Person of Interest: Experimental Investigations into the Learnability of Person Systems

Mora Maldonado, Jennifer Culbertson

Person of Interest: Experimental Investigations into the Learnability of Person Systems

Mora Maldonado
Jennifer Culbertson

Person systems convey the roles entities play in the context of speech (e.g., speaker, addressee). As with other linguistic category systems, not all ways of partitioning the person space are equally likely crosslinguistically. Different theories have been proposed to constrain the set of possible person partitions that humans can represent, explaining their typological distribution. This article introduces an artificial language learning methodology to investigate the existence of universal constraints on person systems. We report the results of three experiments that inform these theoretical approaches by generating behavioral evidence for the impact of constraints on the learnability of different person partitions. Our findings constitute the first experimental evidence for learnability differences in this domain.

Keywords: person systems, pronouns, artificial language learning, linguistic universals, semantics

1 Introduction

Person systems—exemplified by pronominal paradigms (e.g., *I*, *you*, *she*)—categorize entities as a function of their role in the context of speech: the speaker, the addressee, and others, who play no active role in the conversation. As in other semantic domains, it has long been observed that person systems exhibit what appears to be constrained variation across languages: some person systems are frequent, while others are rare or do not occur at all (Cysouw 2003, Baerman, Brown, and Corbett 2005).[1]

This article has profited from the comments and suggestions of Daniel Harbour and Peter Ackema, as well as of two anonymous *LI* reviewers. We also wish to thank the audiences of The Alphabet of Universal Grammar at the British Academy for interesting discussion and Estudio Chirrikenstein for the beautiful illustrations. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 757643).

[1] This kind of typological tendencies has been seen in other semantic domains, involving both content and logical words. For example, crosslinguistic regularities have been argued to provide evidence for a universal basis for color categorization, reflecting properties of the human perceptual system (Kay and Regier 2007, Gibson et al. 2017, Zaslavsky et al. 2019). Similar arguments have been made to explain the distribution of kinship systems across languages (Kemp and Regier 2012). Relatedly, the study of logical words has also revealed that connectives and quantifiers found in natural languages only cover a very small subset of all possible meanings, indicating the existence of semantic universals (Barwise and Cooper 1981, Piantadosi, Tenenbaum, and Goodman 2016, Chemla, Buccola, and Dautriche 2019, Steinert-Threlkeld and Szymanik to appear).

Linguistic Inquiry, Volume 53, Number 2, Spring 2022
295–336

https://doi.org/10.1162/ling_a_00406

Typological regularities of this sort have led linguists to propose universal constraints on possible person systems (Silverstein 1976, Ingram 1978, Noyer 1992, Harley and Ritter 2002, Bobaljik 2008, Harbour 2016, Ackema and Neeleman 2018). Such constraints are often conceived of (either implicitly or explicitly) as reflecting characteristics of human linguistic capacity that have consequences for learning: specifically, they are assumed to delimit the space of hypotheses entertained by the learner (Chomsky 1965; but see, e.g., Perfors, Tenenbaum, and Regier 2011). However, while person has been extensively investigated from formal and typological perspectives, the link between hypothesized universal constraints on person systems and learnability remains largely unexplored (though see Nevins, Rodrigues, and Tang 2015 for an artificial learning approach, and Brown 1997, Hanson 2000, Hanson, Harley, and Ritter 2000, Moyer et al. 2015 on acquisition). In this article, we introduce an artificial language learning methodology to investigate the existence of universal constraints on person systems. First, we summarize some additional theoretical background from which we will derive a number of specific predictions regarding the learnability of these systems.

1.1 The Person Space

As mentioned above, there are three conversational roles typically delimited in the person space: the speaker, the addressee, and others, who are not active participants in the conversation. Following standard assumptions, we represent them as *i*, *u*, and *o*, respectively (e.g., Harbour 2016). From this ontology, we obtain seven logically possible person categories or "metapersons": *i*, *io*, *iu*, *iuo*, *u*, *uo*, *o* (Sokolovskaja 1980, Bobaljik 2008, Sonnaert 2018). Research on the typological distribution of person systems, however, has found evidence that only four of these are grammaticalized as person categories: first exclusive (*i*), first inclusive (*iu*), second (*u*), and third (*o*). This asymmetry can be directly captured by assuming that the speaker and addressee are unique—there are no forms that express uniquely multiple speakers, or uniquely multiple addressees—but there can be an undefined number of others (following Harbour 2016).[2] The meanings expressed by the unattested combinations (*io*, *iuo*, *uo*) can be captured as the interaction between person and number; the four core person categories can each be pluralized by adding extra others (see footnote 2; Boas 1911). Table 1 illustrates these person and number categories (number expressed by the presence or absence of the subscript *o*). To account for the above-mentioned restrictions on person categories, theories of the person space have traditionally posited two primitive binary features: [±speaker] and [±addressee] (or other equivalent notations; Silverstein

[2] This assumption is not trivial. As soon as multiple speakers and addressees are allowed in the ontology, each logically available combination of the three entities—*i*, *io*, *iu*, *iuo*, *u*, *uo*, *o*—should count as a possible person category independent of number. For example, one form would refer to the speaker alone (*i*) and another form to the speaker plus someone else (*io*), each with a plural alternative (*ii* vs. *ioo*). However, this contrast is never grammaticalized: no language distinguishes between plural expressions referring to multiple speakers/addressees and expressions referring to the speaker/addressee plus others. Indeed, this has been formulated as a typological universal: pluralities containing participants (speakers or addressees) are never formally distinguished from pluralities containing others. This universal has been extensively discussed in the literature on person systems (Greenberg 1988, Cysouw 2003, Bobaljik 2008, Wechsler 2010), but is not directly investigated here.

Table 1
Four-person system

	Attested categories	Binary-features account
1 EXCL	i_o	[+speaker −addressee]
1 INCL	iu_o	[+speaker +addressee]
2	u_o	[−speaker +addressee]
3	o_o	[−speaker −addressee]

Table 2
Example personal pronoun systems

	Ilocano		Mandarin		English	
	MIN	AUG	SG	PL	SG	PL
1 EXCL	co	mi	wo	women	I	we
1 INCL	ta	tayo	—	zamen	—	we
2	mo	yo	ni	nimen	you	you
3	na	da	ta	tamen	she/he/they	they

1976, Ingram 1978, Noyer 1992, Bobaljik 2008). As shown in table 1, the interaction between these two binary features predicts all and only the four attested categories.

An example of a language with a four-way person distinction in its pronominal system is Mandarin. As shown in table 2, each person category is expressed by a different pronoun, with additional morphology marking whether the referent is singular or plural. This system has seven forms total, since the inclusive is inherently nonsingular (it necessarily refers to both the speaker and the addressee) and thus always features plural morphology in Mandarin. Another example is Ilocano, which differs from Mandarin in that it makes a minimal/augmented number distinction rather than singular/plural. This system has eight forms, including two distinct inclusive forms, one minimal (*ta* = speaker and addressee only) and the other augmented (*tayo* = speaker, addressee, and others).

In many other languages, the meaning space is partitioned such that not all possible person and/or number categories are expressed by distinct forms. Such languages exhibit homophony. For example, *inclusive* languages differ from *noninclusive* languages (terminology from Daniel 2005) like English, where there is homophony between the first and inclusive persons (in addition to homophony between second person singular and plural; see table 2). Feature-based accounts of person often derive a restricted set of partitions of the person and number space as defined by the presence or absence of homophony among cells in the space. Such theories only derive homophony patterns by contrast neutralization or underspecification: a distinction that is made available by the grammar might not be active in a specific language (Halle and Marantz 1994, Harbour 2008, Harley 2008, Pertsova 2011). Specifically, the set of features in table 1 straightforwardly derives three person homophony patterns on the basis of which contrasts are left underspec-

ified ([±speaker], [±addressee], or both). For example, neutralizing the [±addressee] feature would generate syncretism between 1 EXCL and 1 INCL categories (grouped as [+speaker]), on the one hand, and between 2 and 3 (grouped as [−speaker]), on the other. Other feature-based homophony patterns can also be derived from this system by restricting underspecification to specific natural classes. For example, the aforementioned clusivity distinction is lost when two meanings that share the feature [+speaker] (*i* and *iu*) become indistinguishable (e.g., English). That is, the grouping of 1 EXCL and 1 INCL categories relies on them belonging to the same natural class [+speaker]. These kinds of feature-based patterns are often referred to as systematic homophony.

However, many different homophony patterns have been documented both within and across languages (Zwicky 1977, Corbett, Brown, and Baerman 2002, Cysouw 2003, Baerman, Brown, and Corbett 2005, Baerman and Brown 2013), not all of which can be derived by feature neutralization. In some cases, two or more meanings that do not share any feature are nevertheless expressed by the same form in a given language. This so-called accidental or random homophony is therefore not described in terms of contrast neutralization, as the targeted meanings do not belong to the same natural class (e.g., defined by the features in table 1). Partitions that are not derivable by a theory are often assumed to arise through historical accident, target mainly individual paradigms in a given language, and may be marginal typologically (Halle and Marantz 1994, Pertsova 2011, Sauerland and Bobaljik 2013; but see Cysouw 2003). But whether the typological evidence accords with this prediction is not always clear.[3]

While estimating the frequency distribution over partitions of the person space is complex, a number of theories have been developed to make more fine-grained predictions about possible homophony patterns. Harley and Ritter (2002) put forward a universal feature geometry for person based on three privative features, as illustrated in figure 1. This derives the same set of four person categories as the binary-features account in table 1 but also establishes hierarchical relations between them, in order to make more accurate predictions about the typological frequency of homophony patterns (see also Béjar 2003, McGinnis 2005, Cowper and Hall 2009 for similar approaches). For example, this system derives homophony between 1 EXCL, 1 INCL, and 2 categories and therefore predicts this to arise systematically.

In a similar vein, Harbour (2016) posits a theory specifically designed to capture the robust typological generalization in (1), also known as Zwicky's observation (see Zwicky 1977: 717–719).

[3] As it turns out, determining the crosslinguistic frequency of different partitions of the person (and number) space is not straightforward. It crucially depends on the specific assumptions made about how to count different systems. Some authors (Cysouw 2003, Sauerland and Bobaljik 2013) include individual person-marking paradigms (e.g., verbal agreement and pronominal systems) within a single language. Using this metric for counting, the frequency distribution across languages is extremely skewed: in Cysouw's dataset, of the 4,140 possible partitions of an eight-cell person/number space, only 61 are attested (calculated over 265 paradigms). A slightly different approach that also counts paradigms is given by Baerman, Brown, and Corbett (2005) and Baerman and Brown (2013). Still another possibility is to count on the basis of an abstract notion of person partition, across all the paradigms in a given language. For example, Harbour (2016) proposes a superposition technique: a distinction is neutralized in a language if and only if it is inactive across all paradigms. Under this superposition approach, there are 15 possible partitions of a four-cell person space, but only 5 are attested typologically (with one marginal exception; see appendix A2.1 in Harbour 2016).

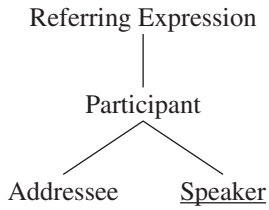


Figure 1
Feature geometry account in Harley and Ritter 2002. The underlined daughter node Speaker represents the default interpretation of the bare Participant node. Languages may or may not allow both dependent nodes to be specified together. If they do, the inclusive category is obtained when Speaker and Addressee are present simultaneously.

(1) Languages that do not have a dedicated phonological form for an inclusive person ('you and us') always assimilate it into the first plural person ('us') and never into the second ('you') or third ('them').

Harbour makes use of two binary features, [±author] and [±participant]. While the features themselves denote lattices containing referential entities (*i*, *u*, *iu*, *o*), the values of the features are modeled as complementary operations on lattices. Because features are similar to functions, languages can differ not only in which features are active, but also in the order of feature composition. A similar approach is taken by Ackema and Neeleman (2018; see also Ackema and Neeleman 2013), with the main goal of accounting for the typological tendency described in Baerman, Brown, and Corbett 2005 and summarized in (2).[4]

(2) Languages that feature homophony between first (inclusive and exclusive) and second plural pronouns ('us' = 'you') and between second and third ('you' = 'them') are far more frequent than those instantiating first-third homophony.

Each of these approaches (which we will discuss in more detail below) introduces different theoretical apparatus to capture these typological observations. There are, however, a number of obvious limitations that make basing theories *exclusively* on typological evidence problematic.

1.2 From Typology to Learning

There is extensive literature now documenting (and in some cases proposing solutions to) the problems posed by typological data samples (for an excellent overview, see Cysouw 2005). For

[4] As an anonymous reviewer points out, the approach taken by Baerman, Brown, and Corbett (2005) is a conservative one: their typological counts are restricted to cases where whole forms are identical, disregarding morpheme syncretism. This might result in the exclusion of morphemes that are syncretic but occur in different combinations. Given that person features are often instantiated in morphemes rather than in whole words, the typological tendency stated in (2) should be considered with some caution.

one, such data are generally sparse, and in many cases the number of languages behind a given typological generalization is quite small. For instance, the largest sample of person/number paradigms, from Cysouw 2003, includes only around 200 languages. Sparse data lead to unreliable estimates of relative frequency, particularly in the tail of the distribution. For example, it is not possible to confidently conclude on the basis of small samples that a given partition is impossible (e.g., see Piantadosi and Gibson 2014, and also footnote 3).

Moreover, typological data are massively confounded: there are many factors that shape typological distributions (e.g., historical accidents, genetic relations between languages, facts about diachrony; see, e.g., Pagel, Atkinson, and Meade 2007, Bickel 2008, Cysouw 2010, Dunn et al. 2011), only a subset of which are relevant for building theories of the generative capacity of the linguistic system. The immediate consequence is that these data sources typically cannot be used to argue for a causal link between the cognitive or linguistic system and particular features of language (e.g., see discussion in Culbertson 2012, Piantadosi and Gibson 2014, Ladd, Roberts, and Dediu 2015).

As a response to these general issues—which are relevant for typological data in any domain—there has been an increasing attempt to bring behavioral data on learning to bear on linguistic theories. Specifically, artificial language learning experiments have now been used to link typological universals to human learning and inference in a number of domains including phonology (e.g., Wilson 2006, Moreton 2008, White 2017, Martin and Peperkamp 2020), syntax (e.g., Culbertson, Smolensky, and Legendre 2012, Tabullo et al. 2012, Culbertson and Adger 2014, Martin et al. 2019), morphology (e.g., Fedzechkina, Jaeger, and Newport 2012, Saldana, Oseki, and Culbertson 2019), and lexical categorization (Carstensen et al. 2015, Chemla, Buccola, and Dautriche 2019); for reviews, see Culbertson 2012, to appear, Moreton and Pater 2012.

The present study uses artificial language learning experiments to test a set of predictions derived from the feature-based theories of person described above. By incorporating this new source of data, we can corroborate—or not—the universal constraints on person partitions hypothesized on the basis of typological data. The article proceeds as follows. In Experiment 1, we establish an experimental setup to test some basic assumptions of feature-based systems, including whether systematic and random homophony are treated differently by learners acquiring a new person system. In Experiment 2, we investigate whether the universal typological tendency known as Zwicky's observation is supported by a learnability advantage, as predicted by Harbour (2016). Finally, in Experiment 3 we explore potential asymmetries in the learnability of different partitions of first, second, and third person categories, as predicted by different theories (e.g., Harley and Ritter 2002, Ackema and Neeleman 2018).

2 Experiment 1: Something about *Us*

2.1 Introduction

While different theories of person have hypothesized different inventories of features to constrain the person space, they all assume the features to be *universal*, that is, part of the human linguistic capacity (Harley and Ritter 2002, Bobaljik 2008, Harbour 2016, among many others). This as-

sumption predicts that all things being equal, humans should have access to, and therefore be able to learn, feature distinctions that are not at play in their native language (first prediction). Feature-based theories also predict that learners should be sensitive to natural classes as defined by feature structure: categories that share a feature should be more readily mapped onto the same phonological form. These systematic homophony patterns are predicted to be (easily) learnable (second prediction) in contrast to *random* homophony, where there is no featural basis for meanings to share a form.

In Experiment 1, we targeted these two predictions by focusing on person categories that involve the speaker (first exclusive and inclusive persons), and their interaction with number features. We investigated the contrasts that arise by the interaction of two binary features, one for person ([±addressee]) and one for number ([±minimal]) (see Noyer 1992, Bobaljik 2008, Harbour 2014, for more developed accounts).[5] The [±addressee] feature ensures a clusivity contrast: it distinguishes between groups of individuals that include both speaker and addressee (i.e., *iu*, iu_o) and those that exclude the addressee (i.e., *i*, i_o). The [±minimal] feature distinguishes the minimal elements that satisfy each person category (i.e., *i*, *iu*) from nonminimal (augmented) pluralities, where the reference may include other(s) as well (i.e., i_o, iu_o). The resulting four-cell partition of this person-number space is given in table 3.

There are multiple partitions of this space as defined by homophony. We focus here on bipartitions, where two pronominal forms cover the four-cell space. One possible bipartition uses only the [±addressee] distinction, thus contrasting exclusive and inclusive, but neutralizing number ("Person-contrast" in table 4a). A second possible bipartition uses the [±minimal] distinction, resulting in one form for both minimal categories and another for both nonminimal (augmented) categories ("Number-contrast" in table 4b). Interestingly, to make this particular distinction, this paradigm also relies on (or presupposes) the [±addressee] person contrast: the individuals *i* (speaker alone) and *iu* (speaker and addressee alone) are only "minimal" if the speaker-addressee dyad (*iu*) is treated as a smallest element that satisfies the [+addressee] feature. That is, the [+minimal] feature can only pick up the smallest element of the inclusive category if there is an inclusive person category to begin with. Thus, we must assume that the [±addressee] contrast is active in this system, even if it is not encoded by different phonological forms. A schematic

Table 3
Reduced person space

	MIN +min	AUG −min
1 EXCL [(+sp)−add]	*i*	i_o
1 INCL [(+sp)+add]	*iu*	iu_o

[5] We are not committed to this specific inventory; our predictions would hold for any theory that posits the contrasts themselves, regardless of the structure of the feature space.

Table 4
Different partitions of a reduced person space

		MIN	AUG
a. Person-contrast (Number homophony)	1 EXCL		
	1 INCL		
b. Number-contrast (Person homophony)	1 EXCL		
	1 INCL		
c. Random	1 EXCL		
	1 INCL		
d. English-like	1 EXCL		
	1 INCL		

derivation of the bipartitions in table 4a and table 4b using [±addressee] and [±minimal] features is given in the online appendix (figure A.1) (https://doi.org/10.1162/ling_a_00406).

Partitions with random, non-feature-based homophony are also possible. For example, exclusive minimal (*i*) and inclusive augmented (iu_o) may share one form, and inclusive minimal (*iu*) and exclusive augmented (i_o) another. This meaning-to-form mapping cannot be expressed in terms of neutralizing a single semantic distinction ("Random-contrast" in table 4c). Put differently, there is no natural class that groups *only* exclusive minimal and inclusive augmented. Note that there are three other random partitions of this reduced person space.

Our experiment targeted English-speaking learners. In order to make clear learnability predictions for this population, it is important to understand how noninclusive systems of first person pronouns (as in English singular *I* vs. plural *we*) are typically derived in feature-based theories. As observed above, inclusive systems may differ in whether they make a number distinction within the inclusive: languages like Mandarin do not, whereas languages like Ilocano distinguish pronouns referring to speaker and addressee alone from pronouns referring to speaker, addressee, and others (see table 2). To account for this variation, theories of number morphology often distinguish the classical singular/plural contrast, based on the [±atomic] feature, from a minimal/augmented one, based on the [±minimal] distinction (Noyer 1992, Harbour 2011, 2014).[6] Informally, the [±atomic] feature picks up elements from the domain as a function of whether or not they have proper parts, whereas the [±minimal] feature picks out the smallest element(s) in the domain. On their own, these two contrasts only yield different partitions for languages that distin-

[6] Another argument that has been used to support the existence of both [±atomic] and [±minimal] contrasts comes from languages that make a singular/dual/plural distinction, but this is not relevant for our purposes (Harbour 2014, Martí 2020).

guish an inclusive person category—namely, for systems where the smallest pronominal referent might be nonatomic (i.e., *iu*). In noninclusive languages, by contrast, the sets of minimal and atomic pronominal referents are identical; both [±atomic] and [±minimal] contrasts will always lead to the same results (see figure A.1 in the online appendix). Here, we follow much previous literature in assuming that minimal/augmented pronominal systems are necessarily inclusive, and we take languages like English, which make only a singular/plural distinction, to be based on the simpler [±atomic] contrast (e.g., Noyer 1992, Harbour 2014, 2016). Under these assumptions, then, both the [±minimal] and the [±addressee] features are nonnative for English speakers in the context of first person partitions. Note, however, that even if languages like English *were* described as making use of the [±minimal] feature, this contrast alone would still be insufficient to derive the bipartitions in table 4a–c, as the [±addressee] feature is also required (see also the discussion in section 2.4).[7]

In this experiment, we therefore test the two predictions laid out above: first, that English speakers should be able to learn contrasts (table 4) that are not directly instantiated in their language but are broadly attested typologically—namely, the inclusive/exclusive distinction and the minimal/augmented distinction (e.g., in Ilocano; Thomas 1955, Cysouw 2003); second, that these unfamiliar feature-based paradigms should be easier for English speakers to learn than random homophony paradigms (table 4c). If these predictions are borne out, then we can conclude that the person-number space is indeed based on a set of universal features, such as those posited by the theories described above. We can then take the next step of testing different predictions made by particular feature-based theories. As a sanity check, we also test whether English-speaking learners are biased in favor of a person system that resembles their own, as in table 4d. If participants perceive the similarity between the new system they are learning and the English person-number system, then this is a good indication that our experimental methodology is successfully engaging the linguistic space we intend.

To test these predictions, we used two complementary artificial language learning paradigms. In Experiment 1A, participants were taught a pronominal system that matched one of the four paradigms in table 4, and were then tested on how accurately they were able to learn each pattern ("ease of learning" paradigm; Culbertson, Gagliardi, and Smith 2017, Tabullo et al. 2012). In principle, two paradigms might be equally learnable (after some amount of exposure), and yet one of them might still be preferred. In Experiment 1B, we therefore investigated differences in the likelihood of inferring a given paradigm in the absence of explicit data ("poverty of the stimulus" paradigm; Wilson 2006, Culbertson and Adger 2014). Participants were trained on only two cells of the paradigm in table 3 and then had to use the forms they had learned to express all the cells in the paradigm. In other words, they had to extrapolate the taught forms to the remaining two categories. How they should extrapolate was ambiguous on the basis of their training, and how they did so indicated which underlying paradigm they had inferred. For example,

[7] Crucially for us, the [±addressee] distinction is nonnative *in the presence of* the [+speaker] feature (i.e., {[+speaker], [±addressee]} is nonnative). Of course, English speakers do have experience with this feature: it is used to distinguish first person (collapsing inclusive and exclusive) from second person.

a participant might be trained on two distinct forms for exclusive minimal (speaker only) and exclusive augmented (speaker plus others), and then tested on how they mapped these forms to the two remaining categories including the addressee. If they used the augmented form for both new categories, then they had inferred an English-like paradigm.

2.2 Methods

Both experiments, including all hypotheses, predictions, and analyses, were preregistered: Experiment 1A (https://osf.io/w8n25) and Experiment 1B (https://osf.io/j2rcn). Materials, data, and scripts are provided at https://osf.io/ca8yp/?view_only=24c66919e3ff41cf8a01f7c328dead6e. All analyses are as per the preregistration unless we say otherwise. These experiments were implemented using the JavaScript library jsPsych (De Leeuw 2015) and presented to participants in a web browser.

2.2.1 Design Participants in Experiment 1A were randomly assigned to one of four conditions: English-like, Person-contrast, Number-contrast, and Random-contrast (see table 4). Participants in all conditions were taught two pronominal forms mapped into four person categories (exclusive minimal, inclusive minimal, exclusive augmented, and inclusive augmented). All conditions instantiated bipartitions of the person space with two-to-one mappings, but differed on which contrast was directly reflected in the forms (and which one was neutralized): an English-like contrast ([±atomic]), a person contrast ([±addressee]), a number contrast ([±minimal]), or a random homophony pattern.

In Experiment 1B, participants were randomly assigned to one of three conditions, illustrated in table 5.[8] Conditions differed in which subset of two first person categories was trained (critical training set) and held out (critical held-out set). This determined which alternative full paradigms were consistent with the two categories participants learned. Condition 1 was consistent with an English-like pattern (or with feature-based homophony). Conditions 2 and 3 were consistent with either a person or a number contrast system (i.e., feature-based homophony), or with a random contrast system (i.e., random homophony).

All participants in both Experiment 1A and Experiment 1B were additionally exposed to another four pronominal forms that mapped into the second and third person singular and plural categories. These forms were used as fillers and were not analyzed.

2.2.2 Materials The same materials were used in Experiments 1A and 1B. In both cases, the language consisted of six different pronoun forms, used for the filler categories (2SG/PL, 3SG/PL), plus the critical first person forms. For each participant, these six lexical items were randomly drawn from a list of eight CVC words created following English phonotactics: *kip*, *dool*, *heg*, *rib*, *bub*, *veek*, *tosh*, *lom*. Items were presented orthographically.

To express the pronoun meanings, we commissioned a cartoonist to draw scenarios involving a family of three sisters and their parents. Each family member had a clearly defined role in the

[8] Two additional conditions were also run, but are not reported here. Since these are orthogonal to the main aim of the experiment, we simply refer to Maldonado and Culbertson 2019 for more details.

Table 5
Conditions in Experiment 1B. There were two training and two held-out categories per condition. Each training category was mapped to a different pronominal form (here called A or B), schematically represented with light and dark gray. Participants had to use the training forms they learned (A or B) to express the held-out meanings (cells with *A or B?*). There are four different paradigms compatible with the training per condition, as specified in the rightmost column.

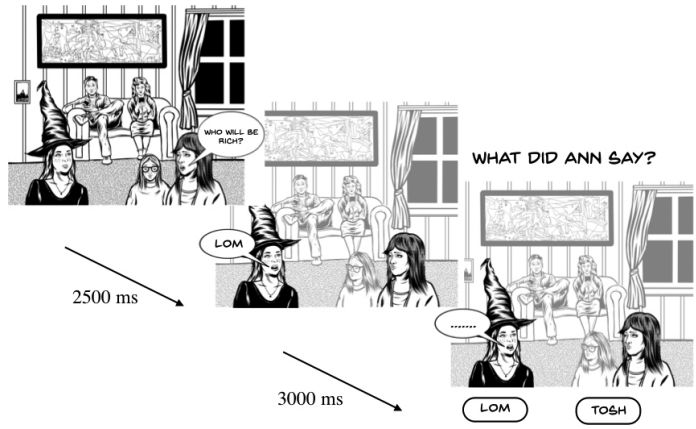
	Mappings			Compatible paradigms
		MIN	AUG	
Condition 1	EXCL	A	B	English-like, Number-contrast, Random × 2
	INCL	A or B?	A or B?	
		MIN	AUG	
Condition 2	EXCL	A or B?	A	Person-contrast, Random × 3
	INCL	A or B?	B	
		MIN	AUG	
Condition 3	EXCL	A or B?	A or B?	Number-contrast, Random × 3
	INCL	A	B	

Table 6
Highlighted family members for each person category. To ensure that forms were not associated with specific quantities, critical augmented categories randomly included one or two additional others. Third person singular meanings were always expressed with a female other.

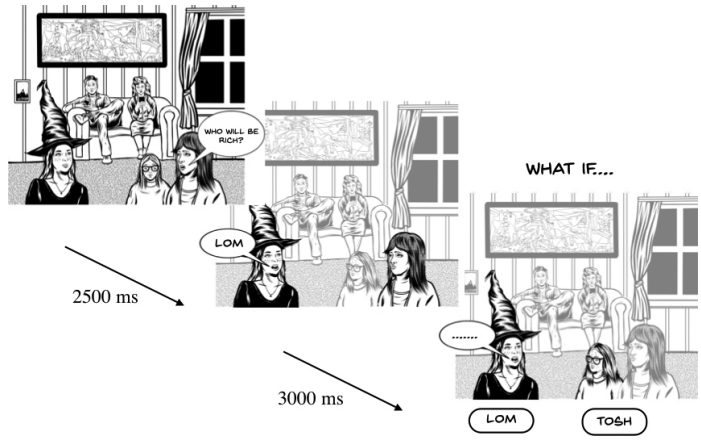
Category	Highlighted set
1 EXCL.MIN	speaker
1 INCL.MIN	speaker, addressee
1 EXCL.AUG	speaker, other(s)
1 INCL.AUG	speaker, addressee, other(s)
2.MIN	addressee
2.AUG	addressee, other(s)
3.MIN	one other
3.AUG	multiple others

conversational context. The two older sisters were speech act participants (in all scenarios, they were either speaker or addressee). The third (little) sister was spatially close, but never a speech act participant. The parents were seated in the background (serving as additional others).

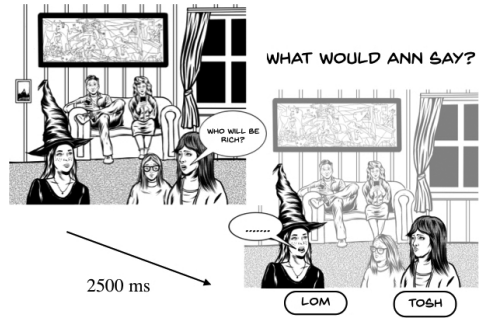
Pronouns were used as one-word answers to questions like *Who will be rich?* Meanings were expressed by visually highlighting subsets of family members, as in table 6. In some cases, more than one pattern of visual highlighting could match the target meaning; options were then selected randomly. An example illustrating the INCL.MIN trial is provided in figure 2. All questions



(a) Exposure (INCL.MIN)



(b) *What if . . .* (EXCL.AUG)



(c) Testing (INCL.MIN)

Figure 2
Illustration of exposure, *what if . . .*, and testing trials. Feedback was presented for 2000 ms after response in exposure and *what if . . .* trials.

were English interrogative sentences of the form *Who will . . . ?*, which were randomly drawn from a list of 60 different tokens.

2.2.3 Procedure The basic procedure was the same in Experiments 1A and 1B. Participants were first introduced to the family, including the names of the sisters, and were told they were going to see the sisters playing with a hat that had two magical properties: whoever wore it could see the future but would also talk in a mysterious ancestral language. Participants were instructed to figure out the meanings of the words in this new language. They were given the hint that the words were not names, and an example trial with an English pronoun (*her*). In addition, the speaker and addressee roles switched several times during the experiment to highlight that the words were dependent on contextually determined speech act roles. This was induced by swapping who had the magical hat.

Experiment 1A had two phases. In the training phase, participants were first exposed to six pronominal forms in the new language, corresponding to the four filler and four critical person categories. Each exposure trial had two parts: a scene where a question was asked, and a scene where the question was answered with a pronoun form in the target language (e.g., figure 2a). To check that participants were paying attention, they were then asked to select the pronominal form they had just seen from two alternatives. There were 12 training trials (2 repetitions per form). After this initial exposure, participants were tested on the trained forms in what we call *what if . . .* trials. *What if . . .* trials consisted of a question-and-answer scene, as in the exposure phase, followed by a "What if?" scene in which a new set of individuals was highlighted. Participants were asked to pick the correct word for that meaning from two alternatives (e.g., figure 2b). There were 32 such trials (3 repetitions per control form, 6 per critical form). Participants were given feedback on their answers. Participants were then given a final, critical test. Trials consisted of a question scene, followed by a scene highlighting the referent(s), but no pronominal form. Participants had to pick the word corresponding to the meaning from two alternatives (e.g., figure 2c). This phase consisted of 24 trials (3 repetitions per form). Participants received no feedback during this phase.

Experiment 1B also had a training and a testing phase. Crucially, during the training phase participants were only trained on the pronouns in the filler and critical training sets (6 person categories). There were 12 exposure trials (2 repetitions per form) and 16 *what if . . .* trials (2 repetitions per filler form, 4 per critical training form). Participants were given feedback on their answers. The critical testing phase included trials for the two remaining critical categories, that is, the held-out set. This phase consisted of 48 trials (6 repetitions per form). Participants received no feedback during this phase.

Both experiments included a *pretraining* phase where participants were exposed only to the three singular person pronouns. This was done to familiarize participants with the setup by using less complex stimuli (i.e., scenes where a single family member was highlighted as the pronominal referent). At the end of both experiments, participants were given a debrief questionnaire, which included questions targeting how they interpreted the meanings they were taught. Importantly, most participants reported having understood the words as pronouns. For example, participants in Experiment 1B (Condition 2) described the meaning of form A as 'me or us not including

you' and the meaning of form B as 'us including you'.[9] More details about the procedure used in these experiments can be found in table A.1 in the online appendix.

2.2.4 Participants A total of 197 English-speaking adults were recruited via Amazon Mechanical Turk for Experiment 1A (English-like group: 48, Person-contrast: 49, Number-contrast: 50, Random: 50). Two further participants were excluded for not being self-reported native speakers of English. Of the 197 participants, 171 (English-like group: 44, Person-contrast: 41, Number-contrast: 41, Random: 45) responded accurately on more than 80% of exposure trials during the training phase and were considered for further analysis, according to our preregistered plan. Participants were paid 2 USD for their participation, which lasted approximately 15 minutes.

A different group of 253 English speakers were tested in Experiment 1B (Condition 1: 87, Condition 2: 87, Condition 3: 79). Per our preregistered plan, participants were excluded if (a) their accuracy rates during exposure training were below 80%, or (b) they had not answered correctly more than two-thirds of the training trials. Note that high accuracy rates on trained critical items are important here because extrapolation of these forms is only interpretable if participants have learned them. This resulted in analysis of 131 participants (Condition 1: 46, Condition 2: 49, Condition 3: 36). Participants were paid 2.5 USD for their participation, which lasted approximately 20 minutes.

2.3 Results

2.3.1 Experiment 1A Figure 3 shows the proportion of correct responses in critical trials per experimental condition (English-like, Person-contrast, Number-contrast, and Random-contrast) during the testing phase. We ran two logistic mixed-effects models (using the lme4 software package (Bates et al. 2014) in R (R Core Team 2018)) to evaluate the effect of the experimental condition on accuracy rates (coded as 0 or 1). Both models included by-participant random intercepts. Here and throughout, the standard alpha level of 0.05 was used to determine significance, and *p*-values were obtained using asymptotic Wald tests.

A first model assessed whether the English-like pattern was learned better than the alternative paradigms. We used treatment contrast coding with the English-like paradigm as baseline; each of the remaining conditions was compared with this fixed level. The model revealed that the proportion of correct responses in the English-like group was significantly higher than chance ($\beta = 1.78$, $p < .001$). Accuracy in the Person-contrast and Number-contrast groups did not differ significantly from that of the English-like baseline (Person-contrast: $\beta = -0.6$, $p = .093$; Number-contrast: $\beta = -0.38$, $p = .28$); however, accuracy in the Random group was significantly lower than the baseline ($\beta = -1.24$; $p < .001$). This matches the visual pattern in figure 3.

[9] Not all participants reported pronouns for these meanings. Interpreting participants' responses in these cases is not straightforward. For example, a highly accurate participant reported the meaning of form A to be 'sisters'. This suggests that questionnaire responses do not necessarily convey what participants have implicitly learned. We therefore use questionnaire responses as a general sanity check but rely on accuracy rates to draw conclusions about participants' performance in the experiment.

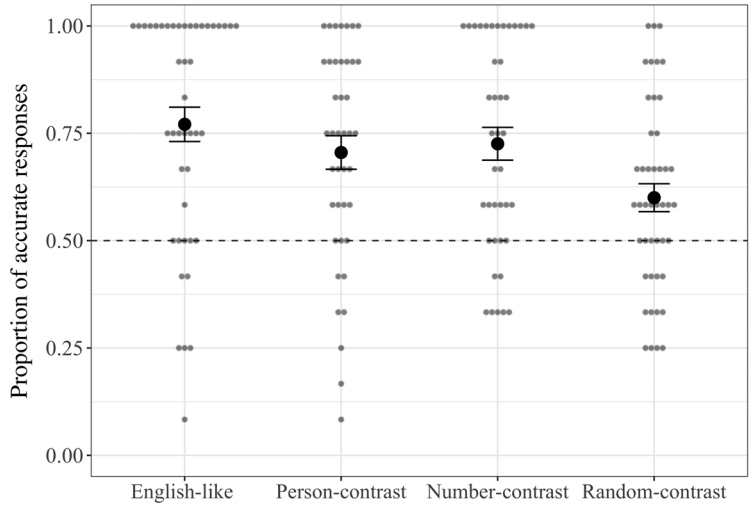


Figure 3
Accuracy rates in critical testing trials by condition in Experiment 1A. Error bars represent standard error on by-participant means; gray dots represent individual participants' means.

We ran a second model to explore the difference between the feature-based and random patterns. The analysis was restricted to Number-contrast, Person-contrast, and Random conditions. We used treatment coding with the Random condition as baseline. The proportion of correct responses in this baseline group was significantly higher than chance ($\beta = 0.58, p < .001$), but significantly lower than in both feature-based conditions (Number-contrast: $\beta = 0.8, p < .01$; Person-contrast: $\beta = 0.62, p = .03$). This suggests that participants trained to make a (nonnative) person or number contrast were more accurate than those trained on a random contrast.

2.3.2 Experiment 1B Recall that participants in Experiment 1B were taught two pronominal forms (coded as forms A and B), which they had to use to describe both a critical trained set and a held-out set of person meanings involving the speaker (levels: EXCL.MIN, EXCL.AUG, INCL.MIN, INCL.AUG). Figure 4 shows the proportion of trials on which participants chose the form A (pronoun) for each critical trained and held-out meaning during the testing phase. Choice of the same form across categories indicates homophony. A visual inspection of figure 4 suggests that participants in Condition 1 consistently used one form for the EXCL.MIN category and the other form for the remaining three categories; this indicates inference of an English-like paradigm. Participants in Conditions 2 and 3 appear somewhat noisier in their responses; however, distinct patterns are evident. In Condition 2, one form is used for the two first person categories, and, at least for some participants, the other form is used for the two inclusive categories (consistent with maintenance of the person contrast, i.e., number homophony). In Condition 3, one form is used for the minimal categories and the other for the plurals (consistent with maintenance of the number contrast, i.e., person homophony). Note, however, that this figure shows by-participant averages for each meaning category rather than which patterns *individual participants* produced.

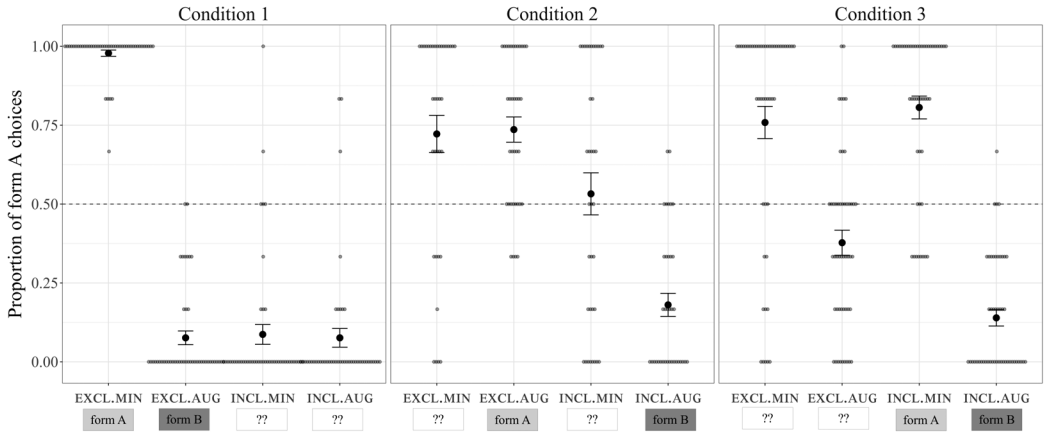


Figure 4
Proportion of form A (as opposed to form B) choices for each first person category during the testing phase in Experiment 1B. Choice of the same form (A or B) across categories indicates homophony. Error bars represent standard error on by-participant means; dots represent individual participants' means.

Figure 4 suggests that there is relatively little variation across participants in Condition 1 compared with the other conditions; almost all participants chose the same form for each category, and they tended to do so categorically. To confirm this statistically, we calculated the *joint entropy* of the held-out set for each individual. This value indicates the degree of uncertainty or variability in each participant's mapping of the trained forms to the held-out categories. Participants who are less consistent in their mapping will have higher joint entropy values. We then fit a linear regression model predicting joint entropy by Condition (3 levels). We used treatment coding, with Condition 1 as baseline. No random effects were included in the model, as each participant had a single joint entropy value. As predicted, joint entropy rates were significantly higher for Conditions 2 and 3 (intercept: $\beta = 0.28$; vs. 2: $\beta = 0.44 \pm .13$, $p < .001$; vs. 3: $\beta = 0.64 \pm .12$, $p < .001$).

A second analysis evaluated whether individual participants in Conditions 2 and 3 were more likely to infer feature-based rather than random patterns (as suggested by figure 4).[10] We calculated the probability that participants were deriving a feature-based pattern (a person contrast in Condition 2 or a number contrast in Condition 3) given their responses to held-out meanings (see figure A.2 in the online appendix). In Condition 2, we computed the probability of choosing form A for the EXCL.MIN and form B for the INCL.MIN; in Condition 3, we computed the probability of choosing form A for the EXCL.MIN and form B for the EXCL.AUG.

