



City Research Online

City St George's, University of London

Citation: Endress, A. (2019). Duplications and domain-generalty. *Psychological Bulletin*, 145(12), pp. 1154-1175. doi: 10.1037/bul0000213

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/22855/>

Link to published version: <https://doi.org/10.1037/bul0000213>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Duplications and domain-general

Ansgar D. Endress
City, University of London

Running head: DUPLICATIONS DOMAIN-SPECIFICITY

Abstract

While specialized, adaptive behavioral traits are ubiquitous in the animal kingdom, at least in humans, there are considerable debates on whether the mind is primarily characterized by various special-purpose, domain-specific mechanisms, or by a few general-purpose, domain-general mechanisms. Drawing from research on artificial language learning, associative learning, serial learning, executive control and formal linguistics, I argue that neither domain-specificity nor domain-generality provide satisfactory descriptions when considering how cognitive mechanisms are implemented. I suggest that some cognitive mechanisms are “*domain-bound*” – they are available in multiple domains (and thus not domain-specific), but not in other domains (and thus not domain-general). Hence, these computations can be performed in many domains but not in others, can be recruited simultaneously by multiple domains, and, across domains, individual abilities with a given computation are relatively uncorrelated. Domain-bound mechanisms have a straightforward evolutionary interpretation: Analogously to the evolution of molecular and morphological structures, cognitive mechanisms can become duplicated over evolution, with independent copies in different domains. This and previous evidence for the importance of duplications for our cognitive abilities call for a revision of the concept domain-generality, suggesting that, in many cases, mechanisms traditionally seen as domain-general might really reflect a collection of local copies of specialized mechanisms.

Keywords: Domain-Generality; Domain-Specificity; Domain-Boundedness; Artificial Grammar Learning; Duplication; Serial Learning; Statistical Learning; Executive Function; Perceptual or Memory Primitives

Duplications and domain specificity

Karl Marx and Adam Smith might differ in their political philosophies, but they agree on one thing: division of labor leads to greater efficiency. Does such a division of labor also exist in the brain? Are there “domain-specific” computational mechanisms that are constrained to operate in only one domain (e.g., face processing) and others in other domains (e.g., language)? Are there “domain-general” mechanisms that can operate in all domains alike? For example, are there domain-specific mechanisms involved in language acquisition (Chomsky, 1980; Fodor, 1983; Lenneberg, 1967), or does language acquisition rely on computational mechanisms available in essentially any domain (Aslin & Newport, 2012; McClelland, Rumelhart, & Group, 1986; Seidenberg, 1997)?

Here, I propose a new view on such questions, whereby cognitive mechanisms can become duplicated over the course of evolution (see also Barrett, 2012; Chakraborty & Jarvis, 2015; Endress, Nespore, & Mehler, 2009; Kaas, 1989; Marcus, 2004; Markram et al., 2004; Samuels, 2000; Shukla, 2005; Sperber, 1994, for related proposals). This allows independent copies of a mechanism to be available in multiple domains, without being fully domain-general. I will first provide several case studies that seem to require duplication accounts of some sort. For example, humans and other animals can detect identity-relations (e.g., noticing that the last syllable is repeated in a sequence such as *dubaba*). While identity-relations seem to be detected by a specialized mechanism, this mechanism is available in many domains and species, from language to vision and olfaction, and from humans to ducklings to bees, and thus cannot be domain- or human-specific (see below for a more detailed review and specific references).

However, even in the domain of language, identity-relations can be computed for only some classes of items but not for others, an unexpected result if the identity-detector were truly domain-general. However, if cognitive mechanisms can become duplicated over evolution, the mechanism detecting identity-relations might be what Endress, Nespore, et al. (2009) called “domain-bound.” There might be distinct and independent instances of the mechanism detecting identity-relations for language, vision, olfaction and so on. (Hereafter, I will use the term identity-detector to refer to an *instance* of the mechanism detecting identity-relations.) Further, I will argue that each of these domains may have several independent identity-detectors and that identity-detection might possibly occur through fairly simple circuit motifs. If so, these identity-detectors would allow identity-relations to be detected in various domains; each of the individual identity-detectors, in contrast, is domain-specific.¹

Following the discussion of identity-relations, I will turn to other case studies from statistical learning to executive function, and suggest that the situation is similar in these cases: mechanisms that are usually considered domain-general might reflect a set of independent, local mechanisms.

1. What is a domain?

The definition of a domain is critical for the following discussion, and helps illustrating some confusion in the literature. From an informational point of view, a domain is a distinct class of information. This class of information is

¹ The term *domain-bound mechanism* is somewhat of an oxymoron. These mechanisms are domain-specific instantiations of mechanisms that perform the same kind of computation. For consistency with earlier papers, however, I will use the term “domain-bound” mechanisms.

recurrent and ecologically relevant in an organism's mental life. As a result, the organism has evolved mechanisms that process this kind of information (see Hirschfeld & Gelman, 1994, for a similar definition), though these mechanisms might come to process different kinds of information than what they evolved to process (Dehaene & Cohen, 2007; Gould & Vrba, 1982; Sperber, 1994). For example, given that faces are a critical source of information in social species, they constitute a recurrent class of information, which might have led to the evolution of mechanisms that deal with this class of information (Bonatti, Frot, Zangl, & Mehler, 2002; Kanwisher, McDermott, & Chun, 1997; Reid et al., 2017; Sugita, 2008; van den Hurk, Van Baelen, & Op de Beeck, 2017, but see e.g., Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Under this view, mechanisms that are domain-general can apply to all domains alike, and mechanisms that are domain-specific process information only in a single domain.

However, there are a number of problems with this definition. First, it leads to a confusion between the computations a mechanism implements, and the implementations themselves (Marr & Nishihara, 1992). For example, some computations we perform on faces might be identical to those we perform on xylophones (e.g., we might encode the relative spatial positions of object parts for either class of stimuli). These computations might well be implemented in distinct brain areas, and might thus be independent. Given the information-based definition above, these computations would thus count as domain-general although the instantiations of these computations are domain-specific, and reside as independent instantiations in whatever brain regions process faces and

xylophones (see also Nozari & Novick, 2017, for further arguments based on monitoring in speech perception and production that we need to distinguish notions of domain-generality based on (i) the similarity of the information that is processed and (ii) the brain regions that are activated by the corresponding processes). As a result, domains cannot be defined exclusively based on the type of information that is processed. A viable definition of a domain needs to take into consideration *how* the information is processed in terms of the underlying implementations.

Second, the definition of a domain above is not fine-grained enough to account for the hierarchical structure of processing. For example, language is clearly a domain, but so is phonology, and, within phonology, prosody. Given that, when we process prosodic information, it comes packaged with phonological as well as linguistic information, a pure information-based definition of domains cannot account for this hierarchical domain structure. To be sure, this is not an inherent problem of an information-based definition of a domain. However, given that the kinds of domains that populate discussions of domain-specificity tend to be fairly broad (e.g., language, face processing, Theory of Mind,...), the kinds of domains that are usually discussed do not reflect the hierarchical structure of processing.

To avoid such problems, I opt for a definition that integrates the information-based notion of domain with the mechanism-based notion of *module*, that is, the set of mechanisms that process the information in a domain (see Samuels (2000) for a similar distinction between what he calls “computational modules” and “Chomskian modules”). Specifically, in the discussion below, a

“domain” refers to a set of implementations of computational mechanisms that process a certain, ecologically recurring class of information, and that, together, fulfill an ultimate function in our interactions with the environment. Examples might include face recognition, object recognition, music or language. A domain is a set of implementations of processing mechanisms that *specifically* operate on the kind of information from which the domain takes its name.

For example, if face recognition is a domain, its constituent mechanisms *specifically* process face-related information. Even if we process relative spatial positions of object parts for both faces and xylophones, the xylophone-related instantiation of this mechanism would not be part of the face-domain. Further, some processes operate on different kinds of information; for example, face and xylophone processing both rely on early visual information processing. These shared processes are not part of the domain of face recognition, because they are not *specifically* involved in face processing.

This notion is similar to Fodorian modules (Fodor, 1983; see also Hirschfeld & Gelman, 1994) except that it is defined in terms of functions and processes rather than in terms of content, and that it is agnostic with respect to the degree of automaticity of a given process (see Sperber, 1994, for a similar view).

Importantly, this definition is hierarchical. Just as every goal can (but need not) have sub-goals, any (ultimate) function can (but need not) have sub-functions. Domains thus can have sub-domains as well. For example, patients with brain damage can have dissociable deficits within the language domain (e.g., problems of language perception vs. production, of grammar vs. word-finding etc.). These functions thus define at least partially independent sub-domains, and

presumably have sub-domains on their own. At least at the genetic level, such a hierarchical modular organization might be crucial to the evolvability of organisms (Nothwang, 2016). Further, if the output of modules generally serves as the input to other modules, such hierarchical processing also increases the flexibility of cognitive operations (Sperber, 1994).²

2. Duplication in evolution

2.1. The importance of duplication in evolution

The proposal below relies on the duplication of cognitive mechanisms. Before presenting my (cognitive) case studies, I will thus briefly outline the role of duplications in evolution. The importance of the duplication of genes, whole genomes and morphological structures has long been recognized in evolutionary biology (Allman & Kaas, 1971; Chakraborty & Jarvis, 2015; Hurley, Hale, & Prince, 2005; Innan & Kondrashov, 2010; Kaas, 1989; Magadum, Banerjee, Murugan, Gangapur, & Ravikesavan, 2013; Ohno, 1970; Taylor & Raes, 2004; J. Zhang, 2003). Innan & Kondrashov (2010) argued that “hardly any aspect of genome evolution or function is not somehow linked to gene duplications, which occur in all kinds of life forms and have taken place since before the last universal common ancestor.” For example, in catarrhines (old-world monkeys and apes),

² To illustrate the increased flexibility due to several levels of representation, Sperber (1994) uses the example of the representations we might maintain of animals. “What you are told about cats is integrated with what you see of cats, in virtue of the fact that the representation communicated contains the concept CAT. But now you have the information in two modes: as a representation of cats, handled by a first-order conceptual module, and *as a representation of a representation* of cats, handled by the second- order metarepresentational module. That module knows nothing about cats but it may know something about semantic relationships among representations” (p. 61, emphasis added).

trichromatic color vision evolved from dichromatic color vision by duplication of an opsin (Jacobs, 2009; Surridge, Osorio, & Mundy, 2003), and mammalian sound localization circuits for high frequencies might be duplicates of earlier circuits (Nothwang, 2016). In fact, just since humans split from chimpanzees, up to 1,800 gene duplications took place in the human lineage (Taylor & Raes, 2004; J. Zhang, 2003).

Such results raise the possibility that duplications are important for the evolution of cognitive abilities as well, and in some form or another, different authors have made similar proposals (Barrett, 2012; Chakraborty & Jarvis, 2015; Endress, Nespor, et al., 2009; Kaas, 1989; Marcus, 2004; Shukla, 2005; Sperber, 1994). For example, Pinker (1998) argued that

“evolutionary change often works by copying body parts and tinkering with the copy. [...] A similar process may have given us our language of thought. Suppose ancestral circuits for reasoning about space and force were copied, the copy’s connections to the eyes and muscles were severed, and references to the physical world were bleached out. The circuits could serve as a scaffolding whose slots are filled with symbols for more abstract concerns like states, possessions, ideas, and desires. The circuits would retain their computational abilities [...]. When the new, abstract domain has a logical structure that mirrors objects in motion [...], the old circuits can do useful inferential work” (p. 355/356).

In line with this speculation, more recent research has shown that entire brain pathways might indeed become duplicated (Chakraborty & Jarvis, 2015; Nothwang, 2016). For example, in vocal-learning bird species, the brain regions

involved in song learning might be duplicates of earlier motor regions, and at least some parrot species and humans might have yet another duplicate (Chakraborty & Jarvis, 2015; Chakraborty et al., 2015). Closer to home, Belyk and Brown (2017) proposed that duplicated brain areas might be important for the evolution of human speech control as well. They noted that humans have two rather than one laryngeal motor areas in each hemisphere and proposed that the more dorsal area evolved either as a duplicate of the (ancestral) ventral laryngeal motor area (and a subsequent migration of the duplicate to its current location) or as a duplicate of a non-vocal motor area (and a subsequent change in the connectivity pattern of the duplicate).

There are also computational reasons for believing that duplicated neural circuits might be used across domains. In machine learning, it has been shown that classification performance in an artificial neural network improves by using connectivity patterns from a network trained on a classification problem in a *different* domain (Caruana, 1997; Donahue et al., 2014). As neural circuits are tuned by evolution, duplicated neural circuit might also be computationally effective in a different domain, though they likely require fine-tuning.

2.2. Functional integration of duplicated structures

Duplicated structures are useful only in so far as they are functionally embedded in the rest of the organism. However, this is often the case (see e.g., Chakraborty & Jarvis, 2015; Hurley et al., 2005, for reviews). In fact, cell types and connectivity are often conserved even if they end up in fairly different structures in different animals (Briscoe & Ragsdale, 2018). For example, birds and other reptiles do not share the mammalian six-layered neocortex. Still, they

share basic circuit types, with some neurons receiving sensory input, others providing output to the brain stem, and still others act as relay between input and output neurons. These basic circuits are conserved due to a conserved network of transcription factors, but can end up in structures that differ vastly across taxa such as the mammalian neocortex and the avian dorsal ventricular ridge. They can also have strikingly different morphology: in the mammalian neocortex, excitatory cells tend to be pyramidal while they have stellate shape in the pallium in birds. Briscoe and Ragsdale (2018) sum up this situation as follows:

“The core circuits of the pallium [in mammals, birds and other reptiles] are conserved, but pallial architectures and the spatiotemporal regulation of cell type specification are not. These latter levels of biological organization can shift over evolutionary time, so long as sensory information gets to the pallium, gets integrated, and elicits an appropriate output. Some transcription factor molecules, such as the hypothesized input cell determinants, exhibit a conserved correlation with defined neuronal cell types across species.”

A more concrete and well-understood case of how duplicated structures can be functionally integrated is the evolution of trichromacy in old-world and Howler monkeys (see e.g., Jacobs, 2009; Surridge et al., 2003, for reviews). As mentioned above, trichromacy evolved from dichromacy through duplication of an opsin. Both the ancestral copy and the duplicate are functional due to a fairly simple mechanism. Both copies of the gene are “nearby,” and their expression is controlled by a common control region. In each cone, this control region thus

stochastically enables expression of only one of the opsins.³ As a result, some cones express the ancestral copy, and others the duplicate.

Another route that leads to (presumably) functional duplicates is the duplication of entire brain pathways. In fact, if entire brain circuits are duplicated by duplicating transcription factors that control both the number and the placement of these circuits within the brain (Chakraborty & Jarvis, 2015; Chakraborty et al., 2015), the resulting duplicates would be expected to be functionally integrated.⁴

2.3. Duplication versus exaptation

It should be noted that the idea of duplication is related to but different from the concept of exaptation (Gould & Vrba, 1982). An exapted trait is co-opted for a new function that is different from the one (if any) for which it has been selected. A classic example involves the evolution of bird feathers. Early birds had probably only limited flight abilities. However, even non-avian dinosaurs were covered in feathers, suggesting that feathers evolved for other

³ In Howler monkeys, the control region was duplicated as well.

⁴ Functionally duplicated structures do not even always require duplicated genes. Trichromacy provides an example for this possibility as well. In primates, the gene for one opsin resides on an autosomal (i.e., non-sex) chromosome, while the gene for the opsin that is duplicated in old-world monkeys resides on the X chromosome. Critically, this latter opsin is polymorphic. As a result, in females whose X chromosomes differ (i.e., who are heterozygous), cones can express the opsin from the non-sex chromosome, or either of the two alleles from the X chromosome. Heterozygous females can thus be trichromates in the absence of a duplicated opsin. Likewise, in zebra fish, mutations (rather than duplications) of a regulatory gene can result in extra copies of functional neurons that drive a behavioral escape response (Liu, Gray, Otto, Fetho, & Beattie, 2003), and even injection of mRNA (the type of RNA based on which ribosomes synthesize proteins) from regulatory genes into embryos is sufficient for the development of functional extra copies of neurons mediating this escape response (Hale, Kheirbek, Schriefer, & Prince, 2004). Hence, duplicated structures are often functional, even when they do not result from duplication at the genetic level.

purposes such as thermal insulation or display (Brusatte, O'Connor, & Jarvis, 2015). As a result, by the time birds evolved from dinosaurs, early birds were likely feathered. Feathers thus were *not* selected for flight in birds, and are rather an *exaptation* for flight (Brusatte et al., 2015; Gould & Vrba, 1982).

Exaptation has also been proposed for cognitive abilities, such as the recycling of originally visual brain areas for reading (Dehaene & Cohen, 2007), of mechanism tracking the number of items for symbolic mathematics (Dehaene & Cohen, 2007; Halberda, Mazocco, & Feigenson, 2008), or of serial memory mechanisms for language acquisition and use (Endress, Nespore, et al., 2009; Fischer-Baum, Charny, & McCloskey, 2011). For example, the “Visual Word Form Area,” a region in the left lateral occipitotemporal sulcus, is specifically activated when reading visual words (Dehaene & Cohen, 2007); this region might have been exapted for reading in humans either because, in non-human primates, it has a pattern of connectivity that makes it particularly conducive for processing visual symbols or because it might be involved in the analysis of visual shapes (Hannagan, Amedi, Cohen, Dehaene-Lambertz, & Dehaene, 2015).

In contrast to exapted traits, duplicated traits first become duplicated. Then, they might retain their original function (or a sub-function thereof), or acquire novel functions either through adaptation or exaptation (though, in the case of gene duplications, it is debated to what extent both copies of the duplicated gene retain a function; M. Lynch & Force, 2000).

2.4. The conglomerate analogy of duplicated structures

Figure 1 presents an analogy to illustrate how these evolutionary concepts help clarifying cognitive concepts such as domain-specificity. In this analogy, the

brain is like a company that manufactures different products in different factories (e.g., motorcycles in a motorcycle factory, and pianos in a piano factory); these factories correspond to domains or modules. However, each factory requires some common functions such as cleaners. If the cleaners are “domain-specific,” they exist only in one of the factories, with no cleaners in the other factories. If the cleaners are “domain-general,” they are located *outside* of the factories (e.g., at the level of the parent company or with a contractor), and enter the factories only to perform their services. Lastly, if the cleaners are domain-bound, each factory has its own cleaners. In contrast, an exaptation would be the equivalent of the factory workers using, say, a shovel to clean the factories; the shovel-wielding workers might be domain-general, domain-specific or domain-bound.

In addition to existing only in a single factory, there is a second way in which cleaners can be domain-specific. All factories might use independent cleaners, but the cleaners might operate differently in different factories. For example, they might use mops in some factories and vacuum cleaners in others. In other words, while their *function* is identical across factories, the *implementation* of this function differs. In evolutionary terms, such a situation might arise when the mechanisms evolved independently in different domains (e.g., through a form of within-species convergent evolution), and, across species, similar functions are sometimes implemented using different computational principles (e.g., auditory source localization in mammals vs. barn owls; Grothe, 2018).⁵

⁵ It is conceivable that some factories share the same cleaners, while other factories have a different set of cleaners (or no cleaners altogether). Psychologically speaking, this might occur for at least five reasons. First, the cleaners might not be a proper part of the factory (or domain

Below, I will provide evidence from behavioral studies that, I argue, are best explained if cognitive mechanisms can become duplicated. I will mostly focus on descriptions at the psychological level, simply because we don't know how the relevant mechanisms are implemented at the neuronal level, let alone how they evolved. However, I will also review some biologically realistic models that might account for some of the case studies below, and argue that these duplicated mechanisms might reflect potentially simple circuits that might well be found in many brain regions.

according to the definition above), but rather be located at the level of the parent company; critically, however, the parent company would send the cleaners only into a sub-set of its factories. Psychologically speaking, this might correspond to a domain-general set of processes with top-down connections to only certain domains, either because some domains are not "connected" to these domain-general processes to begin with, or because the connections were lost due to pruning (see Section 7 below). Second, and relatedly, some lower-level processes might have bottom-up connections to some domains, but not others. For example, even if they represent different domains, face processing and visual object processing both require visual input; in contrast, visual bottom-up input is less important for, say, music processing.

In contrast to the first two possibilities, only a subset of the factories might have cleaners even when the cleaners are part of the corresponding domain. This might arise either because, initially, all factories have domain-bound cleaners but then some factories lose them, or because, initially, no factory has cleaners, but some factories *develop* cleaners from the personnel they already have. Both possibilities are psychologically attested. First, there is evidence, both at the synaptic (Huttenlocher & Dabholkar, 1997) and at the behavioral level (Hannon & Trehub, 2005; Pascalis, de Haan, & Nelson, 2002; Polka & Werker, 1994; Werker & Tees, 1984) that certain abilities that are present early in life are lost in the absence of relevant experience (see Section 7). Alternatively, in its initial stage, none of the factories might have cleaners, but all have employees who have the *potential* and tools to be cleaners; in some factories, these employees might thus become specialized for cleaning, similar to how certain brain regions become specialized for reading when provided with relevant experience such as schooling (Dehaene et al., 2010; Saygin et al., 2016 and Section 7).

Finally, it is also possible that cleaners arise *de novo* only in certain domains.

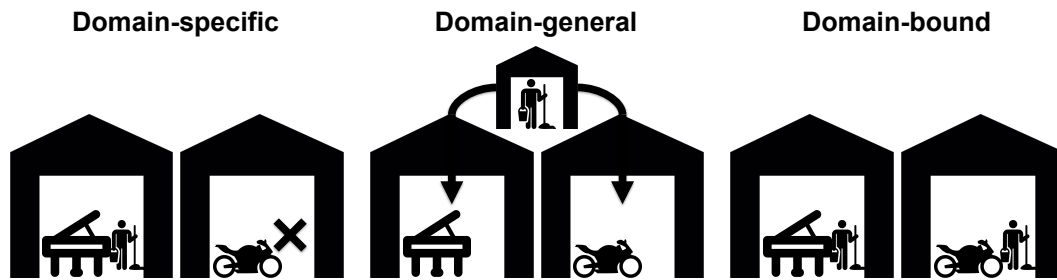


Figure 1: Illustration of the concepts domain-specificity, domain-generality and domain-boundedness, using the analogy of a company with factories that manufacture different products (e.g., a motorcycle factory and a piano factory). All factories require cleaners. (Left) “Domain-specific” cleaners exist only in one of the factories, but not in the other one. (Middle) “Domain-general” cleaners are located *outside* of the factories (e.g., at the level of the parent company or with a contractor), and enter the factories only to clean them. (Right) If the cleaners are domain-bound, each factory has its own cleaners. Using a shovel to clean the factories would be the equivalent of an exaptation.

3. The case of identity-relations

So far, I argued that cognitive mechanisms might (i) be specific to a domain (and thus domain-specific), (ii) exist as a single instantiation that is common to all domains (and thus be domain-general), or (iii) have independent instantiations in different domains (or thus be domain-bound). Further, I argued that the latter possibility would be consistent with the duplication of cognitive mechanisms. I will now review the first cognitive case study for this possibility: the detection of *identity-relations*. In syllable sequences such as *dubaba*, even

seven-month-olds notice that the last two syllables are identical, and extend this identity-relation to new items. Specifically, Marcus, Vijayan, Rao, and Vishton (1999) familiarized seven-months-old infants with syllable sequences such as *ledidi*, *wijeje* and so on (where the last two syllables were repeated, hereafter called an ABB pattern). Following this a familiarization, infants were more familiar with novel sequences with novel syllables that shared the identity-pattern (e.g., *bapopo*) compared to novel syllable sequences that had a different pattern (e.g., *babapo*; hereafter called an AAB pattern). They similarly learned AAB patterns. Infants thus learned an open-ended regularity based on identity-patterns.

3.1. How are identity-relations computed?

According to Marcus et al. (1999), infants learned abstract relationships among variables. For example, if the three syllables in a triplet are represented by three variables X , Y , and Z , the ABB pattern corresponds to the algebraic relation $Y=Z$. In contrast, based on simulations with artificial neural networks, different authors proposed that infants do not need to represent open-ended regularities when learning such patterns, and might just exploit statistical regularities of various sorts (e.g., Altmann, 2002; Christiansen & Curtin, 1999; McClelland & Plaut, 1999; Seidenberg & Elman, 1999; see (Alhama & Zuidema, 2019), for a recent review)

However, identity-relations might be detected by a specialized “repetition-detector.” Specifically, Endress, Dehaene-Lambertz and Mehler (2007) used triplets of piano tones to ask whether participants could learn the identity-based rules ABB and ABA and other very simple rules not based on identity-relations (i.e., low tone-high tone-middle tone vs. middle tone-high tone-low tone). They

showed that learners relying on general algebraic relations between variables should learn both types of structures equally well, and that the statistical models of Marcus et al.'s (1999) data (e.g., Altmann, 2002; Christiansen & Curtin, 1999; McClelland & Plaut, 1999; Seidenberg & Elman, 1999) either predict that both types of structures should be learned equally easily, or that the non-identity-patterns should be easier to learn (depending on the specific models).

In contrast to both classes of models, participants were much better at learning identity-patterns than non-identity-patterns, suggesting that identity-patterns might be detected by a specialized mechanism. Further results showed that Bayesian general-purpose learners (Frank & Tenenbaum, 2011) do not explain the learning of identity-relations either (Endress, 2013). If one adopts the very assumptions that allow Bayesian learners to account for Endress, Dehaene-Lambertz and Mehler's (2007) and other data, one ends up with predictions that are either empirically refuted or highly implausible at best (see Endress, 2013, 2014, for discussion). Identity-patterns thus seem to be processed by some sort of specialized repetition-detector.

Identity-relations play an important role in many languages. For example, Semitic languages such as Arabic and Hebrew have constraints governing which consonants can occur together in a word. These constraints are based on identity-relations among these consonants (e.g., Berent & Shimron, 1997; Frisch, Pierrehumbert, & Broe, 2004; McCarthy, 1979). Other languages use reduplications for grammatical purposes (McCarthy & Prince, 1999). For example, in Marshallese, “takin” means sock, while “takinkin” means to wear socks (Moravcsik, 1978). Such reduplication occur, in some form or another, in

85% of the world's languages (Rubino, 2013).

Despite their importance for language, identity-relations can be perceived in many non-linguistic domains, and by many non-linguistic animals. Humans can compute identity matches for speech syllables, tones and visual objects (Dawson & Gerken, 2009; Endress et al., 2007; Marcus, Fernandes, & Johnson, 2007; Marcus et al., 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007), and are sensitive to the identity-relations from birth (Antell, Caron, & Myers, 1985; Gervain, Berent, & Werker, 2012; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008), though such patterns might be easier to recognize with speech material (Marcus et al., 2007). Bees can compute identity-relations for colors, gratings and odors (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001). Rats and various non-human primate species as well as various bird species can compute such relations for speech and non-speech stimuli (de la Mora & Toro, 2013; Hauser & Glynn, 2009; Murphy, Mondragon, & Murphy, 2008; Neiworth, 2013; Pepperberg, 1987; Smirnova, Zorina, Obozova, & Wasserman, 2015; Yamazaki, Suzuki, Inada, Iriki, & Okanoya, 2012; but see Spierings & ten Cate, 2016; van Heijningen, Chen, van Laatum, van der Hulst, & Ten Cate, 2013, for evidence that these relations are not equally salient to all species), and ducklings and chicks even imprint to abstract identity-relations (Martinho & Kacelnik, 2016; Versace, Spierings, Caffini, Ten Cate, & Vallortigara, 2017, but see Hupé, 2017; Langbein & Puppe, 2017).

At first sight, these results seem to suggest that identity-relations are computed by a species- and domain-general mechanism, that is, a mechanism that has access to all domains and modules. However, at least in humans, a closer examination casts doubt on this hypothesis even for linguistic stimuli. Human

adults and infants appear to readily learn identity-relations with vowels but not consonants, and seem unable to learn identity-patterns carried by syntactic categories.

3.2. Are identity-relations really domain-general? The case from language

3.2.1. The case of vowels and consonants

Adult speakers are better at detecting identity-patterns on vowels than on consonants, to the extent that they fail to detect the patterns on consonants (Toro, Bonatti, Nespor, & Mehler, 2008). In these experiments, participants were familiarized with a sequence of words in which either the vowels or the consonants conformed to an identity-pattern (e.g., words like *tapena* and *bodako*, where the first and the last vowel were identical, or words like *banubE* and *tineto*, where the first and the last consonant were identical). Participants learned the identity-pattern over vowels but not over consonants (Toro, Bonatti, et al., 2008), even when the salience of the vowels was much reduced (Toro, Shukla, Nespor, & Endress, 2008). Similar results have been observed with human infants (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Pons & Toro, 2010).

In contrast, rats learn identity-relations equally well on vowels and on consonants (de la Mora & Toro, 2013), suggesting that consonants do not have an intrinsic acoustic property that makes learning identity-relations particularly difficult.

According to Toro, Bonatti, et al. (2008), the vowel advantage for generalizations is due to the role of vowels in grammar, because vowels might play an important role for grammar, while consonants might be more important for lexical processing (Nespor, Peña, & Mehler, 2003). Given that learning

identity-patterns requires rule-like generalizations, they might plausibly be linked to grammar, and thus show a vowel advantage (but see the next section for evidence that identity-patterns might *not* be linked to grammatical processes). Irrespective of whether this interpretation is correct, these results show that learning identity-relations is much easier for some classes of linguistic stimuli than others, a result that is problematic for a fully domain-general identity-detector.

3.2.2. The case of syntactic categories

While identity-relations can be learned over a variety of stimuli, adult speakers seem unable to learn them over syntactic categories (Endress & Hauser, 2009). Specifically, participants were familiarized with word triplets conforming to either an AAB pattern (noun-noun-verb triplets such as *town-leg-choose* and verb-verb-noun triplets such as *choose-speak-leg*) or an ABB pattern (noun-verb-verb or verb-verb-noun). Following this, they were presented with AAB or ABB test triplets made from new words, and had to decide which one was like the familiarization items. Surprisingly, participants failed to discriminate between “correct” and “incorrect” patterns.

In a desperate attempt to make participants learn these patterns, Endress and Hauser (2009) primed them on nouns and verbs by having participants first categorize words into nouns and verbs; results showed that they performed at ceiling. Then they told them to watch out for a pattern of nouns and verbs, and finally familiarized them with the triplets. Even under these conditions, most participants failed to notice the patterns.

This failure is surprising because control experiments show that all the

building blocks for learning the patterns are in place. For example, participants readily access the categories. They perform at ceiling when categorizing words, and learn other sequential regularities about the categories (e.g., whether triplets start or end with nouns). Further, participants learn identity-relations over non-syntactic categories: When semantic categories were used (e.g., animals and clothes), or phonological categories, listeners readily learned the identity-relations. Nor is the failure due to ambiguities between nouns and verbs specific to English (i.e., that a verb like “run” can be used as a noun as well): Hungarian speakers (where such ambiguities do not exist) show the same pattern of failure.

It turned out that participants could learn the identity-relations over syntactic categories only when the resulting sentences had a syntactic interpretation in terms of subjects, objects and so on. For example, when noun-noun-verb (e.g., baby-water-juggle) and adjective-adjective-noun (e.g., clever-fragile-water) triplets were used for the AAB pattern, participants readily learned the corresponding pattern. While the adjective-adjective-noun order corresponds to a legal English construction, participants also learned such syntactically when they did not conform to English word order. For example, they readily learned ABB patterns such verb-noun-noun or noun-adjective-adjective. In contrast, when the triplets had no obvious syntactic interpretation (as in the AAB patterns verb-verb-noun and adjective-adjective-verb, and the ABB patterns noun-verb-verb and verb-adjective-adjective), participants failed as before.

The kinds of regularities that can be learned with syntactic categories is thus highly constrained (Moro, 2008). Most relevant for the current purposes, there seem to be surprising holes in our ability to learn such structures, a result

that is at odds with views that such relations are computed by domain-general mechanisms. Still, they can be computed in variety of domains, including tones, odors, visual stimuli and linguistic items, a result that is problematic if they were domain-specific.

3.2.3. Are there multiple identity-detectors?

One explanation for these puzzling results involves independent identity-detectors for visual, olfactive and tonal stimuli; within the language faculty, there might be independent identity-detectors for vowels, syllables, semantic categories and so on, but not for syntactic categories. If so, identity-relations can be detected in all domains that have an identity-detector, but not in domains that lack such a mechanism.

Within the domain of language, some support for the idea that the mind is equipped with multiple, independent identity-detectors comes from formal linguistics, where a very similar idea has been proposed to explain the admissible patterns of consonant repetitions in Semitic languages. Specifically, two identical consonants can occur at the end of a word root, but not at its beginning (Berent & Shimron, 1997; Frisch et al., 2004; McCarthy, 1979). The traditional explanation involves two independent computations of identity-relations, one at the phonological level, and one at the morphological level. At the phonological level, consonant repetitions are illegal in word roots (McCarthy, 1979). However, some roots have only two consonants rather than three. This creates a problem at the morphological level, where the standard word pattern has three consonant slots. As a result, one additional consonant must be added, and this consonant will be a repetition of the second consonant, leading to a consonant repetition in the word's

final syllable. Such theories thus assume that there are independent mechanisms sensitive to identity-relations at the morphological and at the phonological level.

Identity-relations might thus be computed by what we called “domain-bound” mechanisms (Endress, Nespore, et al., 2009), consisting of independent copies of mechanisms in different domains with similar computational functions; these copies might have arisen through duplication over evolution.

This view is also made plausible by the observation that identity-relations might be detected using fairly simple neuronal circuits.⁶ For example, Endress (under review) proposed a simple disinhibition-based circuit that might act as an identity detector. This idea is based on the growing evidence for the importance of “*disinhibitory*” circuits (where inhibitory neurons are inhibited by other neurons) in a variety of brain regions and taxa (Chevalier & Deniau, 1990; Goddard, Mysore, Bryant, Huguenard, & Knudsen, 2014; Hangya, Pi, Kvitsiani, Ranade, & Kepecs, 2014; Koyama et al., 2016; Koyama & Pujala, 2018; Lee, Kruglikov, Huang, Fishell, & Rudy, 2013; Mysore & Knudsen, 2012; Pfeffer, Xue, He, Huang, & Scanziani, 2013; Pi et al., 2013; Xu, Jeong, Tremblay, & Rudy, 2013), and in a variety of behaviorally important functions, including attention (van Der Velde & de Kamps, 2001; S. Zhang et al., 2014), gain control (Fu et al., 2014), discriminations of stimulus strength (Machens, Romo, & Brody, 2005; Miller &

⁶ There are a number of biologically realistic models of identity-relations, many of which were developed in the context of memory processing (Arena et al., 2013; Carpenter & Grossberg, 1987; Cope et al., 2018; Engel & Wang, 2011; Hasselmo & Wyble, 1997; Johnson, Spencer, Luck, & Schöner, 2009; Ludueña & Gros, 2013; Wen, Ulloa, Husain, Horwitz, & Contreras-Vidal, 2008). However, they do not explain at least one of two critical features of grammar learning: they either do not generalize to unseen exemplars or they require explicit training with labeled counter-examples (see Endress, under review, for a review).

Wang, 2006), categorization (Goddard et al., 2014; Kusunoki, Sigala, Nili, Gaffan, & Duncan, 2010; Mysore & Knudsen, 2012), behavioral response selection (Jovanic et al., 2016; Zhao et al., 2019), associative learning (Letzkus et al., 2011), plasticity (Fu, Kaneko, Tang, Alvarez-Buylla, & Stryker, 2015) and social behavior (Marlin, Mitre, D'amour, Chao, & Froemke, 2015; Owen et al., 2013).

Endress' (under review) model acts as a feature detector for identical relations. When this circuit is (sequentially or simultaneously) presented with two different items, these items are filtered out through inhibition. In contrast, when two identical items are presented, either sequentially one after another or simultaneously in the same display, the repetition of the items leads to a suppression of the inhibitory filtering. As a result, the repeated items can be propagated for further processing. While there is no direct experimental support for this model yet, it still suggests that simple and ubiquitous circuits might act as identity-detectors.

3.3. Are duplicates hierarchically organized?

The view that the mechanisms detecting identity-relations exist in independent copies appears to be problematic in the light of findings that infants and even bees can generalize identity-relations across modalities (Giurfa et al., 2001; Marcus et al., 2007), and that presenting multimodal items helps infants generalize such patterns (Frank, Slemmer, Marcus, & Johnson, 2009). While such results seem unexpected if an identity-detector were really domain-bound, they might reflect the hierarchical nature of perception and cognition. For example, a repeated syllable is presumably also a repeated auditory object, which is also a

repeated object. If each of these hierarchical levels has its own identity-detector, cross-modal generalizations might be possible (see also Sperber, 1994, for the proposal that the output of modules might be the input of other modules). In line with this view, seven-month-olds can learn hierarchical identity-patterns over identity-patterns (Kovács & Endress, 2014; see also Basirat, Dehaene, & Dehaene-Lambertz, 2014, for related results). For example, when exposed to triplets of syllable triplets such as *dubadu lomomo zavuvu*, infants appear to notice that the last two words have identical identity-patterns.

While it is an open question to what extent duplicate cognitive mechanisms are hierarchically organized, this view would fit well with proposals that novel cognitive functions can arise from the duplication of entire brain pathways (Chakraborty & Jarvis, 2015; Kaas, 1989). If a brain pathway becomes duplicated, its underlying computational functions become duplicated as well. Subsequently, individual computational functions (e.g., the detection of identity-relations) can be lost independently in different duplicates, either over evolution or, as I will argue below, over development.

This leads to the question of whether there are any truly domain-general mechanisms, where a single processor has access to a large number of domains. Three prime candidates come to mind: associative learning, a sensitivity to sequence-edges (i.e., the first and last position; below, edges refer to sequence edges and not visual edges) and executive control functions. All three candidate mechanisms can be used by a variety of species in a variety of domains, and might thus be a species- and domain-general mechanisms. However, I will now argue that even these candidates might not be truly domain-general.

4. Is associative learning domain-general?

If there is a mechanism that is a strong contender to be domain-general, it is presumably associative learning. Associative learning exists in many domains, including vision, olfaction, audition and motor behavior. If we hear, see or feel two objects frequently occurring together, we form associations between them (e.g., Aslin, Saffran, & Newport, 1998; C. M. Conway & Christiansen, 2005; Endress, 2010; Fiser & Aslin, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne & Scholl, 2009). Similar abilities have been observed in many non-human species, including rats, cotton-top tamarins and Zebra finches (Chen & Ten Cate, 2015; Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005). It thus seems plausible that these computations reflect species- and domain-general mechanisms.

However, such results are open to at least two interpretations. On the one hand, associative learning might reflect a single, central and domain-general mechanism that resides at a single location in the brain and that operates on all stimuli among which associations need to be formed. On the other hand, associative learning might be domain-bound: different domains might have their own, independent associative machinery. If so, associative learning should also be available in a variety of domains, in the absence of domain-general machinery.

In the following, I will show that the fully domain-general view of associative learning is untenable: certain associations cannot be learned, even closely related forms of associative learning have dissociable properties and rely on different brain areas and, across domains, associative learning abilities are

largely uncorrelated. I will also discuss a possible neural mechanism of associative learning that supports the domain-bound view as well.

4.1. The case of taste aversion

Animals can associate tastes with visceral sickness and external events such as sounds or light flashes with pain; however, they cannot (or only with great difficulty) associate taste with pain or external events with sickness (Garcia, Hankins, & Rusiniak, 1974, 1976; Garcia & Koelling, 1966). If rats ingest a flavor and then become sick, a single exposure to the flavor is sufficient for the animals to learn to avoid the taste. Likewise, rats readily learn that a sound predicts a shock, but learning is generally more gradual and slower than in the taste aversion case (though, in certain situations, faster learning is possible as well; see e.g., Domjan, 1983). However, animals fail to learn associations between tastes and electroshocks, and between sounds and sickness.

From an evolutionary viewpoint, this pattern of associations seems sensible: Visceral sickness typically results from ingesting toxins, which makes association between tastes and sickness adaptive. Further, one-shot learning would prevent repeated exposure to the toxins. Conversely, physical pain (e.g., from shocks in the experiments, but of course there is no shortage of more ecological ways in which animals can get hurt) is typically caused by external events, making associations between external events and physical pain adaptive. These and many similar results (e.g., Seligman, 1970) suggest that even associative learning does not appear to be fully domain-general, but rather

presents certain specializations.⁷ In fact, at least in *Drosophila*, preferential associations can be evolved in just 40 generations (Dunlap & Stephens, 2014).

In line with this view, some apparently closely related forms of associative learning have different properties even with respect to fundamental principles of reinforcement learning such as blocking and overshadowing.⁸ For example, associative learning of locations shows phenomena such as blocking and

⁷ In the 1970s and 1980s, there have been numerous attempts to find alternative accounts for these results that (1) do not involve adaptive specializations, and (2) fit within general learning theory (see e.g. Domjan, 1983, 2015, for a review of sometimes somewhat complex alternative proposals, and evidence that contradicts at least some of these proposals). One such counterargument relies on earlier experience: maybe rats have learned that tastes are more predictive of sickness than sounds, and vice versa. Critically, however, the preferential associations have been observed in 1 and 5 day-old rats (Gemberling & Domjan, 1982; Gemberling, Domjan, & Amsel, 1980), suggesting a rather limited role of experience. Other developmental studies also reveal the exquisite tuning of preferential associations to the environmental niche that rat pups occupy, and are problematic for most proposals that do not consider preferential associations to be adaptive specializations. For example, before the age of 21-22 days, infant rats are monophagous, and drink only their mothers' milk. They can certainly acquire taste aversions at this age, but if the taste is presented in a suckling context, no taste aversion develops; in contrast, older infants (who also eat non-milk foods) develop taste aversion in a suckling context as well (Martin & Alberts, 1979). Critically, associations between tactile stimuli (e.g., air puffs) and electric shocks are *not* blocked in a suckling context (Alberts & Gubernick, 1984). The most striking aspect of these data comes from pups for whom weaning is delayed by preventing them from exposure to non-milk foods. Specifically, in 26-day-old pups, suckling usually does *not* block taste aversion, but, when weaning is delayed, a suckling context prevents taste aversion in these older pups as well. In contrast, just four hours of experience with (non-toxic) solid food is sufficient to prevent blocking of taste aversion – irrespective of whether this experience is given two days prior to taste aversion training, or two days *afterwards* (Gubernick & Alberts, 1984). To my knowledge, no theory has been developed that explains these dissociations between classical conditioning and taste aversion learning that does not appeal to adaptive specializations.

⁸ Blocking refers to a reduction in the conditioned response when conditioning occurs in the presence of a previously conditioned stimulus. For example, if an animal has already learned that the sound of a bell predicts a reward, and the reward is then paired with the bell and a drum sound presented *simultaneously*, the association between the reward and the drum sound will be much reduced. Overshadowing is a preferential association of the most “salient” stimulus with a reward if several stimuli are simultaneously paired with the reward.

overshadowing when learning occurs with respect to landmarks, but not with respect to boundaries (Doeller & Burgess, 2008): Reward associations with boundaries overshadow associations with landmarks but not vice-versa. Likewise, reward associations with distinct landmarks block each other while associations with boundaries show no blocking. Further, these types of learning rely on different neural substrates (Doeller, King, & Burgess, 2008).

Other cases against the domain-general view involve preferential associations even within the domain of language. For example, associative learning works better over consonants than over vowels (Bonatti, Peña, Nespor, & Mehler, 2005), although the reasons are debated (Bonatti, Peña, Nespor, & Mehler, 2007; Keidel, Jenison, Kluender, & Seidenberg, 2007)

It thus appears that associative learning is not a unitary, domain-general mechanism. Rather, independent associative learning abilities might exist in a variety of domains, perhaps with slightly different properties, which would suggest that they might have been duplicated.

4.2. The case of cross-domain correlations

Further evidence that associative learning might not be a unitary ability comes from Siegelman and Frost's (2015) experiments. They administered a battery of statistical learning tasks, and concluded that they did not reflect a unitary ability (see Frost, Armstrong, Siegelman, & Christiansen, 2015, for a different take on these results from that advocated here).⁹ Specifically, they

⁹ In contrast, when using familiar and meaningful sounds (e.g., the sound of a bell) rather than arbitrary computer sounds, Siegelman et al. (2018) found that visual statistical learning was related to auditory statistical learning.

administered five statistical learning tasks to the same (human adult) participants that differed in whether participants had to compute statistical relations among adjacent syllables, among adjacent non-linguistic sounds, among non-adjacent consonants, or among adjacent visual shapes; they also performed a serial reaction time task. Surprisingly, performance in these different tasks was virtually uncorrelated. This would be unexpected under a domain-general view. After all, if statistical learning reflects a unitary, domain-general mechanism, individuals who are good statistical learners in one domain should also be good statistical learners in another domain. It thus appears that there is no single, overall mechanism that computes associations; rather, associations seem to be computed by more localized mechanisms.

A lack of correlation between statistical learning performance in different tasks does not even require the use of stimuli in different domains. For example, Thompson, Kahn, Falk, Vettel and Bassett (2019) asked participants to learn the statistical structure of a set of fractal pictures. Critically, they told participants either that the fractals represented persons (e.g., avatars on a social media platform) or rock formations as seen through a microscope. Surprisingly, performance in the social condition, where participants thought they were learning information about people, was uncorrelated with performance in the rock condition, suggesting again that separable statistical learning mechanisms operate in different domains.

Such a pattern of results has an intuitive explanation. For the sake of the argument, let us assume that associations are computed according to the principle of “what fires together fires together”, and that this type of correlational learning

reflects long-term potentiation (LTP) at the neuronal level (see M. A. Lynch, 2004, for a discussion of this possibility, but see Gallistel, 2017¹⁰). Given that LTP has been observed in different cell types within different brain regions, albeit with slightly different properties (Malenka & Bear, 2004), statistical learning should occur in different brain regions as well, again with globally similar, but slightly different properties, and, crucially, in the absence of a unitary statistical processor. Rather, all brain regions would have their own associative machinery because they have the machinery required for LTP. Hence, there is no a priori expectation either that all associations are equally easy to compute.¹¹

The situation might be similar with respect to identity-relations: different brain regions might have their own “identity-detectors”, with slightly different properties each. As a result, such mechanisms are neither domain-general nor domain-specific, but rather consist of a collection of localized and domain-specific copies of the same mechanism. We have called this situation “domain-bound” (Endress, Nespore, et al., 2009). Such a situation might arise if key mechanisms are duplicated over evolution, and potentially acquire domain- and species-specific properties; as mentioned above, this possibility is made plausible

¹⁰ Gallistel (2017) argues that associative learning might not rely on LTM, and might rather rely on intracellular processes. While Gallistel (2017) makes a strong case, for the current purposes, I just assume that associative learning relies on *some* widely available mechanism, and use LTM as a label for it.

¹¹ This view also suggests that correlated task performance across different domains would *not* provide evidence for domain-general mechanisms even if such correlations could be found. It is possible that individuals differ in some respect that makes all circuits underlying these abilities more or less effective, even though there might be distinct and independent instantiations of these circuits in different brain areas. In contrast, an absence of cross-domain correlations probably provides evidence against a purely domain-general ability.

by the observation that, as in the case of associative learning, identity-detection might be implemented by fairly simple circuits.

5. Is a sensitivity to sequence-edges domain-general?

Humans, and other animals, are sensitive to the items that occur at sequence edges (i.e., in the first and last position; below edges refer to sequence edges and not visual edges). This sensitivity is probably due to a mechanism of serial memory that encodes the positions of items in sequences relative to the first and last item. For example, in sequences like ABCD, this mechanism retains that A came first, that D came last, and that B and C had some position *relative* to the first and last one (for reviews, see e.g. Fischer-Baum et al., 2011; Fischer-Baum & McCloskey, 2015; Henson, 1998).

Such regularities are extensively used in language. For example, affixes (e.g., the English *-ed* past-tense) tend to occur at word edges, while infixes (e.g., fun-fucking-tastic; McCarthy, 1982) are relatively rare. Further, there are many other linguistic processes that rely on edges, from stress assignment to the coordination of different linguistic hierarchies (see Endress, Nespor, et al., 2009, for a review).

In humans, learning of such affixation-like regularities has been demonstrated for syllable sequences, visual action sequences and non-verbal vocalizations (Endress, Carden, Versace, & Hauser, 2010; Endress & Hauser, 2011; Endress & Mehler, 2009; Endress & Wood, 2011; Marchetto & Bonatti, 2013), though a sensitivity to edges is also important for other aspects of language, notably word segmentation (Monaghan & Christiansen, 2010; Seidl & Johnson, 2006, 2008; Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin,

2011), and the representation of written words (Fischer-Baum et al., 2011; Fischer-Baum, McCloskey, & Rapp, 2010).

As with identity-relations, a sensitivity to sequence-edges is found in many non-human animals, including chimpanzees, cotton-top tamarins, and Zebra finches (Chen, Jansen, & Ten Cate, 2016; Endress, Cahill, Block, Watumull, & Hauser, 2009; Endress et al., 2010). Both cotton-top tamarins and Zebra finches can even learn language-like affixation patterns (Chen et al., 2016; Endress, Cahill, et al., 2009). As a result, a sensitivity to edges is clearly not domain- or species-specific.

However, the patterns found in the world's languages suggest that the sensitivity to edges is not domain-general either. Various theories from formal linguistics assume independent copies of edge-based codes for different linguistic constituents (see Endress, Nespor, et al., 2009, for a review), and these codes account for numerous linguistic regularities (McCarthy & Prince, 1993).

Phonotactic constraints are a case in point. Such constraints determine the permissible phoneme sequences in a language. For example, English words cannot start with the sound at the end of “sing,” and cannot end with the sound at the onset of “hat,” while other languages have different constraints. Many constraints appeal to edges of constituents, and some phonotactic appeal to edges at different levels of the prosodic hierarchy. For example, in some languages such as German, syllable-final obstruents become devoiced in all syllables (e.g., sounds such as /d/ are pronounced as /t/); in other languages such as Polish and Walloon, syllable-final obstruents are devoiced only at the end of (prosodic) words, and thus at the edge of constituents at a different level of the prosodic hierarchy

(Wetzels & Mascaró, 2001). Further, the constraints that can be learned at the edges of words are different from those that can be learned at syllable edges that are not word-edges (Endress & Mehler, 2010). If humans did not have access to multiple and independent representations of edges, such results would be difficult to explain.

Another example for the importance for multiple, independent edge representations comes from alignment theories (McCarthy & Prince, 1993; Nespor & Vogel, 1986). For example, the English plural [s] is a morpheme (in the morphosyntactic hierarchy), but not a syllable (in the prosodic hierarchy). However, even though the two hierarchies are not isomorphic, the right edge of the [s] morpheme is aligned with the syllable comprising the plural [s].

Alignment of edges of different linguistic constituents can explain a variety of linguistic regularities, from syllabification to morphology and the coordination of different hierarchies (McCarthy & Prince, 1993; Nespor & Vogel, 1986). Critically, however, different edges can only be aligned if each linguistic constituent has its own edges to begin with. In the terminology used here, this would mean that each linguistic constituent has its own, domain-bound sensitivity to edges.

There is some indirect evidence for this possibility from brain imaging experiments. Specifically, different English grammatical affixes (i.e., the plural and the third person *-s*) activate different brain regions (Longe, Randall, Stamatakis, & Tyler, 2007), suggesting that there are at least partially independent representations of an [s] in an edge position. However, it is unclear whether such activation differences reflect the affixes per se, or their different grammatical

roles. For example, different brain activations for the inflection of nouns and verbs might not reflect the affixation-pattern per se, but rather the interpretative processes associated with their respective grammatical functions.

Be that as it might, the data reviewed above make it at least plausible that independent sensitivities to edges might exist in different domains. Further, as in the case of the ubiquitous machinery for associative learning, this view is also supported by the observation that circuits that are sensitive to onsets and offsets have been observed in various brain areas.

In the visual modality, ganglion cells responding to onsets and offsets of signals have been reported since Hartline (1938; see also Nelson & Kolb, 2003). In the auditory modality, neurons that respond to sound onsets and offsets are found throughout the auditory hierarchy (see He, 2001, and references therein). Onset and offset responses seem to rely on different inputs, as they are tuned to different frequencies and have a different balance of excitation and inhibition, though the latter might be a consequence of the former (Scholl, Gao, & Wehr, 2010). However, the kinds of edges that are relevant for linguistic alignment theories are often abstract, with few perceptual cues. To be sensitive to such edges, the edges must be computed at more abstract levels as well. As a result, it is an important question for further research whether the neural mechanisms that are used to compute perceptual sequence edges can also be used to compute more abstract edges. However, it is at least plausible that these mechanisms might reflect simple circuits that are duplicated profusely.

6. Are cognitive control mechanisms domain-general?

Another set of mechanisms that might be domain-general comprises attentional

and executive control mechanisms. These mechanisms include working memory, inhibition, and cognitive flexibility (Diamond, 2013). Many of these mechanisms have widespread effects on our mental life and, in some cases, on our non-mental life as well. For example, resistance to temptation in preschoolers (e.g., in the marshmallow test) famously predicts outcomes in variety of domains, from social functioning to academic performance (Mischel, Shoda, & Rodriguez, 1989).

However, as in the other case studies above, such results are open to two interpretations. On the one hand, the mind might feature unitary and domain-general control mechanisms. On the other hand, the control mechanisms might be domain-bound as well, with independent control mechanisms in different domains.

I will now review evidence that is problematic for the domain-general view for two key executive functions: Working Memory (WM) and Inhibition. With respect to WM, different studies suggests that it might not be entirely unitary. With respect to inhibition, performance on different inhibitory tasks is sometimes uncorrelated, some animals might have a domain-specific inhibitory system, and some recent neurobiologically realistic models support the domain-bound view.

6.1. The case of Working Memory (WM)

WM is an Executive Function that is at least partially domain-specific. While WM obviously has a memory component that lets us store information, the most prominent theories of WM hold that items in WM are actively maintained by attentional or executive mechanisms. For example, in Cowan's (1995) model, we retain items in WM by (mentally) attending to them. Accordingly, we have a

capacity limit of three or four items because we cannot attend to more than three or four items simultaneously. Likewise, other authors proposed that what determines WM is Executive Attention (Braver, 2012; Engle, 2002; M. J. Kane, Bleckley, Conway, & Engle, 2001; Michael J. Kane & Engle, 2003; Michael J. Kane, Poole, Tuholski, & Engle, 2006), such that WM scores predict outcomes from IQ (A. R. Conway, Kane, & Engle, 2003) to educational achievement (Alloway & Alloway, 2010; Gathercole, Pickering, Knight, & Stegmann, 2004) to the frequency of mind wandering (Michael J. Kane et al., 2007). In other words, our WM limitations are really limitations of attention or executive function (see, among many others, Braver, 2012; A. R. Conway et al., 2003; Cowan, 1995; Engle, 2002).

If control functions such as executive function or attention were domain-general, one would thus expect WM to be domain-general as well. This, however, does not appear to be the case. In fact, there is substantial evidence that WM is at least in part domain-specific (Baddeley, 1996; Cowan, Saults, & Blume, 2014; Fougner, Zughni, Godwin, & Marois, 2015; Wong, Peterson, & Thompson, 2008; Wood, 2008, 2009). For example, verbal WM is famously independent of WM in other domains (Baddeley, 1996, 2003).

As a result, if the mechanisms at the root of WM are control mechanisms such as attention and executive control (e.g., A. R. Conway et al., 2003; Cowan, 1995; Engle, 2002, but see Banta Lavenex, Boujon, Ndarugendamwