2 in in Turkish pseudo-partitives such as (7).

(7)	*iki*	*kutu*	*elma*
	two	box	apple
	'two boxes of apples'		

Neither NP1 nor NP2 carry a number marker in Turkish pseudo-partitives, and plurality is indicated only by the initial numeral. This holds for both container and measure NP1s and for countable and uncountable NP2s. Kornfilt (1997) argues that the last element of a complex NP constitutes its head, with the preceding elements (i.e., NP1 in pseudo-partitives) serving as modifiers, rendering NP2 the head of pseudo-partitives such as (7). In Turkish, the verb in such cases should thus always be singular. Therefore, Turkish speakers need to acquire the constraints governing number agreement preferences in German through linguistic experience. Although Jessen et al. (2021) found no statistical effects of the age at which their bilingual participants had started acquiring German, here we focus on a more homogeneous group of early bilinguals with high German proficiency.

We examine whether and how agreement choices are affected by conceptual number features of pseudo-partitives, and whether agreement preferences differ between the two speaker groups. We manipulated the semantic type of NP1 by having the NP1 of half of the pseudo-partitives be a container noun, while the other half used measure nouns as NP1s. NP1 and NP2 either matched or mismatched in their grammatical number. This resulted in the following design (also see Section 2): NP1 NUMBER (2 levels: singular, plural) × MATCH (2 levels: match, mismatch) × SEMANTIC CATEGORY (2 levels: container, measure) × GROUP (2 levels: L1 group, bilingual group).

Besides manipulating the semantic category of the first NP, we also collected notional plurality ratings for the pseudo-partitives that each participant saw. This was important since the abovementioned tendencies (container words = one thing, measure words = multiple things) are not absolute, and thus may apply neither to all items nor to all speakers.[3]

We sought to answer the following research questions:

- I. (a) When computing agreement with German pseudo-partitives, do speakers use the first (NP1) or the second noun phrase (NP2) as the agreement controller?
  - (b) Do L1 speakers and early bilinguals differ in their preference for NP1 versus NP2 as the agreement controller?
- II. (a) What is the role of conceptual number in agreement computation (operationalized through different types of NP1 and notional-number ratings)?
  - (b) Does the role of conceptual number in agreement computation differ for L1 speakers versus early bilinguals?

Our predictions are derived on the basis of prior research on agreement computation, with a particular focus on production studies. Most seminal studies on agreement attraction elicited responses by asking participants to orally complete auditorily or visually presented sentence fragments (e.g., Bock and Miller 1991), to repeat a stimulus sentence and supply a continuation (e.g., Bock et al. 2001; Vigliocco et al. 1995), or to describe pictures (e.g., Kandel et al. 2022; Veenstra et al. 2014b); however, much recent work (including the present study) has employed a version of the forced-choice sentence-completion task introduced by Staub (2009), in which participants are presented with preambles and choose a verb form out of several options. Though the task demands of this task are somewhat different from those in more traditional production tasks (e.g., selecting the form vs. generating the verb), key effects from production tasks (e.g., asymmetric attraction effects, notional-number effects, monolingual/bilingual differences) have been reliably found in studies employing forced-choice sentence-completion tasks (e.g., Reifegerste et al. 2020; Smith et al. 2018; Staub 2009, 2010; Villata and Franck 2020), rendering it a suitable task for examining our research questions. See also Veenstra et al. (2014a) for a direct comparison of results from a picture-description task and a forced-choice sentence-completion task using the same stimuli.

On the basis of previous work by Jessen et al. (2021), we expected both German L1 speakers and Turkish-German bilinguals to use NP1 as the primary agreement controller. Effects of NP2 might also play a role, yielding an asymmetrical pattern for mismatching items, with more plural responses following singular (SG)—plural (PL) combinations (*one bag of nuts*) than singular responses following PL-SG combinations (*two pounds of flour*). Such a pattern would also be in line with the singular–plural asymmetry commonly found in agreement-attraction studies. Given that Jessen et al. (2021) found bilinguals to be more strongly affected by NP2 number than German L1 speakers, we might similarly find a greater reliance on NP2 for the bilingual group in our study.

As regards research question II, SG-PL pseudo-partitives containing measure nouns (*one pound of nuts*) are expected to elicit a greater proportion of plural verb choices than those with singular container nouns (*one bag of nuts*) as the former are more likely to refer to several things in the speaker's mind than the latter (Smith et al. 2018), which should in turn be reflected in a greater likelihood for SG-PL measure nouns (vs. SG-PL container nouns) to yield plural verb choices. Similarly, we expected notional-number ratings to predict agreement choices, with more plural verb choices following pseudo-partitives with higher plurality ratings.

Our predictions regarding potential differences in the role of conceptual number for L1 speakers versus early bilinguals are more speculative. While several previous studies have found early bilinguals to be sensitive to notional number, this research has focused on categorical rather than optional agreement marking, and on languages which are very similar with regard to the phenomenon under study (Spanish and English). If Hartsuiker and Barkhuysen (2006) are correct in their assertion that the initial number-marking stage of agreement computation (to which effects of conceptual number have been attributed) is less costly in terms of cognitive resources and less vulnerable to limitations on these resources than the subsequent control (or 'morphing') stage, then bilingual speakers may be more likely than L1 speakers to base their agreement choices on the first stage. This is under the assumption that grammatical computations are cognitively more taxing for bilingual than for monolingual speakers and that thus cognitive resource limitations may affect bilinguals to a greater extent than monolinguals. Such cognitive resource limitations have been proposed as a potential reason for differences between bilingual and monolingual processing in a range of grammatical phenomena, including subject–verb agreement (e.g., Brehm et al. 2022; Cunnings 2017; Foote 2011; Hopp 2006, 2010; Keating 2009; Lehtonen et al. 2023; McDonald 2006; McDonald and Roussel 2010; Sagarra and Herschensohn 2010), and for findings suggesting a greater dependence of agreement processing on domain-general abilities in bilinguals as compared to monolinguals (e.g., Gangopadhyay et al. 2016; Reifegerste et al. 2020; Wood et al. 2021). If such effects are also at play here, bilingual speakers—who were shown by Jessen et al. (2021) to be aware that German pseudo-partitives allow for optionality in number agreement—might show larger effects of conceptual number compared to L1 speakers.

2. Method

2.1. Participants

We present data from 150 participants, of whom 62 were German L1 speakers (henceforth 'L1 speakers') and 88 were Turkish–German early bilinguals (henceforth 'bilinguals'). None of the participants had learned another language besides German (L1 speaker group) or German and Turkish (bilingual group) before the age of 6. All Turkish–German bilinguals had learned German before entering primary school, and all except for three had been living in Germany since birth.[4] All participants had at least 10 years of formal education, ensuring suitable reading abilities; there was no significant difference in educational level between the groups. The bilinguals rated their German skills (overall as well as separately for speaking, listening, writing, and reading) as higher than their Turkish skills; there was no difference in their self-rated enjoyment of using German versus Turkish. See Table 1 for demographic information.

Participants were recruited through the participant database of the University of Potsdam, social media, and word of mouth, and were offered monetary compensation for their participation. All participants provided informed consent to the study.

Table 1. Demographic information on participants, by participant group and by language (where applicable). Standard deviations are provided in parentheses. The column 'Group differences' presents results from *t*-tests for group comparisons on numerical variables (age, language skills, language enjoyment) and results from *chi-square* tests for group comparisons on categorical variables (sex, education, age-of-acquisition).

		L1 Speakers	Bilinguals		Group Differences
Demographic information					
n		62	88		
Age		35.5 (14.3)	27.1 (6.7)		$t(149) = 4.82$, $p < 0.001$
Sex	female	42	65		$\chi^2(1, N = 150) = 0.67$, $p = 0.414$
	male	20	23		
Education (highest degree)	less than 12 years	3	5		$\chi^2(4, N = 150) = 7.56$, $p = 0.109$
	high school diploma	13	29		
	vocational training	4	6		
	Bachelor's degree	16	29		
	Master's degree and above	26	19		
Language information			*German*	*Turkish*	*Differences between languages*
Age-of-acquisition	since birth	–	36	86	$\chi^2(1, N = 88) = 66.79$, $p < 0.001$
	during early childhood (before primary school)	–	52	2	
Language skills (out of 10)	Speaking	–	9.4 (0.8)	8.5 (1.3)	$t(87) = 6.13$, $p < 0.001$
	Listening	–	9.7 (0.8)	9.2 (1.0)	$t(87) = 4.00$, $p < 0.001$
	Writing	–	9.5 (1.0)	8.0 (1.8)	$t(87) = 6.59$, $p < 0.001$
	Reading	–	9.7 (0.7)	8.7 (1.6)	$t(87) = 5.90$, $p < 0.001$
	Overall	–	9.6 (0.7)	8.6 (1.2)	$t(87) = 6.65$, $p < 0.001$
Enjoyment (out of 5)		–	4.4 (0.8)	4.5 (0.6)	$t(87) = 0.43$, $p = 0.669$

2.2. Materials

Our study consisted of two tasks: a binary forced-choice sentence-completion task and a plurality-rating task.

2.2.1. Sentence Completion Task

For the experimental tasks, 40 experimental sentence fragments were created. All experimental stimuli had the same structure: a main clause preamble followed by a declarative complement clause which consisted of a pseudo-partitive subject NP followed by the lexical verb in participle form. The complement clause needed to be completed with a 3rd person auxiliary in sentence-final position, for which participants decided whether it should be inflected for singular or plural number.

Twenty experimental stimuli included pseudo-partitives with container nouns, and 20 included pseudo-partitives with measure nouns. Container nouns are defined in accordance with Smith et al. (2018) as confining their contents within a defined space. All container nouns were marked for plural in the PL-Match and PL-Mismatch conditions. Measure nouns are defined here as standard units of measure. With the exception of *Tonne* ('tonne'), which takes the *-n* plural (*Tonnen*), measure nouns are not marked for plural. Table 2 illustrates our experimental conditions; see Table S1 in Supplementary Materials for a list of all NP1s used in the study.

Table 2. Overview of the conditions.

Semantic Category	Condition	Preamble		NP1	NP2	Verb	Auxiliary	
							SG	PL
Container	Singular-Match	*Oskar glaubt, dass*	*eine*	*Schüssel*	*Joghurt*	*gegessen*	*wurde.*	**wurden.*
		Oskar thinks that	one	bowl	yogurt	eaten	was.	*were.
	Singular-Mismatch	*Oskar glaubt, dass*	*eine*	*Schüssel*	*Beeren*	*gegessen*	*wurde.*	*wurden.*
		Oskar thinks that	one	bowl	berries	eaten	was.	were.
	Plural-Match	*Oskar glaubt, dass*	*vier*	*Schüsseln*	*Beeren*	*gegessen*	**wurde.*	*wurden.*
		Oskar thinks that	four	bowls	berries	eaten	*was.	were.
	Plural-Mismatch	*Oskar glaubt, dass*	*vier*	*Schüsseln*	*Joghurt*	*gegessen*	*wurde.*	*wurden.*
		Oskar thinks that	four	bowls	yogurt	eaten	was.	were.
Measure	Singular-Match	*Sophia sagt, dass*	*ein*	*Pfund*	*Mehl*	*bestellt*	*ist.*	**sind.*
		Sophia says that	one	pound	flour	ordered	is.	*are.
	Singular-Mismatch	*Sophia sagt, dass*	*ein*	*Pfund*	*Nüsse*	*bestellt*	*ist.*	*sind.*
		Sophia says that	one	pound	nuts	ordered	is.	are.
	Plural-Match	*Sophia sagt, dass*	*drei*	*Pfund*	*Nüsse*	*bestellt*	**ist.*	*sind.*
		Sophia says that	three	pound	nuts	ordered	*is.	are.
	Plural-Mismatch	*Sophia sagt, dass*	*drei*	*Pfund*	*Mehl*	*bestellt*	*ist.*	*sind.*
		Sophia says that	three	pound	flour	ordered	is.	are.

Note. English gloss: Container: 'Oscar thinks that one bowl/four bowls of yogurt/berries was/were eaten.' Measure: 'Sophia says that one pound/three pounds of flour/nuts is/are ordered.' NP1 = Noun 1, NP2 = Noun 2. Asterisks indicate grammatically incorrect forms.

Singular NP1s were preceded by *ein* (for masculine and neuter nouns) or *eine* (for feminine nouns), which may be read as either the indefinite article (viz. 'a' in English) or as the numeral 'one.' Numeral quantifiers for plural NP1s varied from two to 500 and were spelled out as words. Singular NP2s were uncountable nouns such as *Mehl* ('flour'), *Kohle* ('coal'), or *Marmelade* ('jam'). It was not possible to use singular countable nouns here because the NP2 in German pseudo-partitives is always plural unless the noun is uncountable. Plural NP2s are countable nouns marked for plural with their respective plural suffix and, where applicable, umlauting (e.g., *Beere-Beeren* 'berry'/'berries,' *Nuss-Nüsse* 'nut'/'nuts'). All NP2s were concrete nouns.

The last element of each sentence fragment was a lexical past participle. This was followed by an auxiliary verb presented in both its singular and plural forms, which the participants were asked to select. Ten measure and ten container items used the auxiliary *sein* with its forms *ist* and *sind* ('to be'; *is/are*) and ten of each category used *werden* (*wurde/wurden*), an auxiliary that can also be used to form the passive in German. Different auxiliaries were used to minimize the possibility that participants recognized patterns in the items or developed response strategies. Note that in our SG-Match conditions, the only grammatically correct choice is a singular auxiliary, whilst in the PL-Match conditions a plural auxiliary is the only correct choice. In Table 2, grammatically incorrect auxiliary forms are marked by an asterisk.

The vocabulary chosen for the experimental items is quite common and should not pose a problem for L1 and proficient L2 speakers of German. Typical German names were chosen for the preambles to keep the participants in monolingual German 'mode' as much as possible (Grosjean 2001).

The 40 experimental sentences (20 measure sentences, 20 container sentences) were distributed across four lists using a Latin Square design; that is, each participant saw five items per condition.

Twenty filler sentences were added, for a total of 60 sentences. All filler sentences started with preambles similar to the ones used in the experimental sentences and ended with an auxiliary, but did not contain pseudo-partitives. Fifteen of these filler sentences were unambiguous regarding the required agreement (e.g., *Annika sagt, dass der Himmel blau ist/*sind*, 'Annika says that the sky is/*are blue'). Five filler sentences contained collection nouns or quantifiers and were ambiguous regarding agreement assignment (*Nicole erklärt, das eine Herde Kühe die Blumen zertrampelt hat/haben*, 'Nicole explains that a herd of cows has/have trampled the flowers').

The 60 sentences (40 experimental sentences, 20 filler sentences) were pseudorandomized, such that no more than two consecutive sentences had the same NP1 NUMBER (singular/plural).

2.2.2. Plurality-Rating Task

In the plurality-rating task, participants were presented with the same pseudo-partitives they had seen during the sentence-completion tasks (i.e., from the same presentation list used in that task) and were asked to indicate whether a given NP (e.g., *ein Pfund Nüsse* 'one pound of nuts,' *vier Schüsseln Joghurt* 'four bowls of yogurt') denoted 'one thing' (*eine Sache*) or 'multiple things' (*mehrere Sachen*).

Mean plurality ratings were calculated for each of the 40 pseudo-partitives in each condition, separately for the L1 speaker group and the bilingual group.

2.3. Procedure

The experiment was conducted online using Google Forms. All instructions in the questionnaire as well as the recruitment email were written in German.

In the first part of the study (the sentence-completion task), participants were instructed to complete sentence fragments by choosing one of two auxiliaries (singular or plural) using radio buttons. Two examples were provided, a container sentence in the SG-Match condition and a filler sentence. Participants were then shown the sixty sentences, one sentence at a time, and were asked to select the auxiliary and continue to the next item by clicking a button. They were able to work at their own pace, but were instructed to follow their gut. It was not possible to return to a previous item to change one's answer. In the second part of the study (the plurality-rating task), participants were asked to gauge whether in their mind each of the forty pseudo-partitives they had responded to in the first part denotes 'one thing' or 'multiple things.' Item presentation order was identical to the first task. The study concluded with a brief questionnaire on demographic information and language background.

2.4. Analyses

The dependent measure was whether the response given by the participant matched in number with the first noun (NP1; i.e., the grammatical subject of the sentence). We calculated mixed-effects logistic regression models (binomial family and bobyqa optimizer) using the lme4 package (Bates et al. 2015).

For the first set of analyses ('Analysis I: Across all items'), the following within-items fixed factors of interest were included in the model: GROUP (2 levels: German L1 speakers, Turkish–German bilinguals), NP1 NUMBER (2 levels: singular, plural), and MATCH (i.e., whether NP1 and NP2 matched in number; 2 levels: match, mismatch). For the second set of analyses ('Analysis II: Semantic Category'), the between-items factor SEMANTIC CATEGORY (2 levels: Container noun, Measure noun) was added. For the third set of analyses ('Analysis III: Plurality Ratings'), SEMANTIC CATEGORY was replaced by the factor PLURALITY RATING (continuous; ranging from 0 to 1).

All continuous predictors were mean-centered; all categorical predictors were assigned sum-coded contrasts (−0.5 and 0.5) (Barr et al. 2013); see Table notes of the result tables for the respective contrasts that were assigned to the relevant levels in each model. Follow-up analyses to interactions with categorical predictors were performed by re-fitting the models with nested effects.

Random factors were participants and items. Following Barr et al. (2013), we started with a maximal random-effects structure and simplified the model in cases of convergence failure. This led to the inclusion of NP1 NUMBER as a by-participant intercept across all analyses.

3. Results

3.1. Analysis I: Across All Items

In a first set of analyses, we examined whether German L1 speakers and Turkish–German bilinguals differed in their response patterns as a function of NP1 NUMBER (singular vs. plural first noun) and MATCH (matching vs. mismatching second noun). See Table 3 for the untransformed descriptive data, Figure 1 for an illustration of these data, and Table 4 for the generalized mixed-effects model fit to the data.

Table 3. Performance in binary forced-choice experiment, by group and condition; descriptive data (mean proportions, standard deviations in parentheses).

Proportion of Responses Matching the Number of NP1		**L1 Speakers**			**Bilinguals**		
		NP1 Number			**NP1 Number**		
		SG	PL	*average*	SG	PL	*average*
Match	match	0.984 (0.126)	0.989 (0.106)	0.986 (0.116)	0.932 (0.252)	0.966 (0.182)	0.949 (0.220)
	mismatch	0.797 (0.403)	0.952 (0.215)	0.874 (0.332)	0.490 (0.500)	0.818 (0.386)	0.654 (0.476)
	average	0.890 (0.313)	0.970 (0.170)	0.930 (0.255)	0.711 (0.454)	0.892 (0.310)	0.801 (0.399)
Effect of MATCH (Difference match vs. mismatch)		0.187	0.037	0.112	0.442	0.148	0.295

Note. NP1 = Noun 1. SG = singular, PL = plural.

Table 4. Performance in binary forced-choice experiment; logistic mixed-effects regression model.

Random Effects		*Variance*	*SD*	**Correlation**
subjects	Intercept	1.2918	1.1366	
	NP1 Number	3.0084	1.7345	−0.15
items	Intercept	0.7172	0.8469	

Fixed Effects	***b***	***SE***	***z*-Value**	***p*-Value**
Intercept	3.4568	0.2019	17.12	<0.001
NP1 Number	1.2450	0.2631	4.73	<0.001
Match	−2.7643	0.1712	−16.15	<0.001
Group	1.9420	0.2699	7.20	<0.001
NP1 Number × Match	1.5601	0.3343	4.67	<0.001
NP1 Number × Group	−0.3849	0.4758	−0.81	0.419
Match × Group	0.2691	0.3292	0.82	0.414
NP1 Number × Match × Group	0.3911	0.6578	0.60	0.552

Note. Formula in R: DV ~ 1 + NP1 NUMBER * MATCH * GROUP + (1 + NP1 NUMBER | participant) + (1 | target). NP1 NUMBER is coded as −0.5 for singular and 0.5 for plural. MATCH is coded as −0.5 for match and 0.5 for mismatch. GROUP is coded as −0.5 for L1 speakers and 0.5 for bilinguals. See Table S2 in Supplementary Materials for follow-up statistics.

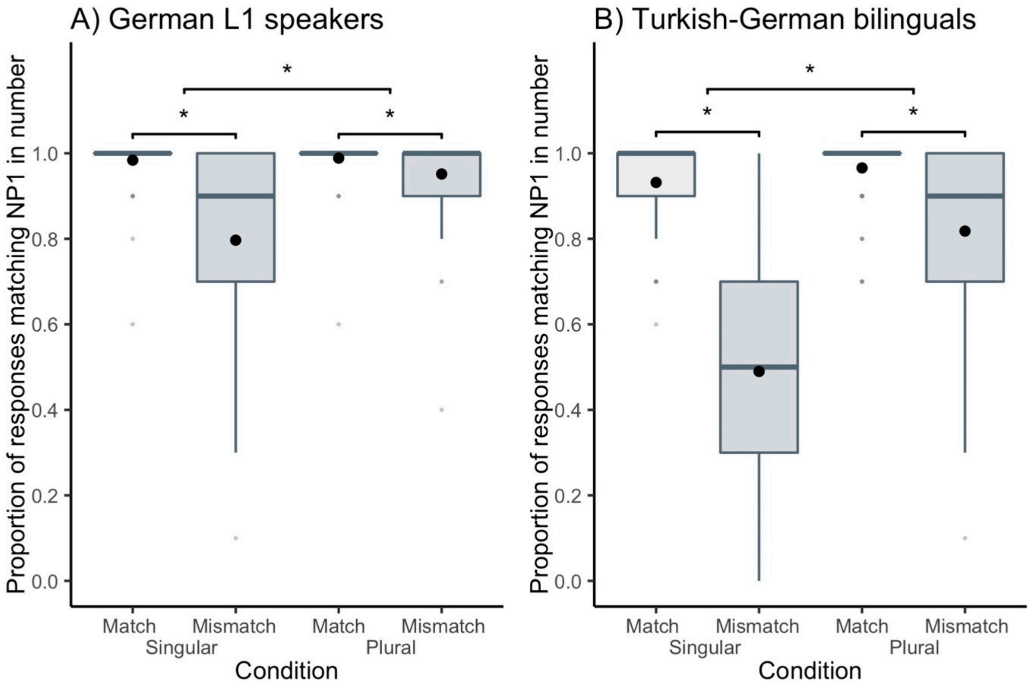

Figure 1. Performance in binary forced-choice experiment, by condition and group. Boxes correspond to the interquartile range (IQR; 25–75%). Whiskers correspond to minimal and maximal datapoints (excluding outliers). Gray dots outside the boxes correspond to outliers (beyond 1.5 × IQR). Black dots inside the boxes correspond to the condition mean. Lines inside the boxes correspond to the condition median. Asterisks indicate statistically significant effects ($p < 0.050$). All plots are created with ggplot (Wickham 2016) and are based on untransformed data, unless otherwise noted.

Responses by the bilingual group (vs. the L1 group) were significantly less likely to match the subject's first noun (NP1) in number, as evidenced by a main effect of GROUP.

However, the response patterns displayed by the two groups of participants (Table S3 in Supplementary Materials) did not significantly differ from each other (and were mirrored in corresponding effects across both groups; see Table 4). Both groups showed a main effect of NP1 NUMBER, with more NP1-matching responses for sentence fragments containing subjects with plural NP1s (PL-Match, PL-Mismatch) than for those with singular NP1s (SG-Match, SG-Mismatch).

Both groups also showed a main effect of MATCH; for sentence fragments in which NP1 and NP2 matched in number (SG-Match, PL-Match), participants were more likely to choose a verb that matched the two NPs in number than they were to choose an NP1-matching responses when NP1 and NP2 mismatched (SG-Mismatch, PL-Mismatch).

Lastly, in both groups these two main effects were qualified by an interaction between NP1 NUMBER and MATCH. While the effect of MATCH affected responses regardless of NP1 number, the effect was larger for pseudo-partitives with singular NP1s (across participant groups: 33.5 percentage points difference) than for those with plural NP1s (across participant groups: 10.1 percentage points difference).[5] There were no significant differences between the two groups in the size of any of these effects (i.e., no interactions with the factor GROUP). See Table 4 for the output of the generalized mixed-effects model across both groups.

To summarize, while the group of Turkish–German bilinguals was overall less likely than the group of German L1 speakers to choose a verb form that matched the NP1 in number (as evidenced by a significant main effect of GROUP), the groups did not differ in their response pattern in this set of analyses (which would have been indexed by an interaction involving the factor GROUP). Both groups chose NP1-mismatching auxiliaries more often when NP1 and NP2 mismatched in number, and this effect was particularly pronounced when NP1 was singular (SG-Match vs. SG-Mismatch) as compared to when it was plural (PL-Match vs. PL-Mismatch). Next, we examined whether group differences might emerge when adding a semantic variable, namely whether NP1 was a Container noun or a Measure noun.

3.2. Analysis II: Semantic Category

In a next set of analyses, we investigated whether the semantic category (Container noun vs. Measure noun) of NP1 affected response patterns. See Table 5 for the untransformed descriptive data, Figure 2 for an illustration of these data, and Table 6 for the generalized mixed-effects model fit to the data.

Table 5. Performance in binary forced-choice experiment, by group, semantic category, and condition; descriptive data (mean proportions, standard deviations in parentheses).

Proportion of Responses Matching the Number of NP1		**L1 Speakers**						**Bilinguals**					
		Container			**Measure**			**Container**			**Measure**		
		NP1 Number			**NP1 Number**			**NP1 Number**			**NP1 Number**		
		SG	PL	*average*	SG	PL	*average*	SG	PL	*average*	SG	PL	*average*
Match	match	0.994 (0.080)	0.994 (0.080)	0.994 (0.080)	0.974 (0.159)	0.984 (0.126)	0.979 (0.143)	0.936 (0.244)	0.980 (0.142)	0.958 (0.201)	0.927 (0.260)	0.952 (0.213)	0.940 (0.238)
	mismatch	0.887 (0.317)	0.984 (0.126)	0.935 (0.246)	0.706 (0.456)	0.919 (0.273)	0.813 (0.390)	0.607 (0.489)	0.934 (0.248)	0.770 (0.421)	0.373 (0.484)	0.702 (0.458)	0.537 (0.499)
	average	0.940 (0.237)	0.989 (0.106)	0.965 (0.185)	0.840 (0.367)	0.952 (0.215)	0.896 (0.305)	0.772 (0.420)	0.957 (0.203)	0.864 (0.343)	0.650 (0.477)	0.827 (0.378)	0.739 (0.440)
Effect of MATCH (Difference match vs. mismatch)		0.107	0.010	0.059	0.268	0.065	0.166	0.329	0.046	0.188	0.554	0.250	0.403

Note. NP1 = Noun 1. SG = singular, PL = plural.

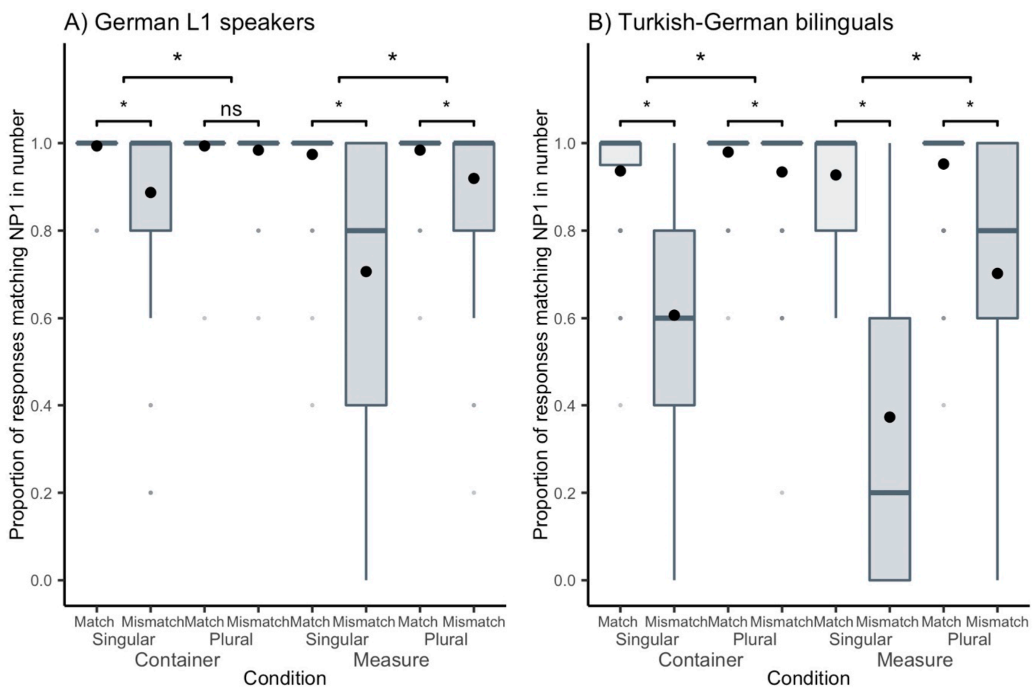

Figure 2. Performance in binary forced-choice experiment, by condition, semantic category, and group. Asterisks indicate statistically significant effects ($p < 0.050$); *ns* indicates statistically non-significant differences ($p > 0.100$).

Table 6. Performance in binary forced-choice experiment; logistic mixed-effects regression model.

Random Effects		***Variance***	***SD***	**Correlation**
subjects	Intercept	1.3547	1.1639	
	NP1 Number	3.1895	1.7859	−0.08
items	Intercept	0.2088	0.4569	

Fixed Effects	***b***	***SE***	***z*-Value**	***p*-Value**
Intercept	3.5134	0.1770	19.85	<0.001
NP1 Number	1.4008	0.2929	4.78	<0.001
Match	−2.6145	0.1922	−13.61	<0.001
Semantic Category	1.3949	0.2369	5.89	<0.001
Group	−2.0331	0.2937	−6.92	<0.001
NP1 Number × Match	1.6816	0.3800	4.43	<0.001
NP1 Number × Semantic Category	0.4737	0.3738	1.27	0.205
NP1 Number × Group	0.8563	0.3696	2.32	0.021
Match × Semantic Category	0.4771	0.5275	0.91	0.366
Match × Group	−0.1622	0.3761	−0.43	0.666
Semantic Category × Group	−0.4106	0.3710	−1.11	0.268
NP1 Number × Match × Semantic Category	0.4226	0.7360	0.57	0.566
NP1 Number × Match × Group	−0.5833	0.7518	−0.78	0.438
NP1 Number × Semantic Category × Group	0.9083	0.7416	1.23	0.221
Match × Semantic Category × Group	0.6765	0.7352	0.92	0.358
NP1 Number × Match × Semantic Category × Group	−0.6707	1.4688	−0.46	0.648

Note. Formula in R: DV ~ 1 + NP1 NUMBER * MATCH * SEMANTIC.CATEGORY * GROUP + (1 + NP1 NUMBER | participant) + (1 | target). NP1 NUMBER is coded as −0.5 for singular and 0.5 for plural. MATCH is coded as −0.5 for match and 0.5 for mismatch. SEMANTIC CATEGORY is coded as −0.5 for Measure nouns and 0.5 for Container nouns. GROUP is coded as −0.5 for L1 speakers and 0.5 for bilinguals. See Tables S5–S9 in Supplementary Materials for follow-up statistics.

Across both groups, we found a main effect of SEMANTIC CATEGORY on the response choices; sentences with Container nouns yielded more NP1-matching responses than sentences containing Measure nouns (0.906 vs. 0.804). However, SEMANTIC CATEGORY did not interact with any of the other factors (see Table 6). As in ‘Analysis I,’ across both participant groups we found significant main effects of NP1 NUMBER (more NP1-matching responses for preambles containing subjects with plural NP1s versus those with singular NP1s), MATCH (more NP1-matching responses for preambles with number-matching vs. mismatching nouns), and GROUP (more NP1-matching responses in the L1 vs. the bilingual group), as well as an interaction between NP1 NUMBER and MATCH (greater effects of MATCH for preambles with singular NP1s than for those with plural NP1s).

When comparing the two participant groups in hypothesis-driven follow-up analyses, L1 speakers and bilinguals showed similarities as well as differences in their response patterns (see Table S5 in Supplementary Materials). The effects found for either group in the first set of analyses (‘Analysis I: Across all items’; a main effect of NP1 NUMBER, a main effect of MATCH, and an interaction between NP1 NUMBER and MATCH; Table S3 in Supplementary Materials) were still present when SEMANTIC CATEGORY was factored in. Importantly, for the bilingual group only, SEMANTIC CATEGORY interacted with NP1 NUMBER (greater effect of NP1 NUMBER for Container vs. Measure nouns) and with MATCH (greater effect of MATCH for Measure vs. Container nouns). See Tables S6–S9 in Supplementary Materials for details on these follow-up analyses. There was no interaction between NP1 NUMBER, MATCH, and SEMANTIC CATEGORY for either participant group.

To summarize, the inclusion of the factor SEMANTIC CATEGORY had only a small effect on overall responses, yielding more NP1-matching responses when the first noun was a Container noun as compared to when it was a Measure noun. Examining the two participant groups separately did reveal a stronger influence of the semantic type of NP1 on the bilingual group’s responses as compared to the German L1 group’s. However, these group-level differences were not reflected in higher-level interactions with the factor GROUP.

As a last step, we performed a finer-grained assessment of the influence of semantics. In the last set of analyses, we replaced the binary factor SEMANTIC CATEGORY with a more direct and nuanced measure of notional plurality, which was collected from the same participants in a second questionnaire after performing the sentence-completion task (see Section 2).

3.3. Analysis III: Plurality Ratings

In our final set of analyses, we examined whether the pseudo-partitive NPs' notional plurality affected participants' verb form choices. After performing the binary forced-choice task, participants rated whether each pseudo-partitive they had previously seen corresponded to 'one thing' (coded as 0) or 'multiple things' (coded as 1). This resulted in mean PLURALITY RATING values for each item in each condition, ranging from 0 (akin to 'unambiguously singular') to 1 (akin to 'unambiguously plural'). Table 7 displays the descriptive data from the plurality-rating task (with higher numbers denoting greater notional plurality), showing relatively small between-group differences in the plurality ratings assigned to the pseudo-partitives. Both groups assigned very low ratings (reflecting a high proportion of 'one thing' responses) to SG-Match (e.g., *eine Schüssel Joghurt* 'one bowl of yogurt') items, while PL-Match items (e.g., *vier Schüsseln Beeren* 'four bowls of berries') received comparatively high ratings (reflecting many 'multiple things' ratings), with the two mismatch conditions (SG-Mismatch: *eine Schüssel Beeren* 'one bowl of berries'; PL-Mismatch: *vier Schüsseln Joghurt* 'four bowls of yogurt') falling in-between.

Table 7. Mean plurality rating (higher numbers = greater notional plurality), by group and condition; descriptive data (means, standard deviations in parentheses). See Table S10 in Supplementary Materials for mean plurality ratings further broken down by Semantic Category.

Plurality Rating		**L1 Speakers**			**Bilinguals**		
		NP1 Number		*Difference between SG and PL*	**NP1 Number**		*Difference between SG and PL*
		SG	**PL**		**SG**	**PL**	
Match	match	0.056 (0.078)	0.796 (0.212)	0.740	0.070 (0.062)	0.742 (0.246)	0.672
	mismatch	0.270 (0.135)	0.690 (0.301)	0.420	0.290 (0.155)	0.599 (0.338)	0.309
Difference between match and mismatch		0.214	0.106		0.220	0.143	

Note. NP1 = Noun 1. SG = singular, PL = plural.

Figure 3 illustrates the relationship between plurality ratings and sentence-completion responses. See Tables 8 and 9 for the generalized mixed-effects model fit to the data.

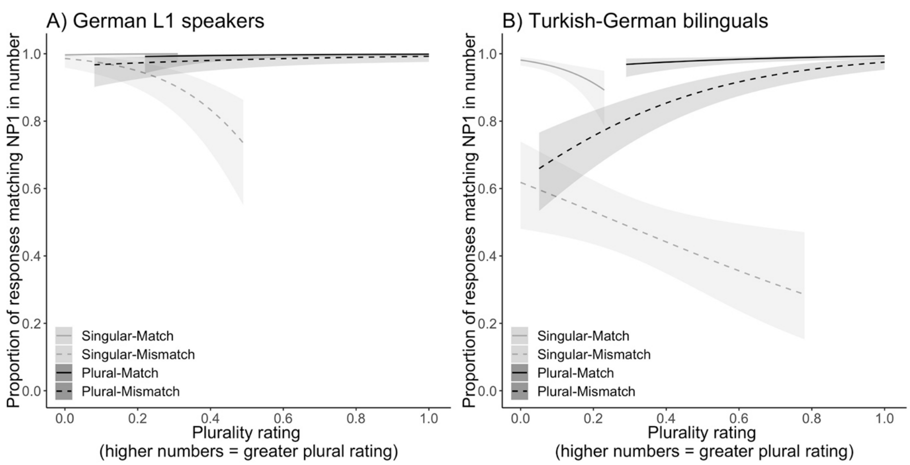

Figure 3. Performance in binary forced-choice experiment as a function of plurality rating, by condition and group. Shaded bands represent standard errors. Plot is based on mixed-effects regression model estimates; see note for Table 8 for details. See Table S11 in Supplementary Materials for follow-up statistics on the effect of PLURALITY RATING for the different conditions.

Table 8. Performance in binary forced-choice experiment; logistic mixed-effects regression model.

Random Effects		*Variance*	*SD*	**Correlation**
subjects	Intercept	1.3400	1.1576	
	NP1 Number	3.0850	1.7566	−0.08
items	Intercept	0.4070	0.6379	

Fixed Effects	***b***	***SE***	***z*-Value**	***p*-Value**
Intercept	3.0147	0.2868	10.51	<0.001
NP1 Number	−0.7626	0.5059	−1.51	0.132
Match	−1.9949	0.4411	−4.52	<0.001
Plurality Rating	0.0053	0.0156	0.34	0.732
Group	−1.9802	0.4887	−4.05	<0.001
NP1 Number × Match	1.1405	0.8795	1.30	0.195
NP1 Number × Plurality Rating	0.0522	0.0307	1.70	0.090
NP1 Number × Group	−0.0261	0.0306	−0.85	0.393
Match × Plurality Rating	0.0544	0.9414	0.06	0.954
Match × Group	−1.8711	0.8632	−2.17	0.030
Plurality Rating × Group	−0.0469	0.0304	−1.55	0.122
NP1 Number × Match × Plurality Rating	0.0440	0.0611	0.72	0.472
NP1 Number × Match × Group	0.4576	1.7295	0.27	0.791
NP1 Number × Plurality Rating × Group	0.1146	0.0608	1.89	0.059
Match × Plurality Rating × Group	0.1580	0.0608	2.60	0.009
NP1 Number × Match × Plurality Rating × Group	−0.2544	0.1217	−2.09	0.037

Note. Formula in R: DV ~ 1 + NP1 NUMBER * MATCH * PLURALITY.RATING * GROUP + (1 + NP1 Number | participant) + (1 | target). NP1 NUMBER is coded as −0.5 for singular and 0.5 for plural. MATCH is coded as −0.5 for match and 0.5 for mismatch. GROUP is coded as −0.5 for L1 speakers and 0.5 for bilinguals. See Table S12 in Supplementary Materials for follow-up statistics.

Table 9. Performance in binary forced-choice experiment; logistic mixed-effects regression model, results for the two participant groups.

Fixed Effects	**L1 Speakers**				**Bilinguals**			
	b	***SE***	***z*-Value**	***p*-Value**	***b***	***SE***	***z*-Value**	***p*-Value**
Intercept	3.0147	0.2868	10.51	<0.001	3.0147	0.2868	10.51	<0.001
NP1 Number	−0.7898	0.8596	−0.92	0.358	−0.7354	0.4603	−1.60	0.110
Match	−1.0593	0.7763	−1.37	0.172	−2.9304	0.3937	−7.44	<0.001
Plurality Rating	0.0288	0.0294	0.98	0.327	−0.0181	0.0091	−1.99	0.046
NP1 Number × Match	0.9116	1.5515	0.59	0.557	1.3692	0.7802	1.76	0.079
NP1 Number × Plurality Rating	−0.0051	0.0585	−0.09	0.930	0.1095	0.0176	6.23	<0.001
Match × Plurality Rating	−0.1051	0.0585	−1.80	0.072	0.0529	0.0174	3.04	0.002
NP1 Number × Match × Plurality Rating	1.1012	0.5521	1.99	0.046	−0.0832	0.0345	−2.41	0.016

Note. Formula in R: DV ~ 1 + GROUP/(NP1 NUMBER * MATCH * PLURALITY.RATING) + (1 + NP1 NUMBER | participant) + (1 | target). NP1 NUMBER is coded as −0.5 for singular and 0.5 for plural. MATCH is coded as −0.5 for match and 0.5 for mismatch. See Table S13 in Supplementary Materials for follow-up statistics.

The two participant groups differed in the way that plurality ratings affected their response patterns, as evidenced by a significant four-way interaction between NP1 NUMBER, MATCH, PLURALITY RATING, and GROUP, in addition to several lower-level interactions involving the factor GROUP (see Table 8).

Follow-up analyses revealed the following. The L1 speaker group exhibited an interaction between NP1 NUMBER, MATCH, and PLURALITY RATING (see Table 9), suggesting differences in the extent to which PLURALITY RATING affected the choice of verb number for NP1-matching and NP1-mismatching subjects. For singular NP1 subjects, there was a main effect of MATCH and an interaction between MATCH and PLURALITY RATING. Following up on this interaction, PLURALITY RATING affected responses in the SG-Mismatch condition (with higher PLURALITY RATING values yielding fewer NP1-matching [i.e., more plural] responses), but not in the SG-Match condition. Moreover, PLURALITY RATING showed no significant effects on either PL-Match or PL-Mismatch sentence fragments. See Figure 3A for a visualization of these effects and Table S11 in Supplementary Materials for relevant statistics.

The bilingual group also showed an interaction between NP1 NUMBER, MATCH, and PLURALITY RATING (see Table 9). Sentence fragments with singular NP1s showed a main effect of PLURALITY RATING as well as an interaction between MATCH and PLURALITY RATING, while sentence fragments with plural NP1s showed no such interaction but only main effects of MATCH and of PLURALITY RATING. Follow-up analyses at the lowest level revealed that PLURALITY RATING affected both SG-Match and SG-Mismatch sentence fragments (with greater PLURALITY RATING values yielding fewer NP1-matching [i.e., more plural] responses), though the effect was greater for SG-Match items than for SG-Mismatch items. PLURALITY RATING furthermore affected both PL-Match and PL-Mismatch sentence fragments (with greater PLURALITY RATING yielding more NP1-matching [i.e., more plural] responses). See Figure 3B and Table S11.

To summarize, the two groups exhibited differential effects of PLURALITY RATING on response patterns: For the German L1 speakers, only responses to sentence fragments in the SG-Mismatch condition (but not in SG-Match, PL-Match, or PL-Mismatch) were affected by PLURALITY RATING; for the Turkish–German bilingual group, responses were affected by PLURALITY RATING in all four conditions. As expected, in all cases greater PLURALITY RATING was associated with a higher proportion of plural responses.

4. Discussion

Using a larger and more homogeneous group of Turkish–German bilinguals than Jessen et al. (2021) did, the present study examined optional subject–verb agreement and the extent to which conceptual number affects participants' choices of agreement controller. In an off-line sentence-completion task, German L1 speakers and early bilinguals were presented with German pseudo-partitives (*eine Tüte Nüsse* 'a bag of nuts') and were asked to choose between a singular or a plural verb form. We manipulated the number of the first NP, whether or not the first and the second NP matched in number, and the semantic category of the first NP (half container words, half measure words), with the assumption that container nouns are more likely than measure nouns to be analyzed as subject heads, and thus to determine number agreement (Smith et al. 2018). The study sought to answer the following research questions:

I. (a) When computing agreement with German pseudo-partitives, do speakers use the first (NP1) or the second noun phrase (NP2) as the agreement controller?
   (b) Do L1 speakers and early bilinguals differ in their preference for NP1 versus NP2 as the agreement controller?

II. (a) What is the role of conceptual number in agreement computation operationalized through different types of NP1 and notional-number ratings)?
   (b) Does the role of conceptual number in agreement computation differ for L1 speakers versus early bilinguals?

In what follows, we will discuss each of these questions in turn, followed by a brief discussion of the study's limitations and our conclusions.

4.1. Pseudo-Partitive Agreement in Native Speakers and Early Bilinguals

In our study, both the L1 speaker group and the early bilingual group preferred NP1 as the agreement controller, as evidenced by a main effect of NP1 NUMBER in both groups. That is, when the preamble contained a singular NP1 (SG-Match, SG-Mismatch: *a bowl of yogurt*, *a pound of nuts*), participants were more likely to choose a singular verb (vs. a plural verb), while the reverse was true for preambles with plural NP1s (PL-Match, PL-Mismatch: *four bowls of yogurt*, *three pounds of nuts*).

Importantly, however, the second noun (NP2) also appeared to affect agreement choices. For preambles in which NP1 and NP2 matched in number (SG-Match and PL-Match: *a bowl of yogurt*, *three pounds of nuts*), participants were more likely to choose a verb that matched the two NPs in number than in cases in which NP1 and NP2 mismatched (SG-Mismatch and PL-Mismatch: *four bowls of yogurt*, *one pound of nuts*), as evidenced by a main effect of MATCH. Notably, NP1 NUMBER and MATCH also interacted with one another, indicating an asymmetrical relationship between NP1 and NP2: while preambles with mismatching NP2s yielded a greater proportion of NP1-mismatching (i.e., NP2-

matching) responses across items, the effect was greater for preambles with singular NP1s and plural NP2s (SG-Mismatch) than for preambles with plural NP1s and singular NP2s (PL-Mismatch). In other words, preambles like *one pound of nuts* yielded more plural responses than preambles like *four bowls of yogurt* yielded singular responses.

Lastly, while the abovementioned effects were not qualified by interactions with the factor GROUP, a main effect of GROUP indicated an overall smaller likelihood for the bilingual group to choose a verb that matched NP1 in number.

Our results replicate findings by Jessen et al. (2021), who had examined the processing of pseudo-partitives with a smaller, less homogenous group, and using different tasks. Across both studies, we find evidence that L1 speakers and bilinguals alike are aware that German pseudo-partitives have two agreement options, but that both groups generally prefer NP1 as the agreement controller. The asymmetrical pattern of NP2 influence was likewise reported by Jessen and colleagues, and is more generally reminiscent of a similar asymmetry found in agreement attraction with L1 speakers and, perhaps less reliably, bilinguals (e.g., Eberhard 1997; Reifegerste et al. 2020; Staub 2009).

4.2. The Role of Conceptual Number in Pseudo-Partitive Agreement

In order to investigate the influence of conceptual number of the pseudo-partitive constructions on SVA, we manipulated the semantic type of NP1, with half being container nouns (*one bowl of berries*), which are argued to be represented as generally referring to actual objects in a speaker's mind, and the other half being measure words (*one pound of nuts*), which are more likely to be represented as quantifiers in speakers' minds (Smith et al. 2018).

One set of analyses ('Analysis II: Semantic Category') operationalized conceptual number as the binary factor SEMANTIC CATEGORY and examined whether an NP1 being a container word versus a measure word yields different agreement patterns. While this factor did not yield changes in significance for any of the effects between this and the previous set of analyses, we found that container nouns elicited overall more NP1-matching verb choices than measure nouns did, in line with Smith et al.'s (2018) findings for Singular-Mismatch items in English pseudo-partitives. SEMANTIC CATEGORY did not show interactions with the factor GROUP, suggesting no systematic differences between the two participant groups in the way that the type of NP1 affected agreement patterns.

Interestingly, however, when examining the response patterns for the two groups separately, differences emerged. While the L1 speaker group still showed no differences in response pattern on the basis of conceptual number (i.e., no interactions involving the factor SEMANTIC CATEGORY), the results for the bilingual group displayed significant interactions between NP1 NUMBER and SEMANTIC CATEGORY and between MATCH and SEMANTIC CATEGORY. Although these group differences were not reflected in higher-level interactions involving the factor GROUP, they may be suggestive of a potentially greater influence of conceptual number on the bilingual group's responses.

For another set of analyses ('Analysis III: Plurality Ratings'), the binary factor SEMANTIC CATEGORY was replaced by the continuous factor PLURALITY RATING. These ratings were computed for each item (separately for the two participant groups) on the basis of plurality ratings that participants gave after finishing the sentence-completion task for the same items they had just seen, and ranged between 0 and 1, with higher numbers denoting greater notional plurality. Unlike the factor SEMANTIC CATEGORY, analyses with PLURALITY RATING revealed a number of differences in response patterns as a function of this factor (i.e., interactions involving the factor PLURALITY RATING), as well as group differences (i.e., interactions involving the factor GROUP). For the L1 speaker group, the only condition to show effects of PLURALITY RATING were preambles in the SG-Mismatch condition (*one pound of nuts*). For the group of early bilinguals, on the other hand, responses to items in all four conditions were influenced by PLURALITY RATING. In all cases, these effects were in the expected direction, with higher plurality ratings being associated with a higher proportion of plural responses and vice versa. This difference in pattern between the groups was present at all levels, from a four-way interaction at the highest level to main effects at the lowest-level by-condition analyses (see Tables 8 and 9 in the paper, and Table S11 in Supplementary Materials). Thus, it appears that during agreement computation early bilinguals were more strongly affected by the notional plurality of the items than German L1 speakers.

One possible explanation for this group difference might be found in Hartsuiker and Barkhuysen's (2006) assertion that of the two proposed stages of agreement processing—conceptually-driven number marking and grammatically-driven number morphing—only the latter is vulnerable to limitations in cognitive resources. If this is true, then circumstances of limited resource availability could provide fertile grounds for the emergence of notional-number effects. Agreement computation may present such a situation for bilingual speakers, as it is argued to be more cognitively taxing for bilingual speakers than for speakers with only one native language (Brehm et al. 2022; Cunnings 2017; Foote 2011; Hopp 2006, 2010; Keating 2009; Lehtonen et al. 2023; McDonald 2006; McDonald and Roussel 2010; Sagarra and Herschensohn 2010). A number of empirical studies have reported greater correlations between domain-general cognitive abilities (e.g., nonverbal working memory or interference control) and performance in tasks tapping morphosyntactic knowledge and processing, including agreement computation, for bilinguals than for monolinguals (e.g., Gangopadhyay et al. 2016; Reifegerste et al. 2020; Wood et al. 2021). Similarly, during grammaticality judgment bilingual participants have been found to show greater neural activation in brain areas associated with executive control than monolingual participants (Prehn et al. 2018).

While we have no direct evidence that the task was indeed more effortful for the bilingual group than for the L1-speaker group in our study, participants' performance in the no-conflict conditions (SG-Match and PL-Match; *one bowl of yogurt*, *four bowls of berries*) point in this direction: when examining only these two conditions, in which either a singular or plural verb is normally required, we find significant group differences, indicating that L1 speakers were more likely than early bilinguals to choose verb forms that matched the two NPs in number (98.6% vs. 94.9% (see Table 3); $b = 1.8075$, $SE = 0.3705$, $z = 4.88$, $p < 0.001$). This group difference in performance in these two conditions (as well as the main effect of GROUP across experimental conditions, which we found in all sets of analyses) suggests that overall the group of early bilinguals may be more prone to ignoring grammatical cues for agreement, possibly due to being more vulnerable to cognitive resource limitations, and thus more susceptible to the influence of the conceptual number of the pseudo-partitive. The observed group difference is in line with the hypothesis that bilinguals weight grammatical cues less strongly in their second language in comparison to (functionally) monolinguals, assigning comparatively more weight to non-grammatical cues instead (e.g., Clahsen and Felser 2018; Cunnings 2017).

Moreover, the above line of reasoning could also explain the finding that L1 speakers' responses were affected by conceptual number only in the SG-Mismatch condition, which likely represented the most challenging condition due to a comparatively salient (qua plural-marking) NP2 competing with a singular NP1. In other words, these results suggest that conceptual number can exert particular effects on agreement computation in cognitively taxing situations, whether the nature of the difficulty is due to the individual or the item.

Our findings may at first glance seem surprising considering some previous research on the effects of conceptual number on SVA, which has reported no differences between monolingual and bilingual speakers (Foote 2010; Nicol et al. 2001; Nicol and Greth 2003) or smaller/absent conceptual-number effects for the latter group (Foote 2010; Hoshino et al. 2010; Wei et al. 2015). Recall, however, that the bilingual speakers in our study had been living in Germany and acquired both languages since birth. This means that our participants are comparable only with the participants in Nicol et al. (2001) and the early bilinguals in Foote (2010), leaving us with the question of why we found differences between our participant groups when these two studies did not. Note that ours and the above studies are not directly comparable, though. First, previous studies have focused on categorical agreement, while the present study examined optional agreement. Second, the above two studies examined speakers of English and Spanish, two languages that are relatively closely related and highly similar in terms of SVA computation. In contrast, the speakers in the present study grew up speaking German and Turkish, which are typologically more distant. Moreover, while German has obligatory SVA, marking both the subject and the verb for the same number feature, Turkish does not, and plural number is usually marked in only one place (e.g., through quantifiers). These crosslinguistic differences in SVA computation may have rendered agreement computation more cognitively taxing for the bilingual participants in our study (as compared to participants in previous studies), potentially giving rise to greater susceptibility to conceptual-number effects in this group.[6]

Importantly, we want to emphasize that if our explanation is along the right lines, we do not suggest that this is a phenomenon that is specific to bilinguals. Instead, we suggest that our findings may highlight a more general relationship between cognitive resources and speakers' reliance on notional number, such that relative resource limitations may result in the balance between conceptual and grammatical forces being weighted more heavily in favor of conceptual forces in bilinguals.[7] If this is the case, we may observe similar instances of using notional number as a "fall back" option when examining agreement computation in L1 speakers that are relatively constrained in their cognitive resources (e.g., in children or older adults).

4.3. Limitations and Future Directions

The study has several limitations, suggesting additional directions for future research. First, the data were collected via the internet, limiting the control that the experimenter has. While certain shortcomings of web-based data collection do not apply to this study (e.g., reaction times were not collected, decreasing the influence of technological or environmental factors that may introduce noise), we still have to rely on participants' self-reported information (e.g., regarding their native language). A lab-based replication of our findings would enhance their reliability.

Second, while we believe that Hartsuiker and Barkhuysen's (2006) claims regarding the greater vulnerability of grammatically-driven number morphing (versus conceptually-driven number marking) provide a promising account for our findings, these explanations remain speculative. Therefore, future studies that directly probe the impact of cognitive-resource limitations on conceptual constraints during agreement computation (e.g., by examining speakers' cognitive abilities, or by introducing a cognitive load manipulation) are clearly necessary to validate this hypothesis and to gain a more complete understanding of the mechanisms underlying the observed effect. Along similar lines, if this explanation is correct, we suggest that examining this topic in less-commonly studied populations typically associated with reduced cognitive abilities, such as older adults (e.g., Christianson et al. 2006), would provide additional insight into the role of cognitive limitations during agreement computation.

Third, our findings regarding the role of conceptual number for SVA may have important methodological implications. While the set of analyses that operationalized conceptual number as a binary factor ('Analysis II: Semantic Category') did not find reliable effects of this variable on response patterns (i.e., interactions involving the factor SEMANTIC CATEGORY), intriguing effects and group differences emerged when conceptual number was operationalized as a continuous factor on the basis of participant ratings ('Analysis III: Plurality Ratings'). This highlights the possibility that norms from "real people" (and ideally from the very individuals that were tested, or at least from individuals stemming from the same population) may be more informative than relying on simple binary categories, especially when examining the effects of complex linguistic properties on language processing. Additionally, such norms can allow for the use of continuous variables in statistical analyses, which both reflect the gradience of a given measure of interest (and should thus be a more reliable reflection of a gradient underlying concept that the measure is supposed to represent), and also facilitates the detection of effects of underlyingly continuous concepts, due to an increase in power (Altman and Royston 2006; Royston et al. 2006).

Thus, future studies examining the role of linguistic properties on language processing may want to strive for the inclusion of rating-based measures (if possible, collected from the participants themselves), which can help uncover effects that may be muddied when assessed as categorical factors. This approach has been successfully employed in the field of lexical processing (see, e.g., Balota et al. (2004), for subjective frequency measures, and Reifegerste et al. (2022), for participant-based age-of-acquisition measures), and could also prove fruitful for sentence-processing research.

Finally, our results have potential theoretical implications as they indicate that the relative strengths of conceptual and grammatical constraints may vary. To capture this variability formally, theoretical models are required that can handle gradient grammatical constraints, such as the Gradient Symbolic Computation framework (Smolensky et al. 2014) or Villata and Tabor's (2022) self-organized processing model. Future studies might want to include a modeling component (as did Jessen et al. 2021; Smith et al. 2018, for agreement) and examine the models' suitability for capturing other linguistic phenomena that involve interacting grammatical and conceptual constraints.

4.4. Conclusions

The present study examined subject–verb agreement in L1 speakers of German and Turkish–German early bilinguals. While most studies have focused on cases of agreement errors, here we examined agreement computation for pseudo-partitives (*eine Tüte Nüsse* 'one bag of nuts'), in which agreement with either the first NP or the second NP is considered licit. Both participant groups showed awareness of the two agreement options but preferred the first NP as the agreement controller, replicating previous findings. Both groups also demonstrated sensitivity to the first NP's semantic category, such that container nouns were more likely to control agreement than measure nouns. Moreover, it appears that the early bilinguals were more strongly affected than the L1 group by the notional plurality of the pseudo-partitive. While our manipulation of semantic category of the first NP (whether it was a container or represented a measure) did not result in reliable group differences in response patterns, plurality ratings ('Does this denote one thing or multiple things?') yielded differential effects on agreement choices for the two participant groups. The bilingual group's responses were strongly impacted by notional plurality, with increases in an item's notional plurality yielding more plural verb form choices across all conditions. Responses by the L1 group, in contrast, were affected by notional plurality only in the most difficult experimental condition.

We argue that these findings might be explained by claims that during agreement computation, conceptually-driven mechanisms (including effects of notional number) are less vulnerable to cognitive-resource limitations than grammatically-driven mechanisms, and that such limitations may affect bilinguals' agreement computation more strongly than is the case for L1 speakers. This may then in turn yield larger effects of notional number on bilinguals' agreement choices than L1 speakers' choices, whose responses are affected by notional number only in the most taxing situations.

Supplementary Materials: The complete set of materials, analysis code, supplementary tables and figures, and the raw data are available on OSF at https://osf.io/73twy/ (accessed on 20 May 2023).

Author Contributions: Conceptualization, C.F.; Methodology, C.F.; Formal Analysis, J.R.; Investigation, A.G.; Resources, A.G. and C.F.; Data Curation, J.R. and A.G.; Writing—Original Draft Preparation, J.R. and C.F.; Writing—Review & Editing, J.R., A.G., and C.F.; Visualization, J.R.; Supervision, C.F.; Project Administration, C.F.; Funding Acquisition, C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Potsdam (Reference number 37/2011).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available on the study's OSF repository at this link: https://osf.io/73twy/ (accessed on 20 May 2023).

Acknowledgments: We are grateful to Anna Jessen for providing advice and guidance on the design and implementation of this research, to João Veríssimo for assistance with the statistical analyses, and to the individuals who participated in the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 The word *eine* (masculine and neuter *ein*) can be interpreted as either the indefinite article 'a' or the numeral 'one.' Importantly, in either case it denotes a singular entity. For brevity, we will use the translation 'one' throughout the manuscript.

2 The statistics reported in the paper are not designed to answer this specific question.

3 For example, measures that are often encountered as containers of that size might yield container readings. Speakers in Europe may more readily interpret a liter of milk as a container of milk of that size than a gallon of milk, while the reverse might be the case for speakers in the U.S. Conversely, certain measure terms are derived from container words (e.g., *Tonne* in German, which means both the container 'barrel' and the measure 'tonne'). Some words may even equally likely represent a container or a measure (e.g., *cup* in English), or their interpretation may depend on the NP2 (e.g., *a cup of tea* vs. *a cup of flour*).

4 The ages-of-arrival in Germany for these three participants were five years, two years, and less than one year of age, respectively.

5 In the interest of conciseness and clarity, statistics for follow-up analyses to interactions at the lowest level are presented in Supplementary Materials (see table notes).

6 It is worth noting that the bilingual group in our study likely differed from the L1 speakers regarding the amount and nature of (especially early-life) exposure to German, as German was not necessarily the (only) language of the household, or they may have been exposed to a different variety of German as compared to the L1 speakers. This point is particularly relevant considering experimental work indicating that language experience can shape agreement preferences via long-term statistical learning of distributional patterns (Haskell et al. 2010). However, it is presently not clear why the language environment our bilingual group was exposed to during childhood might be expected to show distributional properties that render this group more prone to using notional number during agreement computation than the L1 speaker group.

7 We are thankful to an anonymous reviewer for this phrasing.

References

Alarcón, Irma. 2021. Adjectival and verbal agreement in the oral production of early and late bilinguals: Fluency, complexity, and integrated knowledge. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 34: 371–401. [CrossRef]

Altman, Douglas G., and Patrick Royston. 2006. The cost of dichotomising continuous variables. *British Medical Journal* 332: 1080. [CrossRef] [PubMed]

Balota, David A., Michael J. Cortese, Susan D. Sergent-Marshall, David H. Spieler, and Melvin J. Yap. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133: 283–316. [CrossRef] [PubMed]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef] [PubMed]

Bates, Douglas M., Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Bauer, Laurie. 1994. Watching English change: An introduction to the study of linguistic change in Standard Englishes in the twentieth century. *In Watching English Change. London: Routledge*. [CrossRef]

Bock, Kathryn, Kathleen M. Eberhard, J. Cooper Cutting, Antje S. Meyer, and Herbert J. Schriefers. 2001. Some attractions of verb agreement. *Cognitive Psychology* 43: 83–128. [CrossRef]

Bock, Kathryn, and Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23: 45–93. [CrossRef]

Brehm, Laurel, and Kathryn Bock. 2013. What counts in grammatical number agreement? *Cognition* 128: 149–69. [CrossRef]

Brehm, Laurel, Pyeong Whan Cho, Paul Smolensky, and Matthew A. Goldrick. 2022. PIPS: A Parallel Planning Model of Sentence Production. *Cognitive Science* 46: e13079. [CrossRef]

Chen, Lang, Hua Shu, Youyi Liu, Jingjing Zhao, and Ping Li. 2007. ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition* 10: 161–74. [CrossRef]

Christianson, Kiel, Carrick C. Williams, Rose T. Zacks, and Fernanda Ferreira. 2006. Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes* 42: 205–38. [CrossRef] [PubMed]

Clahsen, Harald, and Claudia Felser. 2018. Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition* 40: 693–706. [CrossRef]

Cunnings, Ian. 2017. Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20: 659–78. [CrossRef]

Dekeyser, Xavier. 1975. *Number and Case Relations in 19th Century British English: A Comparative Study of Grammar and Usage*. Amsterdam: De Nederlandsche Boekhandel.

Depraetere, Ilse. 2003. On verbal concord with collective nouns in British English. *English Language and Linguistics* 7: 85–127. [CrossRef]

Eberhard, Kathleen M. 1997. The marked effect of number on subject–verb agreement. *Journal of Memory and Language* 36: 147–64. [CrossRef]

Eberhard, Kathleen M., J. Cooper Cutting, and Kathryn Bock. 2005. Making syntax of sense: Number agreement in sentence production. *Psychological Review* 112: 531–59. [CrossRef]

Foote, Rebecca. 2010. Age of acquisition and proficiency as factors in language production: Agreement in bilinguals. *Bilingualism: Language and Cognition* 13: 99–118. [CrossRef]

Foote, Rebecca. 2011. Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics* 32: 187–220. [CrossRef]

Gangopadhyay, Ishanti, Meghan M. Davidson, Susan Ellis Weismer, and Margarita Kaushanskaya. 2016. The role of nonverbal working memory in morphosyntactic processing by school-aged monolingual and bilingual children. *Journal of Experimental Child Psychology* 142: 171–94. [CrossRef]

Gillespie, Maureen, and Neal J. Pearlmutter. 2011. Effects of semantic integration and advance planning on grammatical encoding in sentence production. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Edited by Laura A. Carlson, Christoph Hoelscher and Thomas F. Shipley. Austin: Cognitive Science Society, pp. 1625–30.

Grestenberger, Laura. 2015. Number marking in German measure phrases and the structure of pseudo-partitives. *Journal of Comparative Germanic Linguistics* 18: 93–138. [CrossRef]

Grosjean, François. 2001. The bilingual's language modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet L. Nicol. Hoboken: Wiley Blackwell, pp. 1–22. [CrossRef]

Hacohen, Aviya, and Jeannette Schaeffer. 2007. Subject realization in early Hebrew/English bilingual acquisition: The role of crosslinguistic influence. *Bilingualism: Language and Cognition* 10: 333–44. [CrossRef]

Hartsuiker, Robert J., and Pashiera Barkhuysen. 2006. Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes* 21: 181–204. [CrossRef]

Haskell, Todd R., Robert Thornton, and Maryellen C. MacDonald. 2010. Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition* 114: 151–64. [CrossRef] [PubMed]

Hennig, Mathilde, Jan G. Schneider, Ralf Osterwinter, and Anja Steinhauer. 2016. Duden—Das Wörterbuch der sprachlichen Zweifelsfälle. In *Der Duden in zwölf Bänden—Das Standardwerk der deutschen Sprache*. Edited by Mathilde Hennig. Berlin: Dudenverlag, Band 9.

Hopp, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22: 369–97. [CrossRef]

Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120: 901–31. [CrossRef]

Hoshino, Noriko, Paola E. Dussias, and Judith F. Kroll. 2010. Processing subject–verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition* 13: 87–98. [CrossRef] [PubMed]

Jackson, Carrie N., Elizabeth Mormer, and Laurel Brehm. 2018. The production of subject-verb agreement among Swedish and Chinese second language speakers of English. *Studies in Second Language Acquisition* 40: 907–21. [CrossRef]

Jaeger, Christoph. 1992. *Probleme der syntaktischen Kongruenz*. Tübingen: Niemeyer.

Jessen, Anna, Lara Schwarz, and Claudia Felser. 2021. Gradience in subject-verb number agreement: Can bilinguals tune in? *Applied Psycholinguistics* 42: 1523–51. [CrossRef]

Juul, Arne. 1975. *On Concord of Number in Modern English*. Copenhagen: Nova.

Kandel, Margaret, Cassidy R. Wyatt, and Colin Phillips. 2022. Agreement attraction error and timing profiles in continuous speech. *Glossa Psycholinguistics* 1: 1–46. [CrossRef]

Keating, Gregory D. 2009. Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59: 503–35. [CrossRef]

Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.

Lardiere, Donna. 1998. Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research* 14: 359–75. [CrossRef]

Lehtonen, Minna, Valantis Fyndanis, and Jussa Jylkkä. 2023. The relationship between bilingual language use and executive functions. *Nature Reviews Neuroscience*. [CrossRef]

Levin, Magnus. 2001. *Agreement with Collective Nouns in English*. Lund: Lund University.

Levin, Magnus. 2006. Collective nouns and language change. *English Language and Linguistics* 10: 321–43. [CrossRef]

Liedtke, Ernst A. 1910. *Die numerale Auffassung der Kollektiva im Verlaufe der englischen Sprachgeschichte*. Königsberg: Karg and Manneck.

Lorimor, Heidi. 2007. *Conjunctions and Grammatical Agreement*. Champaign: University of Illinois at Urbana-Champaign.

Lorimor, Heidi, Carrie N. Jackson, Katharina Spalek, and Janet G. van Hell. 2016. The impact of notional number and grammatical gender on number agreement with conjoined noun phrases. *Language, Cognition and Neuroscience* 31: 646–61. [CrossRef]

Mallinson, Graham, and Barry Blake. 1981. *Language Typology: Crosslinguistic Studies in Syntax*. Amsterdam: North Holland.

McDonald, Janet L. 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55: 381–401. [CrossRef]

McDonald, Janet L., and Christine C. Roussel. 2010. Past tense grammaticality judgment and production in non-native and stressed native English speakers. *Bilingualism: Language and Cognition* 13: 429–48. [CrossRef]

Nicol, Janet L., and Delia Greth. 2003. Production of subject-verb agreement in Spanish as a second language. *Experimental Psychology* 50: 196–203. [CrossRef]

Nicol, Janet L., Matthew Teller, and Delia Greth. 2001. The production of verb agreement in monolingual, bilingual, and second language speakers. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet L. Nicol. Oxford: Blackwell, pp. 117–33.

Prehn, Kristin, Benedikt Taud, Jana Reifegerste, Harald Clahsen, and Agnes Flöel. 2018. Neural correlates of grammatical inflection in older native and second-language speakers. *Bilingualism: Language and Cognition* 21: 1–12. [CrossRef]

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartnik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman House. [CrossRef]

Reifegerste, Jana, Lauren E. Russell, David A. Balota, Daniel Lipscomb, George Luta, Marcus Meinzer, Michael D. Rugg, Kyle F. Shattuck, Peter E. Turkeltaub, Ashley S. VanMeter, and et al. 2022. What's that word again? The contribution of the hippocampus to word-finding declines in aging. Paper Presented at the 2022 Cognitive Aging Conference, Atlanta, GA, USA, April 7–10.

Reifegerste, Jana, Rebecca Jarvis, and Claudia Felser. 2020. Effects of chronological age on native and nonnative sentence processing: Evidence from subject-verb agreement in German. *Journal of Memory and Language* 111: 104083. [CrossRef]

Royston, Patrick, Douglas G. Altman, and Willi Sauerbrei. 2006. Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* 25: 127–41. [CrossRef] [PubMed]

Sagarra, Nuria, and Julia Herschensohn. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120: 2022–39. [CrossRef]

Sato, Mikako, and Claudia Felser. 2010. Sensitivity to morphosyntactic violations in English as a second language. *Second Language* 9: 101–18.

Scontras, Gregory. 2014. *The Semantics of Measurement*. Cambridge: Harvard University.

Selkirk, Elisabeth O. 1977. Some remarks on noun phrase structure. In *Studies in Formal Syntax*. Edited by Adrian Akmajian, Peter Culicover and Thomas A. Wasow. New York: Academic Press.

Shibuya, Mayumi, and Shigenori Wakabayashi. 2008. Why are L2 learners not always sensitive to subject-verb agreement? *EUROSLA Yearbook* 8: 235–58. [CrossRef]

Siemund, Rainer. 1995. "For who the bell tolls"—Or why corpus linguistics should carry the bell in the study of language change in present-day English. *Arbeiten aus Anglistik und Amerikanistik* 20: 351–77.

Smith, Garrett, Julie Franck, and Whitney Tabor. 2018. A self-organizing approach to subject–verb number agreement. *Cognitive Science* 42: 1043–74. [CrossRef]

Smolensky, Paul, Matthew Goldrick, and Donald Mathis. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science* 38: 1102–38. [CrossRef]

Solomon, Eric S., and Neal J. Pearlmutter. 2004. Semantic integration and syntactic planning in language production. *Cognitive Psychology* 49: 1–46. [CrossRef]

Staub, Adrian. 2009. On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language* 60: 308–27. [CrossRef]

Staub, Adrian. 2010. Response time distributional evidence for distinct varieties of number attraction. *Cognition* 114: 447–54. [CrossRef] [PubMed]

Stickney, Helen. 2009. *The Emergence of DP in the Partitive Structure*. Amherst: University of Massachusetts Amherst.

VanPatten, Bill, Gregory D. Keating, and Michael J. Leeser. 2012. Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism* 2: 109–40. [CrossRef]

Veenstra, Alma, Daniel J. Acheson, and Antje S. Meyer. 2014a. Keeping it simple: Studying grammatical encoding with lexically reduced item sets. *Frontiers in Psychology* 5: 783. [CrossRef]

Veenstra, Alma, Daniel J. Acheson, Kathryn Bock, and Antje S. Meyer. 2014b. Effects of semantic integration on subject–verb agreement: Evidence from Dutch. *Language, Cognition and Neuroscience* 29: 355–80. [CrossRef]

Vigliocco, Gabriella, Brian Butterworth, and Carlo Semenza. 1995. Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language* 34: 186–215. [CrossRef]

Vigliocco, Gabriella, Brian Butterworth, and Merrill F. Garrett. 1996a. Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition* 61: 261–98. [CrossRef]

Vigliocco, Gabriella, Robert J. Hartsuiker, Gonia Jarema, and Herman H. J. Kolk. 1996b. One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes* 11: 407–42. [CrossRef]

Villata, Sandra, and Julie Franck. 2020. Similarity-based interference in agreement comprehension and production: Evidence from object agreement. *Journal of Experimental Psychology: Learning Memory and Cognition* 46: 170–88. [CrossRef]

Villata, Sandra, and Whitney Tabor. 2022. A self-organized sentence processing theory of gradience: The case of islands. *Cognition* 222: 104943. [CrossRef]

Wegerer, Martina. 2012. *Die Numeruskongruenz von Subjekt und finitem Verb im Deutschen: Untersuchung der grammatischen Entscheidungsprozeduren bei zweifelhafter Kongruenzrelationen*. Wien: University of Vienna.

Wei, Xiaoyan, Baoguo Chen, Lijuan Liang, and Susan Dunlap. 2015. Native language influence on the distributive effect in producing second language subject–verb agreement. *Quarterly Journal of Experimental Psychology* 68: 2370–83. [CrossRef]

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wood, Taffeta, Amy S. Pratt, Kathleen Durant, Stephanie McMillen, Elizabeth D. Peña, and Lisa M. Bedore. 2021. Contribution of nonverbal cognitive skills on bilingual children's grammatical performance: Influence of exposure, task type, and language of assessment. *Languages* 6: 36. [CrossRef]