We then ran nonparametric Wilcoxon signed-rank tests per condition to determine whether the probability of deriving a feature-based pattern was higher than chance. Given that there were

[10] This analysis diverges from our preregistered plan, which was designed to test this same prediction with a different analysis method. We believe the current analysis is both simpler and more technically sound.

four paradigms compatible with the training in each condition, chance level was set at 25%. The results of these tests indicate that the probability of deriving a person contrast in Condition 2 was not significantly different from chance ($p = .406$), but the probability of deriving a number contrast in Condition 3 was above chance ($p < .001$). The same procedure was followed in Condition 1 with respect to the probability of deriving an English-like pattern. As expected, this probability was significantly higher than chance ($p < .001$).

2.4 Discussion

The main aim of these first two experiments was to test whether learners are sensitive to feature combinations, or contrasts, that are not present in their native language. We exposed English-speaking learners to paradigms expressing four person-number categories in a new language. We focused on systems instantiating either the inclusive/exclusive or the minimal/augmented contrast, which have been argued to have a universal basis in two features, encoding person and number, respectively (e.g., [±addressee], [±minimal]). Our experiments also included an English-like paradigm, as a sanity check. We predicted that participants would find the English-like paradigm easiest to learn, followed by the two non-native-like feature-based homophony patterns, with random homophony being least learnable.

In Experiment 1A, we tested these predictions by training participants on one of four paradigms and comparing how accurately they learned each one. Results confirmed a numeric advantage for native-like pronoun systems: learners found it easier to learn a paradigm with the same structure as English. Interestingly, there was no statistically significant difference in learnability between English-like systems and systems instantiating number or person homophony, suggesting that participants can readily learn other feature-based partitions as well. Results also confirmed that participants trained on a pronominal system with a (nonnative) person or number homophony pattern outperformed those trained on a random homophony pattern. This supports the claim that learners perceive this four-cell person space (i.e., EXCL, INCL, 2, and 3) as the interaction of two distinct features, rather than as a conjunction of four different categories, fully independent from each other (in line with Sauerland and Bobaljik 2013). If learners divide the exclusive, inclusive, minimal, and augmented categories into two natural semantic classes, one for person and one for number, then learning our systematic homophony paradigms simply involves one nonnative contrast each.

These results were for the most part confirmed in Experiment 1B, where participants were trained on ambiguous data that required them to extrapolate trained forms to new meanings. Here, learners were significantly more likely to infer an English-like pattern. They were also more likely to infer a paradigm with a feature-based number contrast (and person-based homophony) than a random contrast, thus making productive use of the nonnative [±minimal] distinction. As noted above, in order to do so—that is, to treat the *i* (speaker only) and the *iu* (speaker and addressee only) references as "minimal"–participants need to be sensitive not only to the minimal/augmented contrast but also to the inclusive/exclusive one, as the dyad of speaker and addressee can only be considered a minimal unit if there is an inclusive category.

We did not find a significant difference in participants' likelihood of inferring a feature-based person contrast (and number homophony) over a random one. In other words, after being trained on an inclusive/exclusive distinction in part of the paradigm (EXCL.AUG and INCL.AUG), participants did not generalize this contrast to the held-out cells of the paradigm. This suggests that although the nonnative clusivity distinction is indeed learnable (in Experiment 1A), English-speaking learners do not necessarily make a productive use of it. The apparent difference between number and person homophony is supported by a post hoc analysis showing that accuracy rates on trained categories (before exclusion) are higher in Condition 3 than in Condition 2 ($p < .001$). The [±addressee] person distinction in pronominal forms was thus harder to learn than the [±minimal] distinction.

One possible explanation for this difference is that it reflects participants' experience with homophony in English: since English encodes a ([±atomic]) number distinction, it is possible to characterize it as a system with (only) person homophony (i.e., a noninclusive language; Harbour 2016). In other words, English speakers have more experience with distinctions in number (in general) than in clusivity.[11]

Alternatively, the fact that English speakers do not generalize this person contrast could be thought of as the result of applying a native constraint against such inclusive systems. In contrast with binary-features accounts, some approaches (e.g., Harley and Ritter 2002) describe noninclusive systems as making use of two privative features (Speaker and Addressee) together with a constraint that prevents these features from cooccurring. Arguably, the apparent difficulty in generalizing the inclusive/exclusive distinction might suggest that English speakers have a harder time learning a distinction that violates a native constraint against the simultaneous specification of the Speaker and Addressee features.[12]

To summarize, this study presents the first experimental evidence for differences in learnability between alternative person paradigms. Unsurprisingly, native-like paradigms are easiest to learn and most likely to be inferred when input is ambiguous. More interestingly, paradigms exhibiting homophony within a natural class are learned (and in some cases inferred) more readily than paradigms with random homophony. In what follows, we build on these basic results to investigate more specific constraints on the person space hypothesized by feature-based theories of person.

[11] Importantly, our findings are not compatible with participants' being *only* sensitive to the specific number distinctions found in English. In particular, both Conditions 2 and 3 involve collapsing categories that are expressed as distinct pronouns—*I* and *we*—in English. While one could in principle argue that this explains why participants in Condition 2 are less likely to collapse the exclusive minimal and augmented categories, this would not explain why participants in Condition 3 readily collapsed the exclusive and inclusive minimal.

[12] Note that the advantage for a paradigm with a number contrast over a person contrast (i.e., for person over number homophony) found in Experiment 1B contrasts with typological data, which suggests that languages with different pronominal forms for exclusive and inclusive persons but no number distinction are more common than minimal/augmented languages that do not make an inclusive contrast (Cysouw 2013). However, these counts are very sparse.

3 Experiment 2: *Within You, Without You*

3.1 Introduction

In a classic paper, Zwicky (1977) made the following observation regarding the crosslinguistic distribution of person systems: in languages that do not distinguish clusivity (e.g., English), the 'you and us' inclusive meaning is always expressed as a form of 'us' and never as a form of 'you' (or 'them').[13]

At first glance, Zwicky's generalization is quite surprising. Most feature-based approaches to person systems (e.g., Bobaljik 2008) assume that the inclusive person shares features with both the first person exclusive (e.g., [+speaker]) and the second person exclusive (e.g., [+addressee]). Indeed, a number of languages have inclusive pronouns that can be morphologically decomposed into first plus second person forms (e.g., Bislama; Harbour 2011).[14] This leads naturally to the expectation that languages should be as likely to assimilate the inclusive with the second person as they are to assimilate it with the first. In contrast, no theory would predict the inclusive meaning to be homophonous with the third person, as the inclusive and the third person do not have any features in common (although see Rodrigues 1990 for a potential exception).

There have been two general approaches to Zwicky's generalization in the literature. The first maintains the traditional set of features, but posits default feature specifications in order to predict an asymmetry between first-inclusive and second-inclusive (Harley and Ritter 2002, McGinnis 2005). Harley and Ritter's (2002) feature geometry account maintains both Speaker and Addressee features as dependent nodes of the feature Participant, but the Speaker feature is considered to be less marked than the Addressee feature. Consequently, in languages without an inclusive distinction, a preference for assimilating the inclusive meaning into the first person is expected, as they share the default feature. Defaults can be overridden; therefore, the second-inclusive homophony pattern can still arise. By contrast, a third-inclusive system is impossible.

The second approach is to use a different set of features. Harbour (2016) posits [±author] and [±participant], denoting the semilattices $\{i\}$ and $\{i, iu, u\}$ respectively, with the values of the features modeled as complementary operations on lattices. While in Harbour's system the inclusive and second person categories do share "ontological" primitives—both of them contain *u*—the absence of a [±addressee] feature—and of a lattice consisting only of $\{u\}$—creates an inherent asymmetry in how speaker and addressee roles can be represented in a person partition.[15] This asymmetry derives Zwicky's observation as a *strong* constraint on possible person systems. A language that makes use of the [±author] feature will have a bipartition of the person space in which *i* and *iu* are homophonous and *u* and *o* are homophonous. Similarly, a language with both [±author] and [±participant] can have a tripartition (if [±participant] composes last; see Harbour

[13] This is a generalization about languages and not about individual paradigms within a language, which might show accidental homophony (see Harbour 2016 and examples therein).

[14] Although note that there are also languages where the inclusive form patterns morphologically with the second person (e.g., Ojibwa pronouns; Harley and Ritter 2002 from Schwartz and Dunnigan 1986).

[15] Harbour's (2016) proposed ontology consists of *egocentrically* nested subsets, such that the smallest subset in the ontology contains the speaker alone—that is, $i \subset i, iu, u \subset i, iu, u, o$. This ontology is shared by Ackema and Neeleman (2018).

2016 for details) in which *i* and *iu* are homophonous (with two additional forms for *u* and *o*), or a quadripartition (if [±participant] composes first). Without a corresponding [±addressee] feature, though, there is no way to have a system that picks out the set including *iu* and *u.* Indeed, tripartitions involving homophony between inclusive and second person or inclusive and third person are equally impossible.[16]

The theories outlined above differ critically in how second-inclusive and third-inclusive are treated. For Harley and Ritter (2002), third-inclusive is singled out as underivable, while second-inclusive is possible but more marked than first-inclusive. By contrast, Harbour (2016) takes as his starting point the idea that only first-inclusive tripartitions can be generated by the grammar. On the basis of typology alone, it is impossible to adjudicate between these theories: both second- and third-inclusive patterns are unattested. Moreover, neither theory provides an explicit mechanism for linking the feature-based representations (and operations) they posit to typology. The implicit link is *learnability*: only a subset of possible person partitions are learnable by humans; alternatively, some are learned more readily than others. In Experiment 2, we investigated learners' sensitivity to predicted asymmetries among noninclusive paradigms. To do this, we used an ease-of-learning design: we trained English-speaking learners on a new language with an inclusive that was a form of *us* (first-inclusive), a form of *you all* (second-inclusive), or a form of *them* (third-inclusive), and compared how well the systems were learned. Given that English features first-inclusive homophony *and* this is the only tripartition systematically attested in typology, learners were predicted to prefer such paradigms over alternatives. Regarding second-inclusive and third-inclusive homophony, if both patterns are directly ruled out by the grammar (as in Harbour 2016), learners were expected to be equally unlikely to learn either of them. By contrast, if learners are sensitive to the featural commonalities between inclusive and second person (e.g., [+addressee]), a second-inclusive system was expected to be easier to learn than a third-inclusive one (Harley and Ritter 2002, McGinnis 2005). This pattern of results would suggest that an asymmetry between first-inclusive and second-inclusive languages should not be encoded as a hard constraint on person systems (contra Harbour 2016).

3.2 Methods

This experiment, including all hypotheses, predictions, and analyses, was preregistered at https://osf.io/5h4m6. All materials, data, and scripts are provided at https://osf.io/p2c4r/?view_only=ba7a554345f84a2dbddfbcfea67691d3. This experiment was implemented using the JavaScript library jsPsych (De Leeuw 2015) and presented in a web browser.

3.2.1 Design Participants were randomly assigned to one of three conditions, summarized in table 7. Participants in all conditions were taught three pronominal forms mapped to four *plural* person categories (first exclusive, inclusive, second exclusive, and third). Note that in this experiment we rely on a singular/plural (not minimal/augmented) number contrast, with no number

[16] An intermediate proposal is offered by Ackema and Neeleman (2013, 2018). In their proposed feature structure, "there is no natural class . . . that comprises the first person inclusive and the second person, but not the first person exclusive" (2013:910). However, second-inclusive patterns can still be obtained by incorporating an impoverishment rule in the system (see footnote 28 for details). This is not possible for inclusive-third homophony.

Table 7
Conditions in Experiment 2. Grayed cells are mapping to a single pronominal form, white cells to different and distinct forms.

a. First-inclusive	1 EXCL	
	1 INCL	
	2	
	3	
b. Second-inclusive	1 EXCL	
	1 INCL	
	2	
	3	
c. Third-inclusive	1 EXCL	
	1 INCL	
	2	
	3	

distinction in the inclusive person (INCL). Conditions differed in whether the inclusive meaning was assimilated into the first person plural (First-inclusive condition), the second person plural (Second-inclusive condition), or the third person plural (Third-inclusive condition).

Participants in all three conditions were also exposed to three distinct pronominal forms corresponding to the first, second, and third *singular* persons. Participants' learning of these forms was used as an exclusion criterion (see below).

3.2.2 Materials The language consisted of six different pronominal forms: three forms were used for the plural pronouns (critical categories), and three different forms were used for the singular pronouns (filler categories). For each participant, these were randomly drawn from a list of eight CVC words (see Experiment 1).

Pronouns were again used as one-word answers to English interrogative sentences of the form *Who will . . . ?*, randomly drawn from a list of 60 different tokens. Meanings were expressed by highlighting subsets of family members, as in table 8. Visual stimuli were the same as in Experiment 1.

3.2.3 Procedure The general backstory was as in Experiment 1. Participants were instructed to figure out the meanings of the words in the new language, and they were told that the words they were learning were not names. As in Experiment 1, the speaker and addressee roles switched several times during the experiment to highlight that the words were context-dependent.

The experiment had two phases, each composed of exposure and testing blocks (e.g., figure 5). Trials in each of these blocks were analogous to those used in Experiment 1, except that participants had to select the correct word for that meaning among *three* different options (not two).

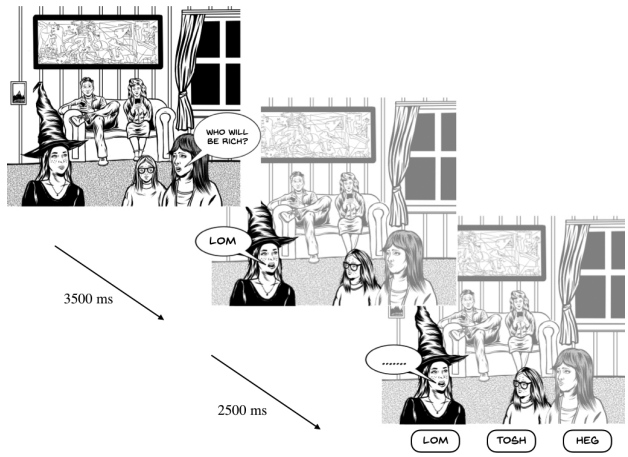
Table 8
Highlighted family members for each person category. 1 EXCL, 2, and 3 plural categories randomly included one or two additional others; the INCL category could refer to speaker and addressee alone or include one or two others as well. In this experiment, there was no distinction between minimal and augmented; rather, there was a distinction between singular/atomic and plural.

Category	Highlighted set
1 EXCL.SG	speaker
2SG	addressee
3SG	one other
1 EXCL.PL	speaker, other(s)
1 INCL	speaker, addressee (other(s))
2PL	addressee, other(s)
3PL	multiple others

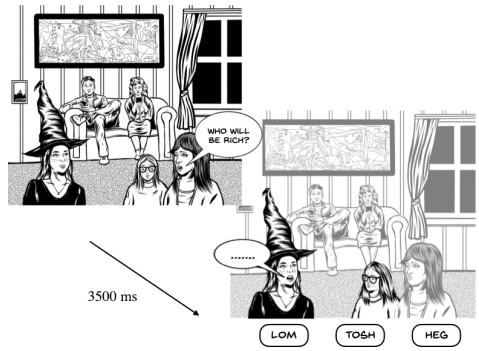
During the first phase, participants were trained and tested on the three singular pronouns. There were a total of 12 exposure and 12 testing trials (4 repetitions per form/meaning). Participants who responded accurately to at least two-thirds of the testing trials in this phase (8 correct responses) moved on to the second, critical phase. This phase was composed of two alternating exposure and testing blocks targeting the mapping between three plural pronouns and four person meanings. There were a total of 24 exposure trials (6 repetitions per meaning) and 48 testing trials (12 repetitions per meaning). More details about the experimental procedure are provided in table A.2 in the online appendix.

The order of presentation of meanings was fully randomized within exposure and testing blocks for each participant. As in Experiment 1, participants were given a debrief questionnaire at the end of the experiment to check how they interpreted the forms they were trained on. For example, participants in the Second-inclusive condition described the meaning of the critical form as 'me and you or you all' or as 'group containing Ann or Mary'.

3.2.4 Participants A total of 320 English-speaking adults were recruited via Amazon Mechanical Turk (First-inclusive group: 109, Second-inclusive: 101, Third-inclusive: 110). This does not include workers who were excluded for not being self-reported native speakers of English (10) and participants who failed to pass an attention check included at the beginning of the experiment (35). This attention check was added because our exit questionnaires in Experiment 1 revealed that many participants had not read the instructions or were bots (Rouse 2015). While these participants were usually excluded by our other criteria, the attention check allowed us to filter them out in advance, distinguishing them from participants who just found the experiment hard. A total of 167 participants responded accurately on more than eight singular testing trials and were allowed to continue with the critical plural pronoun phase, according to our preregistered plan (First-inclusive group: 57, Second-inclusive: 55, Third-inclusive: 55). Participants who finished the two experimental phases (approximately 20 minutes long) were paid 3.5 USD; participants who only completed the first part of the experiment (approximately 5 minutes long) were paid 1 USD.



(a) Exposure



(b) Testing

Figure 5
Illustration of exposure and testing trials for the EXCL.PL category in Experiment 2

3.3 Results

Mean accuracy rates on testing trials during the critical phase are given in figure 6. The effect of condition and block on accuracy rates was analyzed using logistic mixed-effect models with random by-participant intercepts and by-block slopes.[17]

[17] There were two testing blocks in the critical phase, each preceded by an exposure block. Participants were generally expected to improve with accumulated exposure, but this improvement could vary across conditions. Each model included the effect of block on accuracy, as well as the interaction with condition. However, we report here only simple effects regarding the second testing block. The complete model output can be found at https://osf.io/p2c4r/?view_only=ba7a554345f84a2dbddfbcfea67691d3.

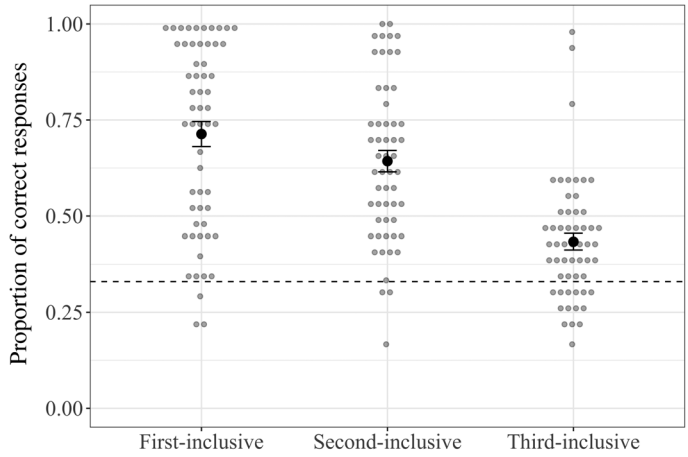


Figure 6
Accuracy rates in critical testing trials by condition in Experiment 2. Error bars represent standard error on by-participant means; gray dots represent individual participants' means. Dashed black line indicates chance value (for three-options forced-choice).

We first compared the First-inclusive with the Second-inclusive and Third-inclusive conditions (contrasts were treatment-coded, with First-inclusive and Block 2 as baselines). The model intercept was significant, indicating that accuracy in the First-inclusive (Block 2) was above chance ($\beta = 1.59$, $p < .001$). In addition, compared with the First-inclusive condition, accuracy was significantly lower in the Third-inclusive ($\beta = -1.83$, $p < .001$) and marginally lower in the Second-inclusive ($\beta = -0.572$, $p = .055$).

We then compared the Second-inclusive and Third-inclusive conditions (contrasts were treatment-coded, with Second-inclusive and Block 2 as baselines). The model intercept was significant, indicating that accuracy in the Second-inclusive (Block 2) was above chance ($\beta = 0.97$, $p < .001$). Accuracy was significantly lower in the Third-inclusive ($\beta = -1.2$, $p < .001$).[18]

3.4 Discussion

These findings confirm that participants are most successful at learning a new language that features homophony between inclusive and first person meanings. This is as expected since this pattern is systematically found in the typology and reflects our English-speaking participants' native system. If Harbour (2016) is correct in deriving a hard constraint on tripartitions that generates only first-inclusive homophony, then we might expect this preference to be very strong indeed. However, the difference in accuracy between the First-inclusive and Second-inclusive conditions was only marginally significant. This result is consistent with the idea that learners

[18] Following our preregistration, we ran a second version of each of these models, restricting the analysis to inclusive meanings (since our hypotheses target the inclusive category specifically). The same pattern of results emerged.

are sensitive to the featural overlap between inclusive and second person, as predicted by Harley and Ritter (2002), McGinnis (2005), and Bobaljik (2008). Participants in the Second-inclusive condition may be treating these as a natural class, relying on a shared feature to learn the partition. This result supports a theory in which first- and second-inclusive tripartitions are both generated by the grammar, but the latter is dispreferred (contra Harbour 2016, and possibly Ackema and Neeleman 2018; but see footnote 16).[19]

Importantly, we also found that learners have a bias against systems that assimilate the inclusive into the third person. This result reveals that second- and third-inclusive systems, despite being (generally) unattested in the typology, are not equal from a learnability perspective: there is stronger pressure against third-inclusive than against second-inclusive homophony. This is again as predicted by theories like Harley and Ritter's (2002), McGinnis's (2005), and Bobaljik's (2008) (but arguably also by Ackema and Neeleman's (2018)), which posit that, unlike first and second person, third person does not form a natural class with the inclusive category (cf., e.g., Rodrigues 1990). As a result, homophony between inclusive and third persons is not predicted to occur systematically, though it might arise accidentally. The learnability cost for third-inclusive systems can therefore be seen as another case of a bias against random homophony.

Returning to Zwicky's observation, our findings suggest that the typological asymmetry between alternative noninclusive systems may not have a *simple* correlate with learning. In the typology, first-inclusive systems are attested systematically, while second- and third-inclusive systems are not; in our experiment, third-inclusive systems were clearly dispreferred, but the advantage for first- over second-inclusive systems was much weaker. While this is consistent with a theory positing weak learning biases (i.e., constraints that can be overridden given sufficient evidence) that penalize third-inclusive most, it still leaves the typological data partially unexplained.

One possibility is that there is an additional weak bias, not at play in our experiments, that further advantages first-inclusive systems. An obvious such candidate is a general *egocentric* bias—that is, increased importance or salience of speakers to themselves (Charney 1980, Loveland 1984, Moyer et al. 2015). If individuals perceive the world as a function of their presence in it, they may be more likely to adopt categorization systems that preserve this distinction. This would lead to an asymmetry between first-inclusive and second-inclusive systems. This bias may be weakened in the context of our experiment, where participants are passive learners and do not themselves feature in the meanings they are learning. We return to this issue in section 5.

[19] Alternatively, as Daniel Harbour (pers. comm.) points out, one could argue that our participants are treating the pronominal system we teach them as an instance of syncretism within an otherwise inclusive language. If this were the case, theories like Harbour's (2016) could still account for our results. In Harbour's system, inclusive languages feature/make quadripartitions of the person space. Underlying quadripartitions might show feature-based syncretism of inclusive and second person in some paradigms, as these categories do share a [+participant] feature. However, given that our participants are speakers of a noninclusive language, it seems very unlikely that they would infer an underlying quadripartition from a three-form pronominal system. Alternatively, because Harbour (2016) posits a shared ontological primitive, *u*, potentially linking inclusive and second person, one could reason that participants in the experiment are relying on this ontological/semantic overlap to learn the system, without paying attention to the features themselves. However, if that were the case, one would need to explain why this does not lead to second-inclusive systems in the typology.

In the next section, we investigate a second typological generalization that appears to challenge the categorical distinction between participants and nonparticipants suggested by Zwicky's generalization.

4 Experiment 3: *I Me Mine*

4.1 Introduction

It has long been observed that there is a fundamental difference between person categories involving speech act participants (first exclusive, inclusive, and second persons) and the third person, which refers to other, nonparticipant individuals. Besides having a *fixed* reference, which does not depend on discourse roles, in a number of languages the third person is treated as morphologically distinct from other person categories (Forchheimer 1953; see summary in Harley and Ritter 2002). These facts have often led researchers to propose that the third person is *unmarked* with respect to the first and second person categories (Benveniste 1971), which are instead encompassed by the same natural class: participant (Hale 1973, Silverstein 1976, Noyer 1992). Here we will focus on one specific instantiation of this proposal, Harley and Ritter's (2002) feature geometry approach.

Harley and Ritter (H&R) capture the intuition that the third person is unmarked by treating it as the default interpretation of the base node (called Referring Expression). First exclusive, first inclusive, and second persons require the presence of the dependent node Participant. For an illustration of H&R's geometry, see figure 1.

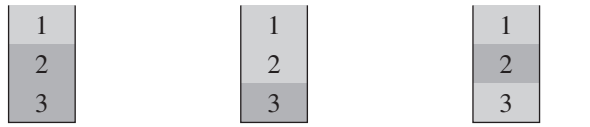
Despite accounting for a number of interesting typological patterns, H&R's system (and others like it; see Béjar 2003, McGinnis 2005) is potentially challenged by the crosslinguistic distribution of person homophony patterns. Specifically, these approaches cannot account for the following typological observation, simplified from (2).[20]

(3) Languages that feature homophony between second and third person categories and between first and second are more frequent than those instantiating first-third homophony.

The numerically higher frequency of first with second (1+2) and second with third (2+3) systems relative to first with third (1+3) systems is illustrated in figure 7. This tendency is found both in free pronouns and in verbal agreement, but it is restricted to nonsingular or number-neutral contexts (i.e., it does not hold for singular cases).[21] Note that 2+3 homophony is also numerically more common than 1+2 patterns. We return to this in section 4.5.

[20] This generalization does not directly concern the inclusive category, as it mostly holds for noninclusive languages, where the inclusive is collapsed with the first person exclusive. For the sake of simplicity, we therefore remove the inclusive from the discussion and refer to both the first person inclusive and first person exclusive as the first (1) person category.

[21] The counts in Baerman, Brown, and Corbett 2005 for personal pronouns and verbal agreement are drawn from two different typological samples, making the counts not fully comparable to each other. In what follows, we specifically focus on pronoun systems.



2+3 homophony	1+2 homophony	1+3 homophony
Pr: 14/200	Pr: 5/200	Pr: 1/200
Agr: 13/111	Agr: 11/111	Agr: 5/111

Figure 7
Illustration of 1+2, 2+3, and 1+3 homophony patterns. Shades of gray indicate forms: cells with the same shade use the same form. Typological counts for pronominal systems (Pr) are from Cysouw's data set (Baerman, Brown, and Corbett 2005, from Cysouw 2003), and counts for verbal agreement (Agr) are from the Surrey Person Syncretism database (Corbett, Brown, and Baerman 2002, Baerman, Brown, and Corbett 2005).

In H&R 2002, 1+2 homophony can arise through neutralization of the Addressee node, and 1+2+3 homophony can arise through complete underspecification of the Participant distinction. By contrast, there is no way for 1+3 or 2+3 homophony patterns to be derived via feature neutralization. This leads to the expectation that 1+2 homophony will arise systematically and will therefore be more common than both 1+3 and 2+3 homophony, which should in turn be equally unlikely to arise (i.e., only through accidental homophony). This does not match up with the typological counts in figure 7.

Ackema and Neeleman (A&N) (2018) propose an alternative theory of person partly designed to better account for the relative frequency of these homophony patterns. A&N redefine the person space in terms of two privative person features, PROX (for *proximate*) and DIST (for *distal*). In line with Harbour (2016), they interpret these features as functions that operate on an input set to deliver a subset as output. The crucial aspect of this account for our purposes is that the semantic specification of these two features implies that one is shared by first and second person (PROX), while the other is shared by second and third person (DIST).[22] The immediate consequence is that both 1+2 and 2+3 homophony can be generated, depending on which one of the two features (PROX or DIST) is left underspecified. In contrast, no feature is shared uniquely by first and third person (while excluding second person), ruling out the existence of a 1+3 homophony pattern.[23]

A&N also predict an asymmetry between 1+2 and 2+3 homophony patterns (Peter Ackema, pers. comm.). In their system, the second person category is the product of applying first PROX

[22] In a nutshell, A&N (2018) assume that PROX and DIST features operate on a set containing all the possible sets of person referents as nested subsets. That is, the set containing all potential referents (S_{iuo}) contains a subset containing only speaker and addressee referents (S_{iu}), which in turn contains a subset containing only speaker referents (S_i). The feature PROX operates on an input set and discards its outermost "layer," whereas DIST selects this outermost layer. It is easy to see that in order to obtain, for example, the first person referent one would need to apply the PROX feature twice, as $\text{PROX}(S_{iuo}) = (S_{iu})$ and $\text{PROX}(S_{iu}) = S_i$.

[23] A similar asymmetry can roughly be derived in Harbour's (2016) theory. Note that this is generally in line with traditional accounts of person systems in which second and third person are grouped as a natural class under the feature [−speaker] (Forchheimer 1953, Noyer 1992, Pertsova 2011).

and then DIST. If a language does not have a spell-out rule for this feature structure, it might in principle collapse second person with first (based on the PROX feature) or with third (based on the DIST feature). However, according to A&N, there is a general principle (the "Russian Doll Principle") such that spell-out rules cannot apply to "inner" features without also mentioning "outer" features, while the reverse is possible. Given that PROX is the innermost feature in the second person structure, 2+3 homophony patterns are predicted to be more common than 1+2 ones.

To summarize, the crosslinguistic observation outlined above is well accounted for by A&N's approach (and Harbour's (2016)), whereas it is problematic for H&R's approach (and similar ones). However, the typological data this generalization is built on are extremely sparse and the magnitude of the differences is small: of approximately 200 languages sampled in Cysouw 2003, the most frequent 2+3 homophony pattern is attested in the pronominal systems of only 14 languages, and the least common 1+3 homophony pattern is attested in 1 (see Baerman, Brown, and Corbett 2005:60). The limitations discussed above for typological data are therefore present here in spades. But the issue is an important one: is the traditional and perhaps more intuitive distinction between discourse participants and nonparticipants central to the organization of the person space? Or is it one asymmetry among potentially many? In what follows, we use the experimental paradigm developed above to bring learnability data to bear on this question.

Using the extrapolation paradigm (cf. Experiment 1B), we measured learners' likelihood of inferring each of the relevant patterns after training on an incomplete paradigm. Participants were taught the meaning of two pronominal forms, which corresponded to a subset of person categories, and they were then tested on how they extrapolated these forms to the remaining category. For example, some learners were taught two distinct forms for first and second person and then were tested on which of those forms they used to express the held-out third person meaning. If they used the first person form to express the new third person meaning, then they inferred a 1+3 homophony pattern. A different pattern of extrapolation would indicate 2+3 homophony, as described in table 9. H&R (2002) predict that learners will be more likely to infer 1+2 homophony relative to both 2+3 and 1+3 homophony. A&N (2018) predict that learners will be equally likely to infer either 2+3 or 1+2 homophony, but less likely to infer 1+3 homophony.

4.2 Methods

This experiment, including all hypotheses, predictions, and analyses, was preregistered at https://osf.io/z9a35. All materials, data, and scripts can be found at https://osf.io/pf7q6/?view_only=96814a5352e54f91a3941b6895e02810. This experiment was implemented using the JavaScript library jsPsych (De Leeuw 2015) and presented in a web browser.

4.2.1 Design Participants were randomly assigned to one of three conditions, summarized in table 9. Conditions differed on which subset of two-person categories was used for training and which category was held out. The training set determined which patterns of homophony participants could extrapolate to: Condition 1 was consistent with 1+2 and 1+3 patterns, Condition 2 with 1+2 and 2+3, and Condition 3 with 2+3 and 1+3. The specific predictions of each account

Table 9
Conditions in Experiment 3. There were two training and one held-out (in boldface) categories per condition. Each training category was mapped to a different pronominal form (A or B), schematically represented with light and dark gray. Participants had to use the training forms they learned (A or B) to express the held-out meaning (cells with *A or B?*). The two rightmost columns state which of the compatible paradigms participants were predicted to infer under Harley and Ritter's (2002, H&R) and Ackema and Neeleman's (2018, A&N) accounts.

	Mappings		Compatible paradigms	H&R	A&N
Condition 1	**1**	A or B?	1+2, 1+3	1+2 > 1+3	1+2 > 1+3
	2	A			
	3	B			
Condition 2	1	B	1+2, 2+3	1+2 > 2+3	1+2 ≈ 2+3
	2	A or B?			
	3	A			
Condition 3	1	B	1+3, 2+3	2+3 ≈ 1+3	2+3 > 1+3
	2	A			
	3	A or B?			

given this design are summarized in the two rightmost columns of table 9. Note that both accounts make the same predictions for Condition 1, but they differ in their predictions for Conditions 2 and 3.

All participants were additionally exposed to two pronominal forms that correspond to the singular alternatives of the plural pronouns they were trained on. The person categories for singular forms were always the same as the critical plural forms for a given participant, and were determined by condition (see table 9). For example, participants in Condition 1 were additionally exposed to second and third person *singular* pronouns.

4.2.2 Materials The language consisted of four different pronominal forms: two plural forms (critical categories) and two singular forms (filler categories). These four lexical items were drawn from the same list of eight CVC items used in Experiments 1 and 2. The reference of the pronouns was expressed by highlighting a subset of family members, as in table 6, except that in this experiment the inclusive category was never expressed. Visual stimuli were the same as in Experiments 1 and 2.

4.2.3 Procedure After being introduced to the general backstory (see Experiment 1), participants were instructed to figure out the meanings of the words in the new language. Participants were given an example trial with an English pronoun (*her* or *me* depending on the condition) that would help them understand that the words they were learning were pronouns. As in the previous experiments, the speaker and addressee roles switched during the experiment to reinforce the context-dependent meaning of the forms.

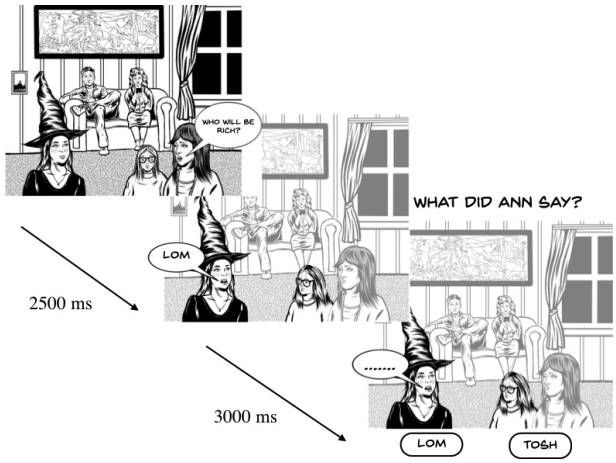
The experiment had two training phases followed by a testing phase, the structure of which was exactly as described for Experiment 1B. The only difference was in the person categories instantiated by the highlighting (see figure 8). Briefly, the two training phases were composed of exposure and *What if . . .* trials; the testing phase involved trials in which a referent set was highlighted and participants had to choose the corresponding form. Participants were given feedback after exposure and *What if . . .* trials, but not after testing trials. The order of presentation of meanings was fully randomized within phases for each participant.

In the first training phase (16 trials), participants were trained on two singular (filler) forms. After this first training, participants were asked to type in a meaning for the two words they had learned, and they were given feedback on their answers. Unlike in Experiment 2, participants were not excluded on the basis of their performance with these singular items in this phase, since they were told what they meant and they were not used for extrapolation. In the second training phase (28 trials), participants were trained on both filler and critical meanings. Finally, in the testing phase the critical held-out meanings were added. There were 24 trials in this phase, 8 of which were repetitions of the held-out meanings. As in previous experiments, participants completed a debrief questionnaire at the end of the experiment. A summary of the procedure is given in the online appendix (table A.3).

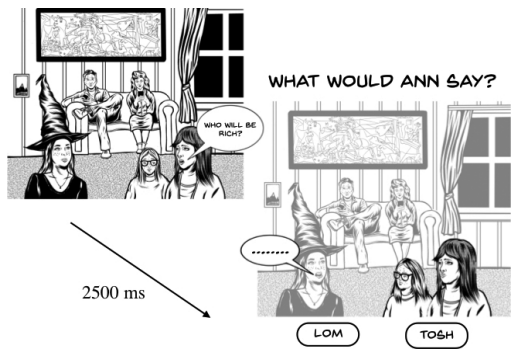
4.2.4 Participants A group of 259 English-speaking adults (Condition 1: 74, Condition 2: 86, Condition 3: 99) who had not participated in one of our previous experiments were recruited via Amazon Mechanical Turk. Five additional participants were excluded for not being self-reported native speakers of English, and 30 for failing to pass the two attention checks included in the experiment: one at the beginning of the experiment (before starting the training), and a second before starting with the testing. Per our preregistration, participants who did not pass both attention checks did not contribute to our sample. Per preestablished exclusion criteria (as in Experiment 1), participants who failed to perform accurately in at least two-thirds of each training category (4/6) during the testing phase were excluded from the analyses. The data from 152 participants were kept for the analyses (Condition 1: 48, Condition 2: 54, Condition 3: 50). All participants were paid 2.5 USD for their participation, which lasted approximately 15 minutes.

4.3 Results

Recall that participants were taught two pronominal forms (coded as forms A and B), which they had to use to describe both a critical set of two trained categories and a third, held-out meaning. Figure 9 shows the proportion of trials on which participants chose the pronominal form coded as A during the testing phase, for each category and condition. For Condition 1, figure 9 shows a mixed pattern of responses for the held-out meaning (first person plural): some participants used the trained second person form (coded as A), some used the third person form (coded as B), and some behaved randomly (i.e., no consistent pattern of response). By contrast, participants in Conditions 2 and 3 appear to have inferred a consistent paradigm: in both cases, participants largely used the same form (coded as A) for second and third person meanings (regardless of what the trained meaning was).



(a) Exposure 1PL category



(b) Testing 2PL category

Figure 8
Illustration of exposure and testing trials in Experiment 3

To compare performance across conditions, we ran a logistic mixed-effects model predicting form A choices in held-out trials by condition (3 levels). Recall that the meaning of the pronominal form coded as A differed depending on condition (see Mapping in table 9). The model used treatment coding, with Condition 2 as baseline, and included random by-participant intercepts.

The model revealed that participants in Condition 2 (baseline) were significantly above chance in selecting form A for the held-out category ($\beta = 4.15$, $p < .001$***). In this condition, the trained meaning of form A was third person, and the held-out category was second person; therefore, this result confirms that participants inferred a 2+3 (rather than a 1+2) homophony pattern. In Condition 3, the trained meaning of form A was second person, and the held-out category was third person. Therefore, if participants in this condition were not significantly differ-

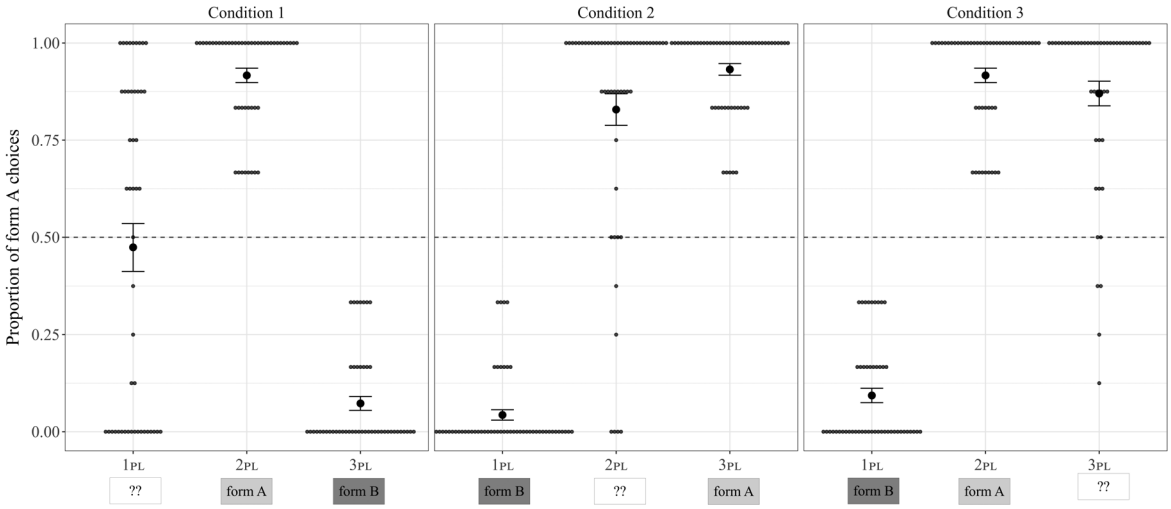


Figure 9
Proportion of form A choices by condition during the testing phase. Choice of the same form across categories indicates homophony. Error bars represent standard error on by-participant means; dots represent individual participants' means.

ent from the baseline, they presumably also inferred 2+3 homophony, and to the same degree as participants in Condition 2. This is confirmed by the model ($\beta = 0.614$, $p = .047$). By contrast, the model revealed a significantly lower proportion of form A responses for held-out items in Condition 1 compared with the baseline ($\beta = -4.8$, $p < .001$***). In this condition, the trained meaning of form A was second person, and the held-out category was first person. Therefore, we can conclude that participants did not infer 1+2 homophony to the same degree that participants in Conditions 2 and 3 inferred 2+3 homophony.

These results were further confirmed by two separate, intercept-only models for Conditions 1 and 3. For Condition 3, the proportion of form A responses was significantly above chance ($\beta = 3.78$, $p < .001$***), again indicating that as in Condition 2, Condition 3 participants inferred a 2+3 rather than a 1+2 system. By contrast, the proportion of form A responses was not above chance for Condition 1 ($\beta = -0.72$; $p = .29$), indicating that Condition 1 participants did not show a clear preference between 1+2 and 1+3 homophony.[24]

4.4 Discussion

Results from Experiment 3 show that participants consistently inferred paradigms that feature 2+3 homophony whenever their training was compatible with this (Conditions 2 and 3). That is,

[24] This is further confirmed by a debrief questionnaire where participants were asked to provide information about the meanings of each of the pronouns. Most participants in Conditions 2 and 3 reported that the second and third person meanings were mapped into the same form.

learners preferred systems that collapse the second and third person categories to alternative homophony patterns. When 2+3 homophony was not available to participants (Condition 1), there was no stable pattern of responses: some participants used the same form for first and second persons, some used the same form for first and third, and some alternated randomly between the two.

This result is in direct contrast to the prediction we derived from Harley and Ritter's (2002) theory. This theory is designed to derive 1+2 homophony, and *not* 2+3 homophony. Further, it does not distinguish 2+3 from 1+3 homophony. Ackema and Neeleman's (2018) theory fares slightly better, as it posits that second and third person form a natural class (DIST) that excludes first person. This predicts that 2+3 homophony patterns should be preferred over 1+3 patterns, in accordance with our results (Forchheimer 1953, Bobaljik 2008; see also binary-features accounts). However, by also positing a feature shared between first and second person (PROX), this theory predicts an asymmetry between 1+2 and 1+3 homophony, not attested in our findings (similar predictions are made by Harbour (2016)). Thus, while the pattern of results in our experiment is quite clear, it does not straightforwardly match the predictions of either account.

Our results, however, do mirror what one might deem the most obvious typological asymmetry in figure 7: paradigms featuring 2+3 homophony are the most common (see table 9). Interestingly, an analogous pattern is found in the distribution of person systems across sign languages, where second and third person are consistently homophonous both in pointing signs and in other grammatical constructions (Meier 1990, Neidle et al. 2000), as well as in the spatial deixis domain in languages across both modalities: most locative systems (e.g., *here*/*there* in English) rely on a distinction between spaces in the vicinity of the speaker (1) and spaces not in the vicinity of the speaker (2+3) (Harbour 2016).[25] This match between our results and typology suggests the possibility that some additional force is at play, which distinguishes first and second person, even if they do form a natural class.[26]

To summarize, recall that at issue here was the special status of the [±participant] distinction in driving homophony. Theories based on the traditional set of binary features, like Harley and Ritter's (2002), are designed to capture this particular natural class, thus predicting systematic

[25] In principle, the English-speaking participants in our experiments might be extrapolating their experience with the locatives *here* and *there*, which make a 2+3 partition of the space, to the person space (Daniel Harbour, pers. comm.). However, in our view this would be rather surprising.

[26] It is worth briefly discussing two alternative interpretations of our experimental results. First, it could be that the little sister was treated as a speech act participant rather than as a third person referent—for example, because she is spatially close to the speaker and hearer in the illustrations (Peter Ackema, pers. comm.). To address this possibility, we reran one condition of this experiment with a different set of images where the little sister was seated in the background together with the parents. The results replicate the findings reported here, suggesting this is not an issue (see the online appendix, figure A.3). Second, participants might think the speaker (the girl with the hat) is in some cases directly addressing the others (e.g., parents) instead of the intended hearer (the girl asking the questions). If an addressee shift of this sort were at play, there would be no "true" third person, explaining why participants derive what seems to be a 2+3 pattern (as suggested by Klaus Abels, pers. comm.). However, in the postexperiment debrief, where participants had to provide meanings for each of the words they had learned, most participants (across all three conditions) described the pronominal form they used to refer to nonparticipants in the conversation (e.g., the parents) as meaning *them* in English. This shows that they were representing one of the pronominal forms as a true third person, and did not interpret the speaker as directly addressing the others.

homophony between speech act participants, that is, first and second person. Our results do not bear this out. Rather, behavior appears to be consistent with a preference to distinguish the speaker from others: person systems that distinguish between participants and nonparticipants—making use of the PROX or [±participant] feature—seem to be dispreferred over systems that make use of the DIST or [±speaker] feature. The preference for 2+3 systems in our experiment might then lie in participants' unwillingness to lose a distinction they have learned between forms that include and forms that exclude the speaker.[27] This bias for speaker distinctiveness was partly discussed in the context of Zwicky's generalization and Experiment 2 above. We therefore return to this issue in the next section.

5 General Discussion

The main aim of the experiments reported here was to bring behavioral evidence from learning to bear on how person systems are represented and whether they are subject to universal constraints on possible partitions. We see our results as making three main contributions: (a) we provide confirmatory evidence that the person space is represented in terms of primitive features, (b) we point to the need for restrictions on patterns of inclusive homophony based on weak biases rather than hard-and-fast constraints, (c) we provide new evidence for an asymmetry between 2+3 and other two-way homophony patterns involving 1 (EXCL+INCL), 2, and 3 person categories. We first summarize each of these in turn and then discuss a number of broader issues raised by our results as a whole.

In Experiment 1, we sought to provide evidence that the person space is represented as an interacting set of universal features. While this idea forms the basis of most theories of person in the theoretical linguistics literature, to date the link with learnability has been left implicit. However, these theories make testable predictions. If learners represent the person space as the interaction of a (primitive) set of person features, then categories that form a natural class should be readily mapped to the same phonological form (as formalized in Pertsova 2011, among many others). The learner must simply determine which feature(s) are underspecified. A set of cells that cannot be characterized as constituting a natural class arguably requires learners to first learn the different categories (i.e., as feature bundles) and then independently learn to map the relevant set to a phonological form. Thus, a feature-based theory of the person space predicts that paradigms with this kind of random homophony should be less readily learned than homophony among meanings with a shared feature. By contrast, if learners perceive the person space as a simple conjunction of four person categories (first, inclusive, second, and third), then there is no basis for predicting that some homophony patterns should be more learnable than others.

The predictions of feature-based approaches to person were borne out in Experiment 1: homophony patterns among forms sharing a feature were indeed easier to learn and were more

[27] The extrapolation paradigm we use is designed to target precisely this kind of effect. An ease-of-learning design, by contrast, may be more likely to tap purely into learners' perceptions of how natural it is for two categories (e.g., first and second persons) to share a form. For some evidence suggesting this, see https://osf.io/9nc36/?view_only=09b7115978e14ae0898d4b8fb47fb4b6.

likely to be inferred when extrapolating a known form to a new meaning. This was the case even though the specific features involved are not active in the participants' native language, suggesting that this is how these systems are represented. These findings draw an obvious parallel with work in phonology showing that rules that can be characterized on the basis of features (or natural classes) are easier to learn than rules that group together a set of featurally distinct segments (e.g., Saffran and Thiessen 2003, Cristià and Seidl 2008, Moreton and Pater 2012). This suggests that feature-based representations are relevant for learning across domains, although the set of relevant features is of course domain-specific.

In Experiment 2, we turned to Zwicky's (1977) well-known observation that person partitions without an inclusive distinction assimilate inclusive with first person (rather than second or third). Different approaches to person treat this apparently strong asymmetry differently. Harbour's (2016) and Ackema and Neeleman's (2018) theories are purpose-built to derive first-inclusive tripartitions only (though Ackema and Neeleman's may also be able to generate second-inclusive homophony).[28] However, approaches like Harley and Ritter's (2002) can derive both first-inclusive and second-inclusive (though the latter is more marked).

Again, translating the predictions of these theories in terms of learnability, we tested whether homophony of inclusive with first, second, or third person was treated distinctly by learners. We found that third-inclusive homophony was particularly problematic, while first-inclusive had only a weak advantage over second-inclusive. In accordance with Harley and Ritter's (2002) theory, this suggests that there is no hard constraint against systems that systematically collapse second and inclusive persons; rather, inclusive can form a natural class with both first and second person, though not (readily) with third.

Finally, in Experiment 3 we provided new evidence for a learning-based asymmetry between partitions with homophony between second and third person (2+3) on the one hand, and both first-second (1+2) and first-third (1+3) homophony patterns on the other. While the former were readily inferred by our participants, the latter were not. This result roughly matches an asymmetry found in the typological distribution of pronominal systems. However, it is problematic under the assumption that all homophony patterns targeting meanings that share a feature should be equally possible. Indeed, theories like Harley and Ritter's (2002), which posit that first and second person form a salient natural class—of speech act participants—would predict that 1+2 homophony should if anything have a learnability advantage. Alternative approaches such as Harbour's (2016) and Ackema and Neeleman's (2018) posit common features between both first and second categories, and between second and third; therefore, their most straightforward prediction is a specific learnability disadvantage for 1+3 homophony. Participants in Experiment 3 instead consistently inferred 2+3 homophony whenever this was consistent with the input they were trained on, while 1+2 and 1+3 were equally dispreferred.

[28] Ackema and Neeleman's (2018) system does derive second-inclusive systems (violating Zwicky's generalization) whenever (a) PROX and PROX-PROX have different phonological realizations, and (b) there is an impoverishment of DIST in the plural when it is a dependent of PROX.

To summarize, our results suggest that learners represent the person space in units that are smaller than person categories (features) and that instantiate natural classes. Natural-class-based similarity therefore clearly plays an important role in determining how humans partition this person space. Indeed, a bias toward patterns based on natural class similarity pushes learners to preferentially collapse categories that share features (when they are required to do so). Our results support the claim that inclusive and second person should be included in the set of natural classes in this domain. However, there is also reason to believe that a bias for keeping the speaker distinct may also be at play. Indeed, our findings are compatible with the idea that natural-class-based similarity, and speaker- (or ego-)based distinctiveness may be two independent forces influencing the learnability of person systems. In Experiment 2, the partition characterized by first-inclusive homophony was learned most readily and is consistent with both of these pressures: it involves homophony among meanings that share a feature, [+speaker] *and* keeps person categories implicating the speaker distinct from other categories. The second-inclusive homophony pattern is the next best, involving homophony among shared meanings but not maintaining speaker categories as distinct from others. In Experiment 3, these two pressures led learners to infer 2+3 whenever they could—again among the options presented to learners in this case, 2+3 homophony is the only one to involve both categories that are highly similar and forms that keep the speaker category distinct. The fit is not perfect: in principle, as noted above, we might have expected the first-inclusive advantage to be stronger, and 1+2 homophony to be preferred over 1+3. One possibility is that our failure to find these differences might be the result of particular features of our experimental design. For example, a difference in learnability between 1+2 and 1+3 might be revealed in an ease-of-learning experiment (see Experiment 2), where resorting to the preferred 2+3 pattern is not possible.

In our discussion of Experiment 2, we suggested that a speaker distinctiveness bias may be the result of the cognitive importance of the ego (see Dixon 1994). Indeed, research on early pronoun acquisition has argued for an egocentric bias, whereby children perceive the world as a function of their presence in it and adopt categorization systems that carry this distinction (Charney 1980, Loveland 1984, Moyer et al. 2015). This is also supported by the fact that perspective-taking appears to be a capacity that takes time to develop; infants and young children do not always succeed on so-called theory-of-mind tasks that require them to recognize that their internal knowledge states are not the same as other people's (see, e.g., Ruffman 2014). If the pressure for keeping the speaker distinct in pronoun systems comes from a general egocentric bias, then we might expect (a) the bias to be stronger in children, and (b) the bias to be stronger when learners are active speakers of the language. The experiments reported here of course involved adults, and our participants were never themselves the speaker. However, further research could test both predictions.

An alternative is to build a speaker-based asymmetry between natural classes directly into a theory of person—at the same level of representation as the domain-specific primitive features. In a nutshell, the idea would be that the natural class "speaker" is somehow represented as special within the person space. The pressure for speaker distinctiveness would then be a special type of natural-class-based similarity that comes not directly from the set of primitive features but

from the particular status of the speaker feature. This sketch is very speculative, and a worked-out implementation is beyond the scope of this article. However, this idea is in the spirit of Harbour (2016) and Ackema and Neeleman (2018), whose theories encode an inherent asymmetry between speaker and addressee, built into the ontology (see footnote 15). Crucially, we would argue that this asymmetry should be treated as a bias, which shapes but does not strictly delimit the space of possible partitions.

Before moving on, we should mention an anonymous reviewer's concerns about whether our results could be accommodated by a non-feature-based approach to person. This is a sensible worry, as it questions whether our approach truly provides evidence that the person space is represented in terms of a set of universal primitives. To address this concern, let us say a few words about whether the results reported here are compatible with putative nonfeatural representations of the person space.

Our experiments show that learners are sensitive to whether certain semantic values (e.g., speaker, addressee) are included in the pronominal reference, regardless of the specific person category. This suggests that person categories form natural classes along these semantic dimensions. Here, we have treated these natural classes as defined by primitive features. Arguably, an alternative, non-feature-based approach would be one that does not decompose person categories into smaller units, but treats them as primitives themselves; for example, first, inclusive, second, and third would be four different primitives, as well as all other plural combinations (see Cysouw 2007 for discussion). We see a problem with this type of alternative view: in order to account for our main findings, this approach would need to explain why it is that learners are more likely to collapse some person categories than others, even when these categories are nonnative.

As a response to this issue, the same reviewer conjectures that some additional pragmatic mechanism might result in more accurate learning predictions. For the time being, however, we remain agnostic about how this proposal could be further developed into a full theory of person that accounts for our results (and for the typology). Instead, we maintain that, to the extent that learners are shown to be sensitive to the semantic overlap between categories, a feature-based account of our results is more parsimonious (as well as in line with already established theories of person).

6 Conclusion

Person systems have been extensively explored from a theoretical standpoint: a number of approaches have been proposed, each of which constrains the set of possible person partitions that humans can represent, with the aim of explaining the prevalence of certain partitions of the person space crosslinguistically. The experiments reported here inform these theoretical approaches by generating direct behavioral evidence for the impact of hypothesized representations and constraints on the learnability of different person partitions. Indeed, our results constitute the first experimental evidence for learnability differences in this domain.

Specifically, we have provided evidence that there is a universal basis for a set of primitives organizing the person space that learners are sensitive to regardless of their native language. Looking more closely into the nature of these primitive features, we have shown that a theory

of the person space needs to account for the semantic similarity between inclusive and second persons, on the one hand, and between second and third persons, on the other. Each of these pairs of categories was treated as a natural class by learners, suggesting that they have features in common (e.g., an addressee feature possibly shared between second and inclusive persons). Besides a preference for feature-based patterns, there was evidence across our experiments that participants have an additional bias toward partitions of the person space where the speaker is distinct from other categories. Thus, even though both speakers and addressees are participants in the conversation, there is an inherent asymmetry in how learners treat them. We sketched two possible accounts of this, the first in terms of a general egocentric bias, and the second in terms of a special type of natural-class-based similarity.

More generally, our results suggest that the experimental methods developed here provide a novel tool for testing theoretically motivated questions about how languages carve up the person space. While these kinds of methods have gained traction in investigating learning biases in a number of linguistic domains, here we have highlighted the sparsity of typological data as underscoring the need for new sources of evidence in building theories of person.

References

Ackema, Peter, and Ad Neeleman. 2013. Person features and syncretism. *Natural Language and Linguistic Theory* 31:901–950.

Ackema, Peter, and Ad Neeleman. 2018. *Features of person: From the inventory of persons to their morphological realization.* Cambridge, MA: MIT Press.

Baerman, Matthew, and Dunstan Brown. 2013. Syncretism in verbal person/number marking. In *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The syntax–morphology interface: A study of syncretism.* Cambridge: Cambridge University Press.

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. Lme4: Linear mixed-effects models using Eigen and S4. *R package version* 1:1–23.

Béjar, Susana. 2003. Phi-syntax: A theory of agreement. Doctoral dissertation, University of Toronto.

Benveniste, Emile. 1971. *Problems in general linguistics.* Miami, FL: University of Miami Press.

Bickel, Balthasar. 2008. A refined sampling procedure for genealogical control. *Language Typology and Universals* 61:221–233.

Boas, Franz. 1911. Introduction to the handbook of North American Indians. *Smithsonian Institution Bulletin* 40:1–83.

Bobaljik, Jonathan David. 2008. Missing persons: A case study in morphological universals. *The Linguistic Review* 25:203–230.

Brown, Cynthia A. 1997. Acquisition of segmental structure: Consequences for speech perception and second language acquisition. Doctoral dissertation, McGill University.

Carstensen, Alexandra, Jing Xu, Cameron Smith, and Terry Regier. 2015. Language evolution in the lab tends toward informative communication. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, ed. by David C. Noelle, Rick Dale, Anne S. Warlaumont, Jeff Yoshimi, Teenie Matlock, Carolyn D. Jennings, and Paul P. Maglio, 303–308. Austin, TX: Cognitive Science Society.

Charney, Rosalind. 1980. Speech roles and the development of personal pronouns. *Journal of Child Language* 7:509–528.
Chemla, Emmanuel, Brian Buccola, and Isabelle Dautriche. 2019. Connecting content and logical words. *Journal of Semantics* 36:531–547.
Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.
Corbett, Greville G., Dunstan Brown, and Matthew Baerman. 2002. Surrey syncretism database. https://www.smg.surrey.ac.uk/syncretism/.
Cowper, Elizabeth, and Daniel Currie Hall. 2009. Argumenthood, pronouns, and nominal feature geometry. *Determiners: Universals and Variation* 147:97–120.
Cristià, Alejandrina, and Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4:203–227.
Culbertson, Jennifer. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6:310–329.
Culbertson, Jennifer. To appear. Artificial language learning. In *Oxford handbook of experimental syntax*, ed. by Jon Sprouse. New York: Oxford University Press.
Culbertson, Jennifer, and David Adger. 2014. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111:5842–5847.
Culbertson, Jennifer, Annie Gagliardi, and Kenny Smith. 2017. Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92:343–358.
Culbertson, Jennifer, Paul Smolensky, and Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122:306–329.
Cysouw, Michael. 2003. *The paradigmatic structure of person marking.* Oxford: Oxford University Press.
Cysouw, Michael. 2005. Quantitative methods in typology. In *Quantitative Linguistik: Ein internationales Handbuch*, ed. by Reinhard Köhler, Gabriel Altmann, and Rajmund G. Piotrowski, 554–578. Berlin: De Gruyter Mouton.
Cysouw, Michael. 2007. Building semantic maps: The case of person marking. In *New challenges in typology*, ed. by Matti Miestamo and Bernhard Wälchli, 225–248. Berlin: De Gruyter Mouton.
Cysouw, Michael. 2010. On the probability distribution of typological frequencies. In *The mathematics of language*, ed. by David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, et al., 29–35. Berlin: Springer.
Cysouw, Michael. 2013. Inclusive/Exclusive distinction in independent pronouns. In *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Daniel, Michael. 2005. Understanding inclusives. In *Clusivity*, ed. by Elena Filimonova, 3–48. Amsterdam: John Benjamins.
De Leeuw, Joshua R. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* 47:1–12.
Dixon, R. M. W. 1994. *Ergativity.* Cambridge: Cambridge University Press.
Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473:79–82.
Fedzechkina, Maryia, Florian T. Jaeger, and Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109: 17897–17902.
Forchheimer, Paul. 1953. *The category of person in language.* Berlin: de Gruyter.
Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T. Piantadosi, and Bevil R. Conway. 2017. Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences* 114:10785–10790.
Greenberg, Joseph H. 1988. The first person inclusive dual as an ambiguous category. *Studies in Language* 12:1–18.

Hale, Kenneth. 1973. Person marking in Walbiri. In *A festschrift for Morris Halle*, ed. by Stephen R. Anderson and Paul Kiparsky, 308–344. New York: Holt, Rinehart and Winston.
Halle, Morris, and Alec Marantz. 1994. Some key features of Distributed Morphology. In *Papers on phonology and morphology*, ed. by Andrew Carnie and Heidi Harley, with Tony Bures, 275–288. MIT Working Papers in Linguistics 21. Cambridge, MA: MIT, MIT Working Papers in Linguistics.
Hanson, Rebecca. 2000. Pronoun acquisition and the morphological feature geometry. *Calgary Working Papers in Linguistics* 22. https://prism.ucalgary.ca/handle/1880/51442.
Hanson, Rebecca, Heidi Harley, and Elizabeth Ritter. 2000. Underspecification and universal defaults for person and number features. In *Actes du Congrès annuel de l'ACL/CLA Annual Conference Proceedings*, 111–122. Ottawa: University of Ottawa, Cahiers Linguistiques d'Ottawa.
Harbour, Daniel. 2008. On homophony and methodology in morphology. *Morphology* 18:75–92.
Harbour, Daniel. 2011. Descriptive and explanatory markedness. *Morphology* 21:223–245.
Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90:185–229.
Harbour, Daniel. 2016. *Impossible persons.* Cambridge, MA: MIT Press.
Harley, Heidi. 2008. When is a syncretism more than a syncretism? Impoverishment, metasyncretism, and underspecification. In *Phi theory: Phi-features across modules and interfaces*, ed. by Daniel Harbour, David Adger, and Susana Béjar, 251–294. Oxford: Oxford University Press.
Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.
Ingram, David. 1978. Typology and universals of personal pronouns. In *Universals of human language.* Vol. 3, *Word structure*, ed. by Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, 213–248. Stanford, CA: Stanford University Press.
Kay, Paul, and Terry Regier. 2007. Color naming universals: The case of Berinmo. *Cognition* 102:289–298.
Kemp, Charles, and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science* 336:1049–1054.
Ladd, D. Robert, Seán G. Roberts, and Dan Dediu. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1:221–241.
Loveland, Katherine A. 1984. Learning about points of view: Spatial perspective and the acquisition of 'I/you'. *Journal of Child Language* 11:535–556.
Maldonado, Mora, and Jennifer Culbertson. 2019. Something about us: Learning first person pronoun systems. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, ed. by Ashok Goel, Colleen Seifert, and Christian Freksa, 749–755. Montreal, QB: Cognitive Science Society.
Martí, Luisa. 2020. Inclusive plurals and the theory of number. *Linguistic Inquiry* 51:37–74.
Martin, Alexander, and Sharon Peperkamp. 2020. Phonetically natural rules benefit from a learning bias: A re-examination of vowel harmony and disharmony. *Phonology* 37:65–90.
Martin, Alexander, Theeraporn Ratitamkul, Klaus Abels, David Adger, and Jennifer Culbertson. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* 5(1). https://doi.org/10.1515/lingvan-2018-0072.
McGinnis, Martha. 2005. On markedness asymmetries in person and number. *Language* 81:699–718.
Meier, Richard P. 1990. Person deixis in American Sign Language. *Theoretical Issues in Sign Language Research* 1:175–190.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25:83–127.
Moreton, Elliott, and Joe Pater. 2012. Structure and substance in artificial-phonology learning. Part I: Structure. *Language and Linguistics Compass* 6:686–701.
Moyer, Morgan, Kaitlyn Harrigan, Valentine Hacquard, and Jeffrey Lidz. 2015. 2-year-olds' comprehension of personal pronouns. https://www.bu.edu/bucld/files/2015/06/Moyer.pdf.
Neidle, Carol, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. 2000. *The syntax of American Sign Language: Functional categories and hierarchical structure.* Cambridge, MA: MIT Press.

Nevins, Andrew, Cilene Rodrigues, and Kevin Tang. 2015. The rise and fall of the L-shaped morphome: Diachronic and experimental studies. *Probus* 27:101–155.
Noyer, Rolf. 1992. Features, positions and affixes in autonomous morphological structure. Doctoral dissertation, MIT.
Pagel, Mark, Quentin D. Atkinson, and Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449:717–720.
Perfors, Amy, Joshua B. Tenenbaum, and Terry Regier. 2011. The learnability of abstract syntactic principles. *Cognition* 118:306–338.
Pertsova, Katya. 2011. Grounding systematic syncretism in learning. *Linguistic Inquiry* 42:225–266.
Piantadosi, Steven T., and Edward Gibson. 2014. Quantitative standards for absolute linguistic universals. *Cognitive Science* 38:736–756.
Piantadosi, Steven T., Joshua B. Tenenbaum, and Noah D. Goodman. 2016. The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review* 123:392–424.
R Core Team. 2018. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.
Rodrigues, Aryon D. 1990. You and I = Neither you nor I: The personal system of Tupinamba. In *Amazonian linguistics: Studies in lowland South American languages*, ed. by Doris L. Payne, 393–405. Austin: University of Texas Press.
Rouse, Steven V. 2015. A reliability analysis of Mechanical Turk data. *Computers in Human Behavior* 43:304–307.
Ruffman, Ted. 2014. To belief or not belief: Children's theory of mind. *Developmental Review* 34:265–293.
Saffran, Jenny R., and Erik D. Thiessen. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39:484–494.
Saldana, Carmen, Yohei Oseki, and Jennifer Culbertson. 2019. Do cross-linguistic patterns of morpheme order reflect a cognitive bias? In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, ed. by Ashok Goel, Colleen Seifert, and Christian Freksa, 994–1000. Montreal, QB: Cognitive Science Society.
Sauerland, Uli, and Jonathan David Bobaljik. 2013. Syncretism distribution modeling: Accidental homophony as a random event. In *Proceedings of GLOW in Asia IX*, ed. by Nobu Goto, Koichi Otaki, Atsushi Sato, and Kensuke Takita, 31–53. Tsu, Japan: Mie University.
Schwartz, Linda J., and Timothy Dunnigan. 1986. Pronouns and pronominal categories in Southwestern Ojibwe. In *Pronominal systems*, ed. by Ursula Wiesemann, 285–322. Tübingen: Gunter Narr Verlag.
Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, ed. by R. M. W. Dixon, 112–171. Canberra: Australian National University.
Sokolovskaja, Natalja K. 1980. Nekotorye semantičeskie universalii v sisteme ličnyx mestoimenij. In *Teorija i tipologija mestoimenij*, ed. by Igor F. Vardul, 84–103. Moscow: Nauka.
Sonnaert, Jolijn. 2018. *The atoms of person in pronominal paradigms.* Amsterdam: Netherlands Graduate School of Linguistics (LOT).
Steinert-Threlkeld, Shane, and Jakub Szymanik. To appear. Ease of learning explains semantic universals. *Cognition.*
Tabullo, Angel, Mariana Arismendi, Alejandro Wainselboim, Gerardo Primero, Sergio Vernis, Enrique Segura, Silvano Zanutto, and Alberto Yorio. 2012. On the learnability of frequent and infrequent word orders: An artificial language learning study. *Quarterly Journal of Experimental Psychology* 65:1848–1863.
Thomas, David. 1955. Three analyses of the Ilocano pronoun system. *Word* 11:204–208.
Wechsler, Stephen. 2010. What 'you' and 'I' mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86:332–365.
White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93:1–36.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.
Zaslavsky, Noga, Charles Kemp, Naftali Tishby, and Terry Regier. 2019. Color naming reflects both perceptual structure and communicative need. *Topics in Cognitive Science* 11:207–219.
Zwicky, Arnold M. 1977. Hierarchies of person. In *Papers from the Thirteenth Regional Meeting, Chicago Linguistic Society*, ed. by Woodford A. Beach, Samuel E. Fox, and Shulamith Philosoph, 714–733. Chicago: University of Chicago, Chicago Linguistic Society.

Mora Maldonado
Centre for Language Evolution
University of Edinburgh
mora.maldonado@ed.ac.uk

Jennifer Culbertson
Centre for Language Evolution
University of Edinburgh
jennifer.culbertson@ed.ac.uk