Number Nativism

Abstract: Number Nativism is the view that humans innately represent natural numbers precisely. Despite a long and venerable history, it is often considered hopelessly out of touch with the empirical record. I argue that this is a mistake. After clarifying Number Nativism and distancing it from related conjectures, I distinguish three arguments which have been seen to refute the view. I argue that, while popular, two of these arguments miss the mark, and fail to place pressure on Number Nativism. Meanwhile, a third argument is best construed as a challenge: rather than refuting Number Nativism, it challenges its proponents to provide positive evidence for their thesis and show that this can be squared with apparent counterevidence. In response, I introduce psycholinguistic work on The Tolerance Principle (not yet considered in this context), propose that it is hard to make sense of without positing precise and innate representations of natural numbers, and argue that there is no obvious reason why these innate representations couldn't serve as the basis of mature numeric conception.

You don't have to be a mathematician to have a feel for numbers.
John Nash

How do humans acquire the capacity to represent precise natural numbers, like SEVEN, ELEVEN, or SIXTY-TWO? That they do is uncontentious. Natural numbers may be our best understood abstract objects. Axioms have been formulated from which their properties can be deduced (Peano 1908). Even primary-school children gain an appreciation for certain mathematical inferences they permit (Butterworth 1999). Yet, no less obviously, numbers are weird things. You can't point at the number seven, or bump into the number eleven, since entities of this sort are not located in space and time. Consequently, it's been hard to see how we could first *learn* of natural numbers, or come to manipulate them precisely, such that consensus would emerge on their basic structural properties (e.g., that each natural number is exactly one greater than its predecessor, or that any two collections of N objects stand in a relation of 1:1 correspondence). In short: This would involve us forming an exact and shared conception of entities that exist outside the physical world with which we interact.

Number Nativism helps defuse these concerns. For the Nativist, our capacity to represent precise natural numbers is not learnt – it is part of our innate biological endowment. This leaves untouched vexed issues concerning the representation of numbers. For instance, one might wonder how a representation, learnt or innate, could refer to numbers if they're located off in Plato's heaven (Benacerraf 1965). Regardless, it's a virtue of the Nativist proposal that it avoids the abovementioned problem of how our internal states might first support precise numerical inferences, given that numbers are entities with which we lack material acquaintance – this is because Nativists deny that our initial capacity to represent and process natural numbers results from learning through observation or acquaintance. And since Nativists deem this innate numerical endowment common to all neurotypical individuals, it's not hard to see why consensus emerges on many of the natural numbers' basic structural properties – this is akin to the way an innate and biologically endowed universal grammar constrains and ensures commonalities among the structural features of natural languages (Chomsky 1986).

Despite these virtues, Number Nativism is an unpopular view. A long and venerable history in the work of Plato to Chomsky notwithstanding, contemporary theories of cognitive development share in the assumption that our capacity to represent precise natural numbers is learnt (Bloom 2000; Dehaene 1997; Hurford 1987), and leading theories of concept learning take the acquisition of precise number concepts to be their central target (Carey 2009; Spelke 2017; Lee & Sarnecka 2010). Indeed, the suggestion that our capacity to represent precise natural numbers is not innate is often deemed non-negotiable – as dictated by "unambiguous evidence" that the acquisition of precise numerical representations is a product of culture and learning environment (Pitt et al. 2022: 371).

I think Number Nativism is in considerably better standing than this modern consensus recommends. After clarifying Number Nativism and distancing it from related conjectures (Section 1), I'll distinguish three arguments which have been seen to refute the view (Section 2). I argue that, while popular, two of these arguments miss the mark, and fail to place any pressure on Number Nativism. Meanwhile, a third argument is best conceived of as a challenge: rather than refuting Number Nativism, it challenges its proponents to provide positive evidence for their thesis and show that this can be squared with apparent evidence to the contrary. In response, I'll introduce psycholinguistic work on The Tolerance Principle (not yet considered in this context), note that it's hard to make sense of without positing innate representations of precise natural number, and argue that there's no obvious reason why these innate representations couldn't serve as the innate basis of mature numeric conception (Section 3).

1. What is Number Nativism?

1.1 Number Nativism

Number Nativism is the view that humans innately represent natural numbers, such as SIX, SEVEN, and TWENTY-THREE, precisely. Specifically, these representations are innate in the sense that they are biologically endowed, and not acquired through a psychological process of learning (Shea 2011).

This does not tie Nativists to the view that these innate resources are present at birth, much as an innate and unlearnt capacity to grow pubic hair does not imply its presence at birth. Nor does it require that they are evolutionarily ancient and/or shared with other animals, much as positing an innate language faculty does not require that chimpanzees can talk. Finally, it does not imply that humans are innately endowed with an indefinitely large number of syntactically atomic symbols, each of which represents a distinct natural number distinctly – i.e., one atomic symbol for 62, another atomic symbol for 63, and so forth, perhaps *ad infinitum*. More plausibly, Nativists allow that humans have an unlearnt capacity to represent a small range of natural numbers (say, ONE, TWO, and THREE – Margolis 2020) plus the representational resources needed to combine these such that larger natural numbers can be defined in terms of these more basic resources (Chomsky 1988; Leslie et al. 2008). This latter possibility – sometimes labelled 'the building blocks model' (Margolis & Laurence 2011) – would, thus, allow humans to represent arbitrarily large natural numbers in terms of their initial unlearnt repertoire (Fodor 1980; Pinker 1994). It thereby contrasts with anti-nativist proposals according to which our first representations of precise natural numbers are discontinuous with the expressive potential of our innate mental representations (Carey 2009; Spelke 2017), at least when these precise natural numbers are larger than ~3 and fall outside the subitizing range (Margolis 2020).

1.2 Number Nativism and Core Cognition

To better understand Number Nativism's claim that humans innately represent *precise* natural numbers, it's instructive to contrast it with a more popular suggestion: that humans are innately endowed with 'core systems', like an approximate and/or small number system, which facilitate basic numerical computations in circumscribed ways.

1.2.1 <u>The Approximate Number System</u>

The approximate number system (ANS) is a psychological system which facilitates numerical discriminations among relatively large collections of items throughout the lifespan (Barth et al. 2003; Izard et al. 2009). While there's disagreement over the system's architecture and algorithms (Dehaene & Changeux 1993; Yousif & Keil 2021) a defining feature of the ANS is that its numerical discriminations are imprecise and conform to Weber's Law. Thus, accuracy is predicted by the ratio between the numerical quantities in question, rather than their absolute difference (Libertus & Brannon 2010). As such, 9 is better discriminated from 10 than 10 is from 11 even though 9 differs from 10 by the same amount as 10 does from 11. What matters is the ratio between these values – the further from 1:1 the better (Figure 1).

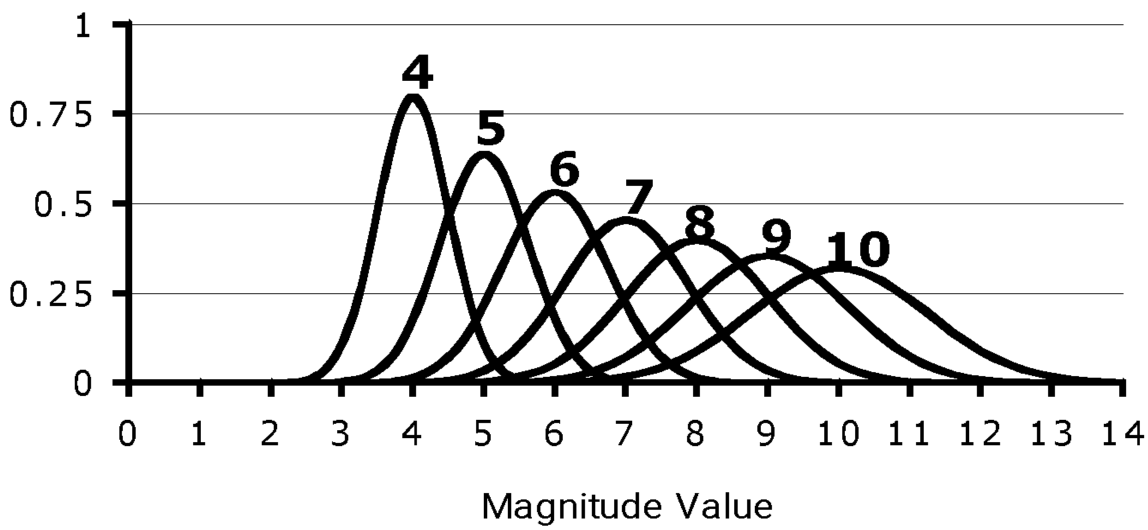

Figure 1. The ANS conforms to Weber's Law in its discriminations: As the magnitude (specified along the x-axis) increases, values which differ by the same absolute amount will be discriminated less-reliably (as illustrated on the y-axis).

The precise natural number representations Number Nativism posits are quite different. Precise natural number representations represent independent natural numbers as categorically distinct: e.g., TEN as categorically different from NINE, ELEVEN, or any other determinately specified number. However, the ANS's conformity to Weber's Law is typically seen to reflect indeterminacy and imprecision in the content of its representations. One reason to say this is that humans will happily apply a single exact number word, like '20', to collections which are discriminably different for their ANS – e.g., to collections of 15 or 25 – when estimating their cardinality (Sullivan & Barner 2013). This indicates that the ANS attributes indeterminate values to the collections it enumerates – for instance, it might attribute to these a numerical range (Ball 2017), a "blur on the number line" (Spelke & Tsivikn 2001), or a probability distribution (Halberda 2016). And since the range of acceptable values that it attributes to a collection increases with collection size (Izard & Dehaene 2008), it seems that the ANS's conformity to Weber's Law results (at least in part) from imprecision or indeterminacy in the content of its representations increasing as values get bigger. In this way, accurately discriminated collections of ten are represented as more different from accurately discriminated collections of nine than accurately discriminated collections of eleven, obscuring the categorical distinctness of each natural number.

This has downstream consequences. Beyond licensing categorical discriminations, precise natural number representations license precise judgements of equinumerosity. If two collections are represented as containing precisely N items, this licenses the deductively valid conclusion that they are equinumerous, irrespective of the number N concerns. Again, ANS representations fall short in this regard, owing to their imprecision (Carey & Barner 2019). Just as representing two collections as roughly 7 (but perhaps 6 or 8) in number fails to establish that they are identical in number, the apparent imprecision of the ANS's representations renders it incapable of establishing 1:1 correspondence among collections. For both reasons, the precise natural number representations Number Nativists posit should be distinguished from innate ANS representations, which represent numbers, but only imprecisely (Clarke & Beck 2021a).

1.2.2 <u>The Small Number System</u>

Beyond an ANS, many theorists posit an innate small number system (SNS). Unlike the ANS, this system facilitates precise numerical discriminations. However, it only does so among small collections, containing <3-4 items.

One reason for positing an SNS is that human adults are highly accurate when perceptually discriminating/enumerating collections containing 4 items or less (Mandler & Shebo 1982; Trick & Pylyshyn 1994). Thus, their small number discriminations, within this subitizing range, show no sign of the ratio dependence that's characteristic of the ANS. Similar results are found in babies. Classic work with 10-month-olds found that (on certain tasks) they would readily distinguish 2 items from 3 items while remaining at chance discriminating 3 items from 6 items (Feigenson et al. 2002). Since the ratio among these latter quantities is larger, it appears that performance was not underwritten by an

ANS, whose discriminations conform to Weber's Law. Rather, infants are seen to possess a separate SNS, which facilitates precise discriminations, but only among collections containing <3-4 items.

What's crucial to note is that, while the precision of the SNS evades certain shortcomings of the ANS, its postulation must still be distinguished from Number Nativism. For a start, most characterisations of the SNS deny that it represents any numerical content whatsoever. Rather, vision scientists propose that its numerical discriminations are implemented by a system of visual indexes, which pick out seen items like a series of pointing fingers implemented in the visual system (Scholl & Leslie 1999). Meanwhile, developmentalists often hold that the SNS employs a system of mental models, or representations of individual items, which are not proprietarily visual and can be flexibly stored in working memory (Simon 1997; Le Corre & Carey 2007). Despite marked differences, these accounts hold that the set size limit of the SNS is implied by the number of visual indexes available to the visual system or the number of mental models that can be stored in working memory. But on neither account does this involve the SNS representing numbers. When two items disappear behind a screen, and infants are surprised when the screen is removed to reveal just one object (Wynn 1992), these proposals explain infants' surprise by suggesting that they have tracked or represented two individuals/items behind the screen when only one individual/item was found. In other words, they appeal to the tracking or representation of items/individuals, rather than those items'/individuals' numerical quantity. In this way, the postulation of an SNS, thus construed, differs rather dramatically from Number Nativism – a thesis which posits innate *representations of number*.

Admittedly, some defend a richer interpretation of the SNS, arguing that it produces a small stock of precise number representations, pertaining to subitizable values – ONE, TWO, and THREE (Hurford 1987; Wynn 1992; Margolis 2020). On these accounts, the SNS's inability to discriminate larger quantities does not result from working memory limitations, or a fixed number of visual indexes. It results from the SNS's inability to represent values >3. Even so, the existence of an innate SNS still falls short of Number Nativism. This is because it does not imply a capacity to represent larger natural numbers precisely, such as SEVEN, ELEVEN, or SIXTY-TWO. If nothing else, additional resources are needed to enable these smaller values, encoded by the SNS, to be summed or recombined (Chomsky 1988; Leslie et al. 2008a), such that the summing and recombination of these representations enables larger values to be represented without expanding on the expressive capacities of our innate representational repertoire (Fodor 1975; 1980; 2008). Suffice to say, the acquisition of these additional resources – for instance, grasp of the successor function, enabling ONE to be added to any successive value – is typically considered a hard-won developmental milestone (Carey & Barner 2019). So, while Number Nativism is consistent with holding that precise natural number representations of SEVEN, ELEVEN, and SIXTY-TWO, are constructed from the small number representations an SNS produces (plus additional resources – e.g., SUCCESSOR FUNCTION), it's agnostic on this and would go beyond the postulation of an SNS, regardless.

2. Arguments Against Number Nativism

I'm not the first to distinguish Number Nativism from the postulation of an ANS and/or SNS. Proponents of these 'core systems' typically distance themselves from Number Nativism, recommending that they will enable large and precise natural number concepts (e.g., SEVEN) to be *learnt* (Carey 2009; Spelke 2017). Whether or not this is correct (Fodor 2010; Rey 2014), there's a widespread assumption that our initial capacity to represent precise natural numbers outside the subitizing range is not innate, given strong (Spelke 2017) or "unambiguous evidence" (Pitt et al. 2022: 371) that our capacity to do so depends on culture and learning environment. So, while there are strong reasons to posit an innate ANS and SNS, there is supposed to be similarly strong reason to deny that humans are innately equipped to represent large natural numbers precisely.

This is a mistake. To see why, I'll now distinguish three arguments which motivate these anti-nativist conclusions. While two of these arguments feature prominently in the literature, they miss the mark, and fail to place pressure on Number Nativism. Meanwhile, a third argument is best understood as a challenge – rather than showing Number Nativism to be false, it challenges the Nativist to provide positive evidence for their hypothesis and show that this evidence can be squared with apparent evidence to the contrary. This challenge will be considered in Sections 3, where I introduce psycholinguistic work on The Tolerance Principle and argue that it is hard to make sense of without positing innate representations of large and precise natural numbers.

2.1 <u>The Argument from Anthropology</u>

One argument, which is often said to refute Number Nativism, concerns cross-cultural work. Here, it's claimed that monolingual speakers of languages which lack precise number words – such as Pirahã, which simply contains words for *one, two,* and *many*, or Munduruku, which lacks words for numbers bigger than 5 – remain systematically incapable of representing precise natural numbers outside their linguistic count range (Gordon 2004; Pica et al. 2004). Indeed, it is work of this sort which provides the "unambiguous evidence", mentioned above, that our capacity to represent precise natural numbers is dependent on culture and learning environment and, hence, not innate (Pitt et al. 2022: 371).

Take Pica and colleagues' (2004) pioneering work with the Munduruku. These researchers reported that monolingual speakers of Munduruku were in possession of an intact ANS but remained systematically incapable of counting collections of dots outside their linguistic count-range; at best, they could rely on heuristic strategies such as "matching their fingers and toes to the sets of dots" (500). Around the same time, Gordon (2004) reported similar findings with the Pirahã, motivating his Whorfian hypothesis that the ability to conceive of precise number depends on one's native language.

More recently, Pitt and colleagues (2022) reported converging evidence with speakers of Tsimane. Unlike Munduruku and Pirahã, Tsimane has a fully productive system of number words. Nevertheless, Tsimane individuals' education and knowledge of these number words varies considerably. Taking this into account, Pitt and colleagues tested participants on "a simple numerical matching task" and reported that their discrimination of "exact cardinalities" was "limited to the number words [each participant] knew" (371). For instance, participants who only knew words for 1-7 only made accurate and precise discriminations among these values. So, unlike the ANS, and its representations of approximate number (which these researchers consider an innate human universal), this was seen to provide "unambiguous evidence that large exact number concepts are enabled by language" and therefore emerge as a complex product of one's culture and learning environment. But, if our initial acquisition of precise natural number representations is a result of culture and learning environment, rather than innate biological endowment, Number Nativism is false. This is The Argument from Anthropology.

How convincing is this argument? Despite its prominence in the literature, *not very*. It's been questioned whether the abovementioned groups are truly bereft of the precise number words these Whorfian conclusions assume (Nevins et al. 2008). Moreover, the argument may be criticised for conflating numerical performance (or lack thereof) with numerical (in)competence (more on this below). But even putting these issues aside, the conclusion that our capacity to represent precise natural numbers depends on language/culture neglects straightforward evidence to the contrary.

In one study, Butterworth et al. (2008) tested indigenous children from the Walpiri and Anindilyakwa communities of Northern Australia. Like Munduruku and Pirahã, Walpiri and Anindilyakwa are limited in the precise number words they contain (roughly speaking, they only have words for *one two* and *many*, and *one two three* and *many*, respectively). Even so, Butterworth and colleagues found that monolingual children from both communities would succeed in accurately identifying precise quantities up to 9 (the largest value tested) and be able to later recall these quantities in a memory task. They were also found able to accurately perform a cross-modal matching task, matching a precise number of seen counters to a precise number of tones in a heard sequence, and to correctly answer simple addition and division tasks concerning precise quantities. Importantly, these children performed no worse than English speaking children, and their performance showed no discontinuity at or around the subitizing threshold, indicating that performance was not the result of their SNS or ANS. These results would, thus, seem to contradict Pitt and colleagues' assertion that precise numeric conception depends on one's possession of precise number words.

Related results are not hard to find. Follow up studies with the Walpiri and Anindilyakwa, replicated Butterworth et al.'s findings. For instance, Reeve et al. (2018) found that language failed to predict performance on a novel addition task and, further, showed that residual differences in performance between English speaking children and members of the above communities were better predicted by differences in visuo-spatial working memory than numerical understanding. Similar results have also emerged in the Amazonian communities which originally motivated The Argument from Anthropology. In one study, Frank et al. (2008) failed to replicate Gordon's (2004) original results with the Pirahã, finding that monolingual speakers of Pirahã *could* perform precise numerical discriminations when the tasks used did not require participants to *remember* numerical quantities. Likewise, Izard et al. (2008) found that monolingual speakers of Munduruku would succeed in making precise numerical discriminations when tasks were simplified or appropriately framed. For instance, where Pica et al.'s (2004) study tested the Munduruku's capacity for precise enumeration using a subtraction task, wherein large numbers were subtracted from one another to produce values under five (i.e., values that could be stated using the Munduruku's limited number vocabulary), participants in Izard et al.'s task were tested on a simpler match-to-sample task – they were simply asked to match the number of seeds in a collection to the number of dots on a screen. In this simpler task, they "spontaneously" succeeded, answering accurately and precisely, with occasional errors poorly predicted by the acuity of their ANS (Izard et al. 2008: 501).

What, then, should make of The Argument from Anthropology? At best, we should recognise that the evidence supporting it is complicated. It does not straightforwardly refute Number Nativism. But, putting things more forcefully, the conflicting findings described in this section plausibly evince Number Nativism. If a genuine absence of precise enumeration is hard to find, even in isolated communities with few number words and little in the way of formal mathematical education, this motivates thinking that the acquisition of precise natural number concepts is probably a human universal, just as Number Nativism boldly predicts.

2.2 The Argument from Linguistic Development

The Argument from Anthropology does not warrant the rejection of Number Nativism it is sometimes claimed to mandate. Indeed, cross-cultural work may ultimately evince Number Nativism. Nevertheless, The Argument from Anthropology is not the only reason Number Nativism is rejected. A similarly popular Argument from Linguistic Development is also advanced in this connection. While this argument has a similar flavour to the Argument from Anthropology, it appeals to an independent body of research, concerning children's protracted development on number discrimination tasks.

To this end, proponents of the Argument from Linguistic Development often emphasise the Give-N task in which children, who have begun to utter some portion of the count-list in their native language ('one, two, three, four…'), are asked to give an experimenter or puppet a specific number of items (e.g., six toys). A robust result is that children's performance on this task proceeds slowly through several stages. First, children become "one-knowers" who reliably give one item when asked but randomly produce >1items when asked for any larger number of items. Next, children become "two-knowers" who reliably pass one item when asked or two items when asked but randomly produce >2 items when asked for larger quantities. Eventually, after passing through several such stages (e.g., becoming "three-knowers" etc.), children become "cardinal-principle knowers" who appreciate that each successive value in their count list refers to a number that is one greater than its predecessor (Wynn 1990; 1992; Lee & Sarnecka 2010). Since performance on this task predicts performance on other numerical comprehension tasks, like the "What's on this card?" task, in which children are asked to report the number of items seen on a card (Le Corre et al. 2007), a standard interpretation is that each knower-level reflects mastery of a new natural number. One-knowers have learnt to represent ONE precisely but lack the capacity to represent TWO, while two-knowers have learnt to represent

ONE and TWO precisely but lack the capacity to represent THREE, etc. Thus, performance on the Give-N task is taken to show that the count list "is first mastered much as children learn to recite the alphabet, that is, without attributing any significance to the order" (a bit like the ordered, but meaningless, rhyme "eenie-meanie-miney-mo") and "that knowledge of the counting principles is not innate, but rather constructed as a result of children's attempt to make sense of the verbal count list" (Le Corre & Carey 2008: 651).

The problem with this assessment is that the argument (as formulated) conflates the comprehension of precise number words with the acquisition of precise number concepts (Margolis 2020). The above results show that children enjoy a slow and protracted development learning to use and respond to number labels in natural language – e.g., the word 'seven', such that they can reliably produce seven items when asked. However, this is something on which Number Nativism is silent. A child in possession of an innately endowed concept SEVEN might simply struggle to match this up with a precise number word in natural language. Thus, "children's slow, step-by-step learning of number words could reflect their difficulty mapping their number concepts onto language, rather than limitations to the number concepts themselves" (Spelke 2017: 151). Indeed, this is more than a possibility. For we have just found reason to posit a disconnect between precise number language and precise numeric conception when we found that monolingual speakers of languages lacking precise number words nevertheless possess a capacity for precise enumeration (Butterworth et al. 2008). Such a disconnect is even independently motivated, e.g., by research finding that certain autistic individuals, whose acquisition of language is severely impaired, nevertheless succeed in learning mathematics and often develop superior calculation skills to neurotypicals (Cowan & Frith 2009).

It will be replied that these suggestions fail to explain why children are quite so slow to acquire number words, given evidence that other word meanings are acquired rapidly (Carey 2009; Samuels & Snyder 2024: Ch.3.3). For instance, two-year-olds can learn novel nouns from "a single ambiguous exposure" (Spiegel & Halberda 2011), while the learning of number words takes years. If children were already in possession of the relevant number concepts, shouldn't the comprehension of number words emerge quicker than that?

I think not. The reason is that word-referents are not created equal. Rather, the 'fast mapping' of novel words onto pre-existing concepts is subject to independently supported constraints (Medina et al.

2010), like a 'whole object bias' which disposes children to assume that novel words label whole observable objects (Markman 1991). The upshot is that children are slow to learn the meanings of words referring to the parts, predicates, and properties of objects, even after concerted teaching efforts (Hansen & Markman 2009), especially when these properties and predicates are abstract. Accordingly, 3-year-old children struggle to acquire words for categorical colours, when these are used to describe objects (Landau & Gleitman 1985), even though young infants possess a well-established capacity to discriminate colour categories (Bornstein 1976).

What's important to note is that we should expect the same points to apply to number words. After all, the whole (abstract) object, *six,* is not something we can point at when we teach children the referent of 'six'. What we can do is point at collections containing six items and state of each collection that 'the items are six in number'. But when we do 'six' is used to describe rather than label the collection. Thus, it functions like the colour terms children are so slow to acquire (owing to the *whole object bias*).[1] Consistent with this suggestion, children grasp precise number words at around the same time they grasp categorical colour words (4 years of age – Landau & Gleitman 1985), even though number is less visibly salient than colour. It is also consistent with the fact that prior to grasping the meanings of the colour words, children behave like subset knowers: For instance, three-year olds use the word 'blue' in grammatically and semantically apt ways (e.g., in sentences like "The cup is blue") but apply it to red and blue items with equal likelihood (Rice 1980). Thus, Number Nativists find independent reason to reject the assumed connection between numerical language and numerical conception implied by the Argument from Linguistic Development, and the implication that innate or early emerging number concepts would facilitate a fast mapping of these onto number words.

2.3 The Argument from (Non-Linguistic) Performance

The Arguments from Anthropology and Linguistic Development are unconvincing. This is significant, since they're probably the most common reasons for thinking Number Nativism implausible. Even so, it might be replied that their failings serve to highlight the real reason Number Nativism is falsified: that children also fail to perform precise numerical discriminations in *non-linguistic* tasks.

[1] This predicts that if children were not subject to the hypothesized word-learning biases, learning number words would come quickly. It is notable, therefore, that isolated adult populations *do* seem to learn (first) number words rather quickly. For instance, Ken Hale, the great documenter of Warlpiri and other Australian languages, wrote that "the English counting system is almost *instantly* mastered by Warlpiris who enter into situations where money is important" (Dixon 1980: 107-8, *my emphasis*).

This is the conclusion drawn by Spelke (2017). To make her point, Spelke emphasises a single study by Izard et al. (2014). In an initial experiment, children (just under three-years-old) were presented with a tree containing 6 branches – they were then presented with either 5 or 6 puppets, each of which was placed on an independent branch of the tree. The experimenter proceeded to emphasise that each puppet only occupied one branch and that one branch was empty if and when it was (i.e., in the 5-puppet condition). They then asked each child to help place the puppets into an opaque box and "rock them to sleep". Finally, the children "woke the puppets up" and placed them back on the branches of the tree. Importantly, however, when the experimenter secretly removed the sixth puppet from the container in the 6-puppet condition (leaving just 5 puppets in the box, when the children should have expected 6) the children spent considerably longer rechecking the box for a final puppet than in a 5-puppet condition (where they shouldn't have expected an additional sixth puppet). It was as if children had enumerated the puppets at the beginning of the experiment, represented them as being exactly 5 or exactly 6 in number (depending on the condition), and were then surprised to find a different number of items returned from the opaque container.

Spelke rejects this numerical interpretation. She considers a subsequent manipulation on the above task. In this manipulation, half the children saw one puppet being removed from the container and then returned to it. Having done so, children who were assigned to a 6-puppet condition continued searching for a sixth puppet when one had been secretly removed from the container (leaving just 5 items). This is as we would expect if children were representing and keeping track of the precise number of puppets observed (and, hence, that the 5 puppets produced $\neq$ the 6 puppets in the box). However, when one puppet was removed from the container and replaced by a featurally indistinguishable (yet numerically distinct) doppelganger, in full view of the children, children failed to search longer for a missing sixth puppet in the 6-puppet condition than in an otherwise identical 5-puppet condition. In other words, children's ability to track the number of puppets went away. To make sense of this result, Spelke proposes that children were not considering the number of puppets on the tree or in the box after all; rather, they had simply remembered the specific individuals originally located on the tree and were subsequently seeking to place *those* individuals back on the tree.

My own view is that Spelke's explanation of this result is, itself, problematic. Wouldn't remembering 6 distinct individuals exceed 3-year-olds' working memory limit? After all, Spelke has herself proposed that working memory limitations explain why children fail catastrophically on subitizing tasks involving >3 items belonging to the same set (Feigenson et al. 2004). I'll put this to one side. Suppose we accept Spelke's interpretation of Izard et al.'s result: What would this show? It would show that children neglect to represent or use precise numerical information in a task where adults might. But plainly, this would not show that they lack the competence to do so.

Consider the experiments described in Section 1. When describing the empirical evidence for an SNS, we noted Feigenson et al.'s (2002) study in which 10-month-old infants accurately discriminated 2:3 items but were at chance discriminating 3:6. This is striking since 3:6 differs by a larger absolute amount and a larger Weber fraction. Indeed, it differs by a ratio which younger infants are known to reliably discriminate using their ANS, even when tested on a small number of trials (Libertus & Brannon 2010). Thus, children in Feigenson et al.'s task failed to discriminate 3 from 6 (i.e., exhibited a failure of performance in this task) even though they possess a well-established competence to do so using their ANS. For everything that has been said, we should recognise that the same *might* apply to the study Spelke emphasises when arguing that precise representations of natural numbers are lacking in young children; even if children did not represent or make use of precise numerical information this does not show that they lack the competence to do so.

This is not to suggest that Spelke's argument (which I'll now call The Argument from Performance) holds no weight. My point is that it must not be treated as a deductive proof that number representations are absent in the children tested, let alone a deductively valid refutation of Number Nativism. It's best construed as a challenge. Spelke is not confusing absence of evidence for evidence of absence. Her point is that we seem to lack positive reasons for thinking precise natural number representations present in young human children, even when we go looking for these and design tasks to avoid taxing language and other extraneous resources. Pending positive reasons of this sort, the postulation of innate natural number representations could then seem extravagant. Much as we should not posit innate knowledge of quarks, carburettors, or (Plato notwithstanding) the Pythagorean theorem if we have no evidence or reason to do so, a persistent failure to find evidence of the capacity to represent precise natural numbers in young children should increase our confidence that this much is lacking in the organisms with which we're concerned.

However, far from providing "unambiguous evidence" against Number Nativism, which mandates an account on which our initial capacity for precise natural number representation is learnt, this Argument from Performance is straightforwardly defeasible. Really, it just *challenges* Nativists to identify positive evidence that humans innately represent precise natural numbers and show that this can be squared with the above results. In response to this challenge, I'll now consider recent work in psycholinguistics, which has not yet been discussed in this context.

3. The Tolerance Principle and Number Nativism

The preceding arguments are something of a grab-bag, but there's a common thread. Each argument highlights an absence of numerical performance where Number Nativists, or some caricature thereof, might expect to find this. But while these claims are often overblown (as in The Argument from Anthropology) or problematically conflate the mental representation of numerical quantities with their linguistic expression (as in The Argument from Linguistic Development), it's true that they present Number Nativists with a challenge. The challenge is to provide positive evidence for innate representations of precise natural numbers and show that this evidence can be squared with the failures of numerical performance described above.

To this end, I'll now introduce psycholinguistic research on The Tolerance Principle (3.1). I'll argue that this is hard to make sense of unless young human infants have an innate (and heretofore unacknowledged) capacity to represent large natural numbers precisely (3.2). I'll then argue that there is no obvious reason why these innate representations couldn't form the basis of mature numeric conception (3.3).

3.1 <u>The Tolerance Principle</u>

'The Tolerance Principle' (TP) is a theorem which purports to specify when children will and won't endorse rule-like generalizations in language, treating these as productive rules to be generalized to new cases.

Consider the "add '-ed' to make verbs past tense" rule in English. This rule applies to many cases: if we add 'ed' to verbs such as 'stop', 'crouch', 'berate', and 'manage' we correctly produce their past-tense counterparts, 'stopped', 'crouched', 'berated', and 'managed'. And, sure enough, this is something young children readily appreciate, productively applying "add '-ed'" to novel cases.

A problem arises when we note that such rules admit exceptions: 'hold' becomes 'held' (not 'holded'), 'go' becomes 'went' (not 'goed'), etc. Irregular transformations of this sort are found in all natural languages (Sapir 1928) and often concern some of the most common words therein (Pinker 1999). Yet, despite their token frequency, young children draw a categorical distinction between these (type infrequent) irregular transformations and the (type frequent) regular transformations they observe. So, while overgeneralizations from (relatively common) regular transformations are among the most frequently documented errors in child morphology (Yang 2002), overgeneralisations from (type infrequent) irregular transformations are virtually absent (Yang 2016a). Thus, it is only when a rule/generalization applies to a sufficiently large proportion of types within the target domain that it "earns" its productive keep (Aronoff 1976; Plunkett & Marchman 1993; Bybee 1995).

This raises the question: How common must these generalizations be? When does a linguistic regularity hold sufficiently often that it is something from which children productively generalize?

Many answers to this question seem possible. We could imagine that rule-like generalizations are treated as productive iff they apply to >50% of cases in the target domain, or just in case they apply more often than chance. TP recommends a more nuanced answer: It proposes that a rule is treated as productive iff its treatment as such speeds up the average time with which it is accessed given two independently motivated background assumptions: *The Elsewhere Condition* and *Zipf's Law*.

> **The Elsewhere Condition** states that for a productive rule (e.g., "add '-ed'") to be applied to a given token (e.g., a given verb), exceptions to the rule must be considered and rejected as irrelevant, in series, and in rank order of frequency. Accordingly, the Elsewhere Condition states that entries in the target domain are processed one-by-one, as follows:

If A then B

If C then D

…

Otherwise: Apply the productive rule-like generalisation where ‘A’ and ‘B’ refer to transformations of the most common irregular in the target domain, ‘C’ and ‘D’ refer to transformations of the second most common irregular in the target domain, and so forth, and where it is only after all such irregulars have been considered and ruled out in series that the productive rule is applied. (This may sound like a bold conjecture. However, the Elsewhere Condition has been widely supposed in linguistic theorising [Anderson 1969; Halle 1997; Brown & Hippisley 2012] and it is supported by a host of empirical considerations. For instance, it is supported by the fact that the speed with which irregular transformation rules are processed, in both linguistic production and comprehension tasks, is inversely proportional to their rank-order frequency in the language [Marlsen-Wilson & Tyler 1997; Clahsen 1999; Pinker & Ullman 2002] and by the observation that irregulars, which do not permit productive generalisation, are processed faster than regulars [Yang 2016a: Section 3.3.2]. Moreover, it is worth noting that while the implementation of The Elsewhere Condition strikes critics as cumbersome, requiring that we laboriously represent the number of times each word in the lexicon has been tokened, efficient algorithms have been posited, showing how simple heuristics yield similar results [Rivest 1976; Sleator & Tarjan 1985; Yang 2016a].)

Zipf's Law refers to the (more mysterious) fact that when measured values in a target domain are sorted in decreasing order, the value of the n^{th} entry is often inversely proportional to n, such that: The most common type in the target domain appears roughly twice as often as the second most common type in the target domain, three-times as often as the second most common type in the target domain, and so forth (Zipf 1949). While Zipfian distributions pop up in surprising places (e.g., predicting city size in many countries) they are best established in the case of language. Thus, in the Brown Corpus for American English, the most commonly appearing word "the" accounts for roughly 7% of all word occurrences, while the second-most-common word "of" applies to roughly 3.5% of word occurrences, and so on (Fagan & Gençay 2010). In other words, the second most common word appears roughly half as often as the first, and the third most common word appears roughly a third as often as the first, etc.

Of course, Zipf's Law is an approximation – albeit an approximation that holds across frequency distributions of many kinds, including countless sub-domains of language (ibid.). What matters here is that it enabled Charles Yang (with help from Sam Guttman) to formulate TP: proving that if linguistic rules are processed in accord with The Elsewhere Condition, and entry types follow a Zipfian distribution in their token frequency, then the encoding of the productive rule will increase the mean speed with which entries are processed just in case:

$$E \leq \text{Ø}_N \text{ where } \text{Ø}_N := N/\log N$$

where 'E' refers to the number of exceptions to the rule-like generalisation and 'N' refers to the total number of entry types known in the target domain. Put differently, TP proposes that a rule-like generalisation should be treated as productive iff the number of exceptions to the generalization does not exceed the total number of types encountered in that domain divided by the natural logarithm of that total (Yang 2018). Conformity to this rule ensures that (on average) entries are processed quicker, if we assume the Elsewhere Condition and a Zipfian distribution.

Working through the details of Yang's proof is beyond the scope of this paper (see Yang 2016a or 2018 for details). What matters here is that TP makes precise predictions about when children should productively generalise rules they have encountered. For instance, given a rule R that applies to a target domain with 9 items, TP predicts that 4 (or 4.096 or Ø9=9/ln9) exceptions can be tolerated before its treatment as a productive rule becomes less efficient than storing individual items on a case-by-case basis in one long look-up table. Thus, strict adherence to TP implies that only when the number of exceptions sits below this threshold will the rule be deemed productive, and categorically so (i.e., generalized to 100% of novel cases). Meanwhile, a larger target domain containing 20 types will allow that more of these (up to 6) might violate a productive rule therein. In any case, TP predicts that the relative proportion of tolerable violations will decrease as domain size gets bigger (Figure 2). This has the intriguing consequence that productive rules are easier to learn when one knows less words in a language (Yang 2016a: 66).

N of 10	Allows 4 tolerable exceptions	40%
N of 20	Allows 6 tolerable exceptions	30%
N of 50	Allows 12 tolerable exceptions	24%
…		
N of 1,000	Allows 144 tolerable exceptions	14.4%

Figure 2: The proportion of tolerable exceptions the Tolerance Principle permits decreases with the number of types within the domain.

So far, TP might sound like a normative thesis, about when children *should* deem linguistic rules productive (given certain background assumptions). Thus, it might seem orthogonal to the project of describing actual human psychology. What's remarkable, is that TP has been found to predict actual patterns of productive generalization with astonishing accuracy.

Demonstrations of this astonishing accuracy take several forms. An initial indication that TP may be descriptively adequate is that it predicts the frequency with which regular and irregular word types appear in corpus data, across many languages, including Polish, Russian, German, North American English (Yang 2016a), Cree (Henke 2022), and Early Modern English (Ringe & Yang 2022). TP is also found to explain otherwise puzzling discrepancies in cross-cultural linguistics. For example, it's been noted that children begin productively applying the rules of their native languages' count lists at different points. Thus, children do not begin to productively apply the rules of the English count list until they have learnt to count to 72 (at which point they have a eureka moment: "the next number must be 'seventy-three', just like how 'sixty-three' followed 'sixty-two'!" – Fuson et al. 1982). In a stunning vindication of his theory, Yang (2016b) showed that, given the number of exceptions to the productive rules governing the English count list, this is the threshold at which TP predicts productive generalization to obtain. He also noted that the threshold is lower in Mandarin – where productive generalisation takes off at 40 – and again showed that this is what TP predicts.

More relevant for us are a suite of carefully controlled developmental studies. In one study, Schuler et al. (forthcoming, Exp.1) presented children, aged 5-7, with 9 noun types from one of two artificial languages. In either case, the experimenter produced both the "singular" and "plural" form for each entry. In one language, 5 of the nouns shared a plural suffix ('ka') while 4 did not. In this condition, TP predicts that children should productively generalize the regularity, since it predicts an allowance of 4.096 (Ø9=9/ln9) exceptions before productivity breaks down. Meanwhile, the second language contained just 3 nouns with the shared suffix 'ka' – while this made 'ka' more common than any other suffix, it left 6 exceptions to the rule (well above the threshold of 4.096). Sure enough, when children were exposed to the first language, they applied the "add '-ka'" rule to novel nouns 92% of the time (which was not significantly different from 100%, indicating categorical application). Meanwhile, in the second language, children applied the "add '-ka'" rule just 16.9% of the time (which was not significantly different from chance, despite 'ka' appearing 3x more than any other suffix encountered).

The fact that children productively generalised, categorically, and in line with the thresholds predicted by TP, bears out bold predictions of Yang's theory. However, a second experiment from Schuler et al.'s paper was even more telling. This second experiment was identical to the first except the token frequency of nouns in the artificial languages varied dramatically. This allowed that regular nouns could appear with low frequency and irregular nouns could appear with high frequency, making for a more ecologically valid test set which closely resembles the frequency distributions found in (some) natural language domains (Pinker 1999; Yang 2016a).

Under these conditions, it initially appeared that TP was falsified. Specifically, it was found that children tested on a language with 5 regular nouns and 4 exceptions applied the regular form just 54.63% of the time (significantly less than 100%, as TP predicts). However, the standard deviation of this result was large, and when results were analysed at the level of individual children, the categorical nature of their responses remained apparent: 16 of 20 children's responses were categorical, with 5 children effectively generalising the "add '-ka'" rule 100% of the time, and 11 failing to apply any observed suffix more often than chance.

This prompted the experimenters to consider the number of noun-types that each child had actually remembered (based on a rating test that was conducted after all trials). Taking this into account, the results proved to align with TP after all. For instance, children who only remembered 8 of the 9 noun-pairings, including all 4 exceptions, did not productively generalise the regular rule significantly more often than chance. This is what TP predicts, since the tolerance threshold for a domain with 8 noun types is 3.85 (and thus <4), not the 4.096 exceptions tolerated in a domain with 9 types. Indeed, TP's predictions bore out in this way in 15 of the 16 children whose behaviour was categorical. So, when children's word retention was taken into consideration, categorical generalisations conformed to TP, with 15 of 20 children effectively and categorically distinguishing (e.g.,) sets with 9 types and 4 exceptions from sets with 8 types and 4 exceptions.

Of course, this study probed 5–7-year-olds – children who have had plenty of time to learn stuff. However, similar results are found in young infants (Gómez & Lakusta 2004; Koulaguina & Shi 2013;

Koulaguina & Shi 2019). Perhaps most dramatically, Shi and Emond (2023) tested non-Russian-speaking 14-month-olds. These infants were exposed to 16 three-word sentences of Russian, which either conformed or failed to conform to a movement rule (ABC → BAC vs. ABC → ACB). In a domain containing 16 types, TP predicts that productive generalisation should occur if there are <5.77 exceptions. Thus, in a first experiment, where 11 sentences conformed to a movement rule and 5 did not, TP predicts that generalisation would occur. Sure enough, infants in this experiment looked significantly longer when a subsequent test stimulus failed to follow the rule that was implicit in 11 of the 16 exemplars. A second experiment then tested 14-month-olds on an identical training set comprising 16 sentence types, except that here only 10 sentences conformed to the rule, leaving 6 exceptions (i.e., a number now exceeding the TP threshold). Under these conditions, infants did not look significantly longer when the subsequent test item failed to conform to the rule than when it did not. Thus, looking behaviour implied that 14month-olds distinguish collections containing 11 regulars and 5 exceptions from collections containing 10regulars and 6 exceptions, in harmony with the predictions of TP, given a common set size of 16.

3.2 From TP to natural number

Work from various laboratories, utilising various experimental methods, bears out fine-grained predictions of TP. Most importantly, work with young children suggests that their tendency to treat linguistic generalisations as productive conforms to the categorical thresholds predicted by TP. I'll now argue that this is hard to make sense of unless these children possess a facility representing precise natural numbers, in a manner that is almost certainly innate. My argument for this claim proceeds in four steps. After making this argument, I'll explain how and why these innate representations could serve as the innate basis of mature numeric conception (Section 3.3).

Step 1: TP requires representing absolute quantities

Recall that TP thresholds are not a constant ratio of rule-conforming to rule-violating types. Thus, TP does not predict that children will treat a rule or generalization as productive iff it applies to (e.g.) ½, or ¾, of types within the target domain. Rather, TP predicts that the proportion of tolerable violations to a productive rule decreases with domain size. For instance, in a domain with 10-word-types, 4 (i.e., 40%) of these may violate a rule that is treated as productive; meanwhile a domain of 20-word-types can tolerate 6 (i.e., just 30%) exceptions, with larger domains tolerating a smaller proportion of exceptions still (Figure 2).

This raises the question: how could these precise thresholds be identified without first representing the quantity of word types in the domain? Since TP thresholds are not a constant ratio but are, instead, specified by a computational operation over the total (ever evolving) quantity of types encountered within the target domain, it's hard to see how the threshold could be identified without explicitly encoding the total quantity of types in that domain. And once the TP threshold has been identified, it's hard to see how one could identify whether a rule has/has not exceeded this threshold, such that productivity will/won't ensue, unless the absolute quantity of types that do/don't violate the rule are encoded. *Prima facie*, these quantities must be stored and represented.

Admittedly, this modest suggestion might be resisted by those advocating associationist theories of productivity (Goldberg 2019; Rummelhart & McClelland 1986). On these accounts, thresholds for productivity are set by associative mechanisms which eschew the need to represent quantities entirely. Instead, linguistic rules become more strongly associated with entries in the target domain following exposure to these. And when associations become suitably strong, productive generalization ensues. But not because the system has kept a record of the total quantity of types within the target domain, or the total quantity of rule-conforming/rule-violating types therein.

The trouble is: associative models fail to capture the thresholds for productivity observed in children (i.e., that predicted by TP). Associative models *do* explain how children might apply productive rules categorically. However, they overgeneralise from irregular forms inordinately more often than children (Marcus 1995; Yang 2016a; Yang 2018). In this way, associative models do not provide an empirically adequate alternative to a model on which TP thresholds are identified (and compared to quantities of rule-conforming and/or rule-violating types within the domain) via explicit representations of domain size and quantities of rule-conforming/rule-violating types therein.

<u>Step 2: These quantities include natural numbers</u>

If absolute quantities are represented in the service of implementing TP, which types of quantity? If TP requires representing the absolute quantity of types in the domain, and the absolute quantity of rule-conforming and/or rule-violating types therein, such that these representations can inform downstream computations, to which quantities do these representations refer? My suggestion is that, minimally, these quantities must include natural numbers.

Firstly, it's unclear what non-numerical quantities could support implementation of TP. As we've seen, TP identifies a threshold which is determined by the number of types in the target domain, and a rule is treated as productive just in case the number of rule-conforming types fails to exceed this threshold. To implement TP, such that these predictions are borne out, therefore requires that these numbers are tracked in some non-accidental way.

Tracking numbers is not the same as representing them, however (Butterfill 2018). To illustrate, consider a situation in which TP thresholds are identified by a module which uses non-numerical quantities as a proxy for number. For instance, the module might represent the total amount of time spent encountering word types in a domain and use this as a stand in for number. Provided that each word type is encountered roughly as often and is tokened for roughly the same duration as any other, a system which merely represents duration might effectively track the relevant numbers, enabling it to implement TP without representing these; after all, represented duration would hereby vary as a linear function of the numbers in question.

The trouble is: this heuristic strategy won't work in scenarios where TP succeeds. For one thing, we have seen that word frequencies follow Zipf's Law, such that common words appear exponentially more often than uncommon words. A system which merely uses total word duration as a proxy for number would, thus, be poorly modelled by TP. And indeed, duration and other potential confounds (e.g., cumulative loudness[?]) were effectively controlled for in the experiments described in Section 3.1. For instance, Schuler et al.'s second experiment manipulated token frequency such that regular types could appear with low frequency and irregular types could appear with high frequency. This ensured that the time spent encountering word types (or, e.g., their cumulative loudness?) could not be used as a proxy for number in the abovementioned ways. Conformity to TP obtained, nonetheless.

This brings me to my second point: Beyond the fact that it's unclear how performance in the abovementioned experiments would be explained by a system which simply represented non-numerical quantities, the implementation of TP involves representing quantities with properties that are unique to discrete numbers.

Consider Frege's (1884) insight that numbers differ from other quantities in their second-order character (i.e., in that they can only be assigned relative to a sortal). If I point at the boots in my closet and ask "How many?" Frege would note that this question is ill-posed. To answer it, we need to decide if we're interested in enumerating the individual boots in the closet, the pairs of boots in the closet, or the different boot types contained in the collection. In any case, the sortal needs specified since 16 individual boots might only amount to 8 pairs of boots, or 1 type of boot. Thus, the way in which we individuate the items has an impact on the number we attribute to the collection. What Frege observed is that non-numerical quantities are not like this. If we want to know how much the boots weigh, or what their volume is, it won't make any difference how the collection is carved up: irrespective of whether the collection is thought of as constituting a bunch of individual boots, pairs of boots, or types of boots it will take up the same amount of space in my closet and register the same weight on my scales. Numbers are, thus, distinctive among quantities for having a *second-order character*.

Frege's observation is pertinent in the current context since the quantities that are tracked and represented when implementing TP have a second-order character. The threshold specified by TP is set by the (whole) number of *types* in the domain, rather than their *token* frequency. Indeed, this is illustrated in the above experiments which control for this latter variable. This suggests that implementing TP involves tracking and representing quantities of words relative to the sortal *linguistic* (e.g., word) *type*, and this proceeds in abstraction from quantities assigned via the deployment of other sortals, like *linguistic* (e.g., word) *token*. Echoing Frege, this is tantamount to saying that the quantities being represented include *numbers* (including whole or natural numbers) *of linguistic types*.[2]

<u>Step 3: Natural numbers are represented precisely</u>

Implementing TP involves representing absolute (natural) numbers of types within a target domain and absolute (natural) numbers of types which do/don't violate a rule-like generalisation. Nevertheless, this is consistent with these representations representing numbers imprecisely. For instance, it's been argued that the ANS represents natural numbers because non-numerical confounds fail to explain its performance, and because ANS representations track quantities with a second-order character (Clarke & Beck 2021a). Even so, this is consistent with the claim that the ANS represents

[2] TP could implicate representations of rational numbers also. For instance, in a domain with 9 types, the relevant mechanisms might represent the TP threshold as 4.096. I simply take this to be an open possibility, however: In principle, a psychological system might round down to the nearest whole number (i.e., 4) and use this whole number to specify the discrete quantity of violations that can be tolerated.

natural numbers imprecisely (Clarke & Beck 2021b). Since the tasks described in Section 3.1 involve children (even young infants) keeping track of quantities outside the subitizing range, you might think performance was underwritten by an innate ANS, representing natural numbers *imprecisely*.

This would be a mistake. The tasks described in Section 3.1 involved children discriminating numbers of linguistic types with a level of precision that's unheard of in ANS tasks, despite hundreds (if not thousands) of papers probing ANS acuity. Consider Schuler et al.'s finding that, when memory retention was considered, 15/20 children performed productive generalizations in line with TP, even when frequency distributions made (token) irregular transformations more frequent. As noted, this involved children discriminating domains containing 8 types (of which 4 were rule-violating) from domains containing 9 types. Similarly, Shi and Emond found that 14-month-olds were sensitive to the difference between 10 rule-conforming and 6 rule-violating types and 11 rule-conforming and 5 rule-violating types in a domain with 16 types. Indeed, this latter result is particularly impressive since it didn't take account of individual variations in sentence retention (something Schuler et al. showed to be relevant), suggesting that – if anything – these results underestimated numerical competence.

Either way, ANS acuity lags way behind these thresholds for discrimination. Until recently, many claimed the hardest ratio an adult human's ANS could discriminate is 7:8 (Carey 2009: 295). While these claims are overblown – recent studies show that adults can distinguish 50:51 ratios above chance, given an enormous number (e.g., 400) of trials (Sanford & Halberda 2023) – children in the abovementioned studies were effectively tested on a single trial. For instance, in Shi and Emond's (2023), 14-month-olds encountered one collection of 16 Russian sentence transformations and effectively discriminated a case in which 10 conformed to a rule from a case in which 11 did. In other words, they performed better than adults have ever been observed to perform using their ANS, even though ANS acuity is considerably lower in children (Libertus & Brannon 2010).

It might be replied that this falls short of establishing that the representations implicated in TP represent *specific* natural numbers *precisely* – perhaps they still represent numbers imprecisely, just less imprecisely than the ANS. I'd counter that this reflects an unwarranted scepticism which is undermined by reflection on the abovementioned tasks. As we saw in Section 1, precise representations of specific natural numbers bear hallmarks that approximate number representations lack. For instance, they license robust judgements of equinumerosity. The abovementioned work on

TP supports thinking the representations in question meet this criterion. Experiments such as Schuler et al.'s and Shi and Emond's contrasted performance across conditions in which domain size was held constant while the number of rule-violating/conforming types varied. Using this methodology, Schuler et al.'s Experiment 1 found that productive generalisation would ensue when the quantity of rule-violating types was even marginally below the threshold set by TP (e.g., 4 exceptions in a domain with 9 types and 4.096 legal exceptions) but not if this threshold was crossed. This is consistent with thinking that a common threshold was identified across both conditions, irrespective of the number of rule-conforming/rule-violating types in the domain, and that domains were thus represented as equinumerous. Similar points apply to Shi and Emond's work with 14-month-olds on a domain size of 16 types. And this seems particularly compelling when we note that modest changes to the represented domain size affected these results dramatically, precisely as we would expect if different quantities were treated as categorically distinct. For instance, in Schuler et al.'s second experiment, children who simply remembered 8 tokens in a target domain of 9 behaved as if there was a tolerance threshold of 3.85 (in line with the predictions of TP). This highlights the fact that small differences in domain size, yielding small differences in TP threshold, impact children's behaviour markedly and categorically. It would be miraculous that these fine-grained predictions were robustly borne out, if children weren't representing numbers precisely in the first place.

Step 4: This is innate

Finally, these representations are almost certainly innate. For a start, TP seems to be a human universal. It predicts patterns of linguistic development across languages – e.g., points in the English and Mandarin count list where productive generalisation ensues, owing to the idiosyncratic quantity of rule-violating number words in either language (Yang 2016b). Similarly, corpus data suggests that it governs languages like Polish, North American English and German (Yang 2016a), extinct languages like Early Modern English (Ringe & Yang 2022), and Indigenous languages like Cree (Henke 2022). For this reason, TP is not some curiosity of anglophone children, or WEIRD communities (Henrich et al. 2010) – it emerges irrespective of culture and learning environment. And while this insensitivity to culture and learning environment would be surprising on the view that TP is learnt, it's precisely what nativism predicts. Just as an innate and biologically endowed universal grammar constrains (and ensures commonalities among) the structural features of all natural languages (Chomsky 1986), the innate implementation of TP predicts that it be a human universal.

This brings me to my second point: It's just very hard to see how TP could be learnt by young children, especially if children lack the resources required to represent precise numbers. In a standard poverty of the stimulus argument, the nativist argues that some psychological competence is innate on the grounds that there was not enough data in the environment for it to have been learnt. For instance, it is argued that certain 'deep' grammatical principles are innate because children essentially never violate these, despite having not been exposed to enough well-formed strings (or enough guidance as to what constitutes an ill-formed string) to distinguish these from scratch (Chomsky 1965). Such conclusions are, of course, disputed, owing to controversies concerning the amount of data grammatical learning requires (Perfors et al. 2011) or the amount of data children are exposed to in development (Butterfill 2020: 97-100). Regardless, the underlying reasoning seems particularly solid in the case at hand. Children do learn the superficial grammatical principles of their natural language through observation, and subsequently perform generalisations from what they have observed (e.g., overgeneralising the "add '-ed'" rule in English). This is what TP explains. These overgeneralisations might even get corrected, e.g., by a benevolent pedant who tells them not to say 'holded'. But what seems grossly non-obvious, is what would even serve as available evidence to the child that TP itself reflects a maximally efficient principle by which to store lexical entries such that these inductive generalisations would be made (or not) to begin with; not least because TP emerges in young infants, seems to be applied categorically without any discernible process of trial and error, and goes largely unnoticed by adults – i.e., it's the sort of thing whose unpacking required careful mathematical analysis by university professors. For these reasons, it seems far-fetched to suppose that TP, and the precise representations of natural number that are needed for its implementation, are learnt – a point which is, of course, consistent with the principle's genesis in the context of generative linguistics where it is expected to guide the child's search through (an innately constrained) hypothesis space (Yang 2016a).

3.3 From TP to Numeric Conception

Summarizing the discussion so far:

> Number Nativism is an unpopular view. But objections to it are unpersuasive (Section 2) and received theories of development seem to underestimate our innate numerical competences. For when we consider work on TP, we find reason to posit innate and precise representations of natural numbers (Section 3.2), which cannot be assimilated to the representations of an ANS and/or SNS.

In this final subsection, I'll tentatively argue that there is no obvious reason why our basic conceptual grasp of natural number could not be identified with these innate representations, or otherwise defined in terms of them. If true, we not only lack good reason to reject Number Nativism; we also possess evidence for the innate resources from which a plausible Number Nativism might be constructed.

To some ears, this final suggestion will sound odd. Just consider how we might square the results of Section 3.2, evincing innate natural number representations, with the results described in Section 2, concerning children's protracted failures to precisely enumerate collections. *Prima facie,* the answer could seem straightforward. The tasks described in Section 3.2 all involved children discriminating precise numbers of linguistic types. They, thus, arose in the domain of language acquisition. We might, then, suppose that the precise natural number representations involved in TP are proprietary to the language faculty. Since none of the tasks described in Section 2 involved enumerating word or sentence types, it's plausible to think they employed different cognitive resources entirely (consider Mandelbaum [2015] for relevant discussion).

So far so good. But here's the worry: If TP and its associated numerical resources are proprietary to the language faculty, how could they form the basis of the natural number concepts we come to employ in flexible thought?

In framing this worry, my interlocutors have tended to stress that the proprietary representations and processes of the language faculty are *sub-personal* (see: Drayson 2014). Hence, when one learns a grammatical rule in natural language, their language faculty extracts the natural number of types in the target domain, and divides that number by its natural logarithm, all in the service of TP (Section 3.2). But this is not something *the person does*, nor something they possess any awareness of. But if the person is oblivious to these sub-personal states and happenings, how could they provide them with a basic conception of the contents being manipulated (precise natural numbers)?

This is a common worry, but it's a red herring. Let's suppose, if only for the sake of argument, that the innate representations of precise natural number, implicated in TP, are *purely* sub-personal – states that can, in no way, be attributed to the person. This need not render them irrelevant to the acquisition of person-level natural number concepts. For the crux of the issue is not whether these innate representations are person level; it's whether these innate representations are, or could become, accessible to systems involved in concept use and concept formation, such that our mature number concepts might be defined in terms of these (see Section 1.1).

On this point, Nativists find grounds for optimism. This is because, closer inspection reveals that TP is probably not proprietary to the language faculty. Rather, TP is probably employed by a wide range of consumer systems, including systems of central cognition, associated with our conceptual grasp of the world. One reason to say this is that TP isn't really a theorem about language acquisition. It's a theorem which specifies when rules, in general, should be treated as productive – at least when access is governed by The Elsewhere Condition and frequency distributions conform to Zipf's Law. This is important, since Zipfian distributions pop-up all over the place, not just in language (Auerbach 1913), and The Elsewhere Condition specifies a general processing architecture, apt whenever exceptions must be ruled out before productive rules are applied. Accordingly, Newport notes that "The Tolerance Principle not only accounts for findings in scores of languages; it also makes new predictions about nonlinguistic concepts – that is, when a generalization will occur in inductive learning, within languages and beyond".

Of course, the extent to which TP applies outside of language is an empirical question. Nonetheless, extant results support its generality. Yang (2016b) reports that TP makes accurate, quantitative predictions about numerical understanding (e.g., children's explicit grasp of numerical succession). Moreover, it's been noted that TP can explain puzzling effects in the heuristics and biases literature.[3] For example: since TP tolerates a larger proportion of exceptions (and thus predicts *more* productive generalization) when domain size is small, it can account for classic small sample biases in decision making – e.g., why it is that humans overestimate the likelihood that a tossed coin is biased, when they observe a small sample of trials, and see it land one way >50% of the time (Tversky & Kahneman 1974). It likewise explains why biases would emerge most readily when one has limited experience with a group (Bohnet & Chilazi 2019). While these connections are currently the subject of investigation (C. Yang, pers. comm.), they motivate the possibility that TP does not simply serve to process regularities within the language faculty's proprietary domain (e.g., regularities in morphology or sentence structure). Instead, it's also used to assess productive regularities, quite generally, as when

[3] Consistent with this, it's worth noting that TP does not *simply* characterise children's learning – it is also evident in adults. For while adults typically engage in probability matching in tasks of the sort described in Section 3.1, researchers have tended to find that a subset of adults *always* behave in accord with TP during these tasks (Schuller et al. forthcoming). This suggests that TP continues to influence learning and generalization throughout the lifespan.

one is predicting whether subsequent coin tosses will land heads/tails or whether a stereotyped minority job candidate will perform well. If true, this suggests that TP and its associated numerical representations interact with a wide range of systems and representations, including those involved in domain general cognition and conceptual knowledge.

Of course, this doesn't prove that the innate representations of natural number involved in TP serve to define, or otherwise be identified with, mature concepts of number. But if it is acknowledged that TP and its associated machinery interact with a wide range of consumer systems (including conceptual systems, involved in domain general assessments of productivity – e.g., about coin tosses), and the representations and algorithms involved are used by a myriad of systems outside the language faculty, it is hard to see any obvious reason why its representations could not, *potentially*, be used by conceptual systems to define new concepts, or otherwise be accessed by these systems in the generation of number concepts, irrespective of their status as sub-personal. Which is to say: This possibility deserves further consideration. *Theories of numerical development needn't simply avail themselves of the ANS and/or SNS* (*pace* Carey & Barner 2019; Gallistel & Gelman 1976; Spelke 2017).

This is worth noting, since the proposal that the representations of number involved in TP form the basis of mature numeric conception offers to make sense of an otherwise puzzling nexus of results. We began this paper by noting that it's surprising that broad consensus would emerge on the most basic structural features of the precise natural numbers if our initial grasp of these is learnt. Indeed, this now seems *more* puzzling. After all, we've seen evidence that a basic grasp of precise natural number emerges in diverse communities which vary dramatically in their numerical vocabularies and mathematical education (Section 2.1) and noted that precise natural numbers are an abstract and visually non-salient entity, unlike many of the linguistic categories that children quickly learn to label (Section 2.2). The view that natural number concepts are identified with, or otherwise defined in terms of, computationally accessible representations of natural number which *begin their life* as *purely sub-personal symbols*, offers to shed light on this latter fact while illuminating how and why consensus emerges (somewhat uniformly) on the basic structural features of the numbers in question.

In saying this, it's important to note that there is no comparable reason to suppose that the internal operations of our language faculty would provide an innate conceptual grasp of other abstract principles of generative grammar. Unlike the numerical representations involved in TP, there's no reason to suppose that these principles are accessible to a wide range of (conceptual) subsystems, and there is little plausibility to the idea that a conceptual grasp of these principles emerges independently of culture and learning environment (compare Section 2.1). Similar points apply to the proprietary states of other domain-specific faculties and algorithms, like those comprising the visual system.

We also shouldn't assume that the above considerations overgeneralize, showing that since TP is made use of by myriad psychological faculties it could ground conceptual knowledge of our DIVIDING QUANTITIES BY THEIR NATURAL LOGARITHM (i.e., of the TP theorem itself). For while TP implicates *representations* of natural number (Section 3.2), many faculty psychologists would urge that the inflexible transformations such representations drive tend to be built into the architecture of the faculties themselves (Pylyshyn 2003; Quilty-Dunn & Mandelbaum 2018). If true of TP, the transformation algorithm in question will not be explicitly represented and will thus be unable to define concepts of natural logarithms (or the like), even in principle.

Finally, we should acknowledge that, in allowing TP's numerical representations to be made use of by myriad bits of the mind (including processes involved in conceptual thought) we do not incur an obvious tension with children's protracted failures on the task(s) described in Section 2.3. Nativists have always posited discontinuities in the accessibility of our concepts throughout development. This is why nativists about BELIEF (e.g., Onishi & Baillargeon 2005) find it no embarrassment that children suffer a slow and protracted development on certain versions of the false belief task (Wimmer & Perner 1983). It is also why Plato found it no worry that the slave, discussed in the *Meno*, was seen to possess an existing grasp of the Pythagorean theorem even though he was unable to access this information in context prior to Socrates' questioning. To assume that this is implausible in the case of number, without argument, is to beg the question against thousands of years of nativist thinking.

4. Conclusion

I have argued for a modest conclusion, but one with the potential to radically reorient thinking on numerical development. Number Nativism is often considered hopelessly out of touch with the empirical record. However, I hope to have made room for its consideration. Objections to Number Nativism are unpersuasive, the innate numerical competences of young children are richer than current theories recognize, and these competences implicate precise representations of natural number which could (potentially) serve as the basis of mature numeric conception. Add to this the fact that the seeming ubiquity of precise enumeration across cultures is straightforwardly explained by the Nativist (Butterworth 1999), coupled with the fact that Nativism explains why broad consensus emerges on basic structural properties of the natural numbers (e.g., that each natural number is exactly one greater than its predecessor – something which learning accounts have been criticised for not even *trying* to explain [Rey 2014]), it's hard to shake the feeling that Number Nativism remains a leading theory of numerical development.

Works cited:

- Anderson, S. R. (1969). *West Scandinavian vowel systems and the ordering of phonological rules*. PhD thesis, MIT.
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press, Cambridge, MA.
- Auerbach, F. (1913). *Das Gesetz der Bevölkerungs konzentration*. Breslau: W.
- Ball, B. (2017). On representational content and format in core numerical cognition. *Philosophical Psychology*. 30(1–2), 119-39.
- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*. 86(3), 201–21.
- Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*. 70(19), 661–79.
- Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.
- Bohnet, I. & Chilazi, B. (2019). Overcoming the Small-N Problem. *What Works: Evidence-Based Ideas to Increase Diversity, Equity, and Inclusion in the Workplace*. UMass Amherst Guide.
- Bornstein, M. H. (1976). Infants' recognition memory for hue. *Developmental Psychology, 12*(3), 185–191.
- Brown, D. & Hippisley, A. (2012). *Network morphology: A defaults-based theory of word structure*. Vol 133. Cambridge University Press.
- Butterfill, SA. (2018). Tracking and representing others' mental states. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 269–279). Routledge/Taylor & Francis Group.
- Butterfill, SA. (2020). *The developing mind: A philosophical introduction.* Routledge Press.
- Butterworth, B. (1999). *The mathematical brain.* Macmillan Press.
- Butterworth, B., Reeve, R. et al. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *PNAS,* 105(35), 13179-84.
- Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425– 455.
- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Carey, S., & Barner, D. (2019). Ontogenetic Origins of Human Integer Representations. *Trends in Cognitive Sciences*. 23(10):823-35.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press, Cambridge, MA.
- Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.
- Chomsky, N. (1988). *Language and problems of knowledge: The Managua Lectures*. The MIT Press.
- Clarke, S. & Beck, J. (2021a). The number sense represents (rational) numbers. *Behavioral and Brain Sciences,* 44, E178.
- Clarke, S. & Beck, J. (2021b). Numbers, numerosities, and new directions. *Behavioral and Brain Sciences,* 44, E205.
- Clahsen, H. (1999). Lexical entries and rules of languge: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22:991–1069.
- Cowan R., & Frith C. (2009). Do calendrical savants use calculation to answer date questions? A functional magnetic resonance imaging study. *Phil. Trans. R. Soc. B*. 364:1417–1424.
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics: How the Mind Creates Mathematics*. Oxford University Press.
- Dehaene S, Changeux JP. (1993). Development of elementary numerical abilities: a neuronal model. *J Cogn Neurosci*. 5(4):390-407.
- Dixon, R. (1980). *The languages of Australia.* Cambridge Core: Cambridge University Press.

- Drayson, Z. (2014). The personal/sub-personal distinction. *Philosophy Compass.* 9(5): 338-46.
- Shi, R. & Emond, E. (2023). The threshold of rule productivity in infants. *Frontiers in Psychology.* 14.
- Fagan, S. & Gençay, R. (2010). An introduction to textual econometrics. In A. Ullah & D. Giles (eds.) *Handbook of Empirical Economics and Finance,* Routledge.
- Feigenson, L., Carey, S. & Hauser, M. (2002). The representations underlying infants' choice of more. *Psychological Science.* 13(2):150-56.
- Fodor, J. (1975). *The language of thought.* MIT Press.
- Fodor, J. (1981). The present status of the innateness controversy. In his *RePresentations: Philosophical Essays on the Foundations of Cognitive Science.* MIT Press.
- Fodor, J. (2008). *LOT 2.* MIT Press.
- Fodor, J. (2010). Woof, woof. *Times Literary Supplement*, October 8: 7–8
- Frank MC, Everett DL, Fedorenko E, Gibson E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition.* 108(3):819-24.
- Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*, Breslau: W.
- Fuson, KC., Richards, J., & Briars, DJ. (1982). The acquisition and elaboration of the number word sequence. In *Childrens logical and mathematical cognition,* pp.33-92. Springer.
- Goldberg, A. (2019). *Explain me this: creativity, competition and the partial productivity of constructions.* Princeton: Princeton University Press.
- Gómez, RL., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science,* 7:567-80.
- Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, 306(5695), 496–499.
- Halberda, J. (2016). Epistemic Limitations and Precise Estimates in Analog Magnitude Representation. In *Core knowledge and conceptual change* (pp. 171–190).
- Halle, M. (1997). Distributed morphology: Impoverishment and fission. *MIT working papers in linguistics.* 30:425-49.
- Hansen, M.B.; Markman, E.M. (2009). Children's use of mutual exclusivity to learn labels for parts of objects. *Developmental Psychology.* 45 (2): 592–596.
- Henke, R. (2022). Rules and exceptions: a tolerance principle account of the possessive suffix in north east Cree. *J. Child Lang.* 50, 1119-54.
- Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33:1-75.
- Hurford, JR. (1987). *Language and Number: The Emergence of a Cognitive System.* Blackwell.
- Izard, V., Pica, P., Spelke, E. & Dehaene, S. (2008) Exact Equality and Successor Function: Two Key Concepts on the Path towards Understanding Exact Numbers, *Philosophical Psychology*, 21:4, 491-505.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *PNAS*, 106(25), 10382– 10385.
- Izard, V., Steri, A., & Spelke, E. (2014). Toward exact number: Young children use one-to-one correspondence to measure identity but not numerical equality. *Cognitive Psychology.* 72:27-53.
- Koulaguina, E., & Shi, R.. (2013). Abstract rule learning in 11-and 14-month-old infants. *Journal of Psycholinguistic Research*, 42(1), 71-80.
- Koulaguina, E., & Shi, R. (2019). Rule generalization from inconsistent input in early infancy. *Language Acquisition*, 26(4), 416-435.
- Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.* Harvard University Press.
- Le Corre M, Carey S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition.* 105(2):395-438.
- Le Corre, M., & Carey, S. (2008). Why the verbal counting principles are constructed out of representations of small sets of individuals: A reply to Gallistel. *Cognition, 107*(2), 650–662.
- Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science, 34*(1), 51–67.

- Leslie, A. M., Gelman, R., & Gallistel, C. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, 12, 213–218.
- Libertus ME, & Brannon EM. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science.* 13(6):900-906.
- Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1–22.
- Marcus, GF. (1995). The acquisition of the English past tense in children and multi-layered connectionist networks. *Cognition*. 56(3):271-9.
- Margolis, E. (2020). The Small Number System. *Philosophy of Science*, 87(1), 113–34.
- Margolis, E., & Laurence, S. (2011). Learning Matters: The Role of Learning in Concept Acquisition. *Mind & Language*, 26(5), 507–39.
- Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman, J. P. Byrnes, S. A. Gelman, J. P. Byrnes (Eds.), *Perspectives on Language and Thought: Interrelations in Development* (pp. 72-106). New York, NY US: Cambridge University Press.
- Marslen-Wilson, W. D., & Tyler, L. K. (1997). Dissociating types of mental computation. *Nature, 387*(6633), 592–594.
- Medina TN, Snedeker J, Trueswell JC, Gleitman LR. (2011). How words can and cannot be learned by observation. *PNAS*. 108(22):9014-9.
- Nevins, A., Pesetsky, D. & Rodrigues, C. (2008). Piraha˜ exceptionality: A reassessment. *LingBuzz*. Tromso: Universitetet i Tromsoe.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258.
- Peano, G. (1908). *Formulario Mathematico.* Turin: Bocca Freres.
- Perfors, A., Tenenbaum, JB., Griffiths, TL., & Xu F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition.* 120(3):302-21.
- Pica, P., Lemer, C., Izard, V., & Dehaene. S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, 306(5695), 499–503.
- Pitt, B., Gibson, E., & Piantadosi, S. (2022). Exact number concepts are limited to the verbal count range. *Psychological Science,* 33(3): 343-77.
- Pinker, S. (1994). *The Language Instinct.* Willian Morrow & Co.
- Pinker, S. (1999). *Words and rules: The ingredients of language.* Basic Books, NYC.
- Pinker, S. & Ullman, MT. (2002). The past and future of the past tense. *Trends in Cognitive Sciences,* 6(11):456-63.
- Plunkett, K. & Marchman, V. (1993). U-Shaped learning and frequency effects in a multi-layered perception. *Cognition.* 38(1): 43-102.
- Pylyshyn, Z. (2003). *Seeing and Visualising: It's not what you think*. MIT Press.
- Quilty-Dunn, J. & Mandelbaum, E. (2018). Inferential Transitions. *Australasian Journal of Philosophy*, 96:3, 532-547.
- Quine, W. (1960). *Word and Object.* MIT Press: Cambridge.
- Reeve, R., Reynolds, F., Paul, J., & Butterworth, B. (2018). Culture-Independent Prerequisites for Early Arithmetic. *Psychological Science,* 29(9): 1-10.
- Rey, G. (2014). Innate *and* learned: Carey, mad dog nativism, and the poverty of stimuli and analogies (yet again). *Mind & Language,* 29:109-32.
- Rice, N. (1980). *Cognition to language.* Baltimore, MD: University Park Press.
- Ringe, D. & Yang, C. (2022). The threshold of productivity and the 'irregularization' of verbs in Early Modern English. In L. Betterlou et al. (Eds.) *English Historical Linguistics: Change in structure and meaning*. John Benjamins Publishing.
- Rivest, R. (1976). On self-organizing sequential search heuristics. *Communications of the ACM,* 2:63-7.
- Rummelhart, D. & McClelland, JL. (1986). On learning the past tenses of English verbs. In McClelland, JL. & Rummelhart, D. *Parallel Distributed Processing: Explorations and the microsctructure of cognition,* Vol 2. 216-71. MIT Press: Cambridge.
- Samuels, R. & Snyder, E. (2023). *Number Concepts: An Interdisciplinary Inquiry.* Cambridge University Press.
- Sanford, E. and Halberda, J. (2023). Successful discrimination of tiny numerical differences. *Journal of Numerical Cognition*.

- Sapir, E. (1928). *Language: An introduction to the study of speech.* Harcourt Brace, NY.
- Scholl, B. J., & Leslie, A. M. 1999. "Explaining the Infant's Object Concept" In *What Is Cognitive Science?,* ed. Ernest Lepore and Zenon Pylyshyn, 26-73. Oxford: Blackwell.
- Schuler, K.D., Yang, C., & Newport, E.L. (forthcoming) Children form productive rules when it is more computationally efficient. *Journal of Memory and Language.*
- Shea, N. (2011). New concepts can be learned. *Biology & Philosophy,* 26: 129-39.
- Simon, T.J. (1997). Reconceptualizing The Origins of Number Knowledge: A "Non-Numerical" Account. *Cognitive Development*, 12.3: 349-72.
- Sleator, D. & Tarjan, R. (1985). Amortized efficiency of list update and paging rules. *Communications of the ACM.* 28(2):202-38.
- Snyder, E. (2017). Numbers and Cardinalities: What's Really Wrong with the Easy Argument for Numbers? *Linguistics and Philosophy.* 40: 373-400.
- Spelke, E. (2017). Core Knowledge, Language, and Number. *Language Learning and Development*, 13:2, 147-170,
- Spelke, E., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition,*78, 45-88.
- Spiegel C, Halberda J. (2011) Rapid fast-mapping abilities in 2-year-olds. *J Exp Child Psychol.* 109(1):132-40.
- Sullivan, J., & Barner, D. (2013). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology*, *66*(2), 389-402.
- Trick, LM. & Pylyshyn, ZW. (1994). Why are small and large numbers enumerated differently? *Psychological Review.* 101(1):80-102.
- Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science,* 185(4157): 1124-31.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs. *Cognition*, 13(1): 103-28.
- Wynn K. (1990). Children's understanding of counting. *Cognition.* 36(2):155-93.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750.
- Yang, C. (2002). Knowledge and learning in natural language. Oxford University Press, Oxford.
- Yang, C. (2016a). The price of linguistic productivity. Cambridge, MA: The MIT Press.
- Yang, C. (2016b). The linguistic basis of next number. *LingBuzz.* Tromso: Universitetet i Tromsoe.
- Yang, C. (2018). A user's guide to the tolerance principle. *LingBuzz.* Tromso: Universitetet i Tromsoe.
- Yousif, S.R., & Keil, F.C. (2020). Area, not number, dominates estimates of visual quantities. Scientific Reports, 10(1), 13407.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley, Cambridge, MA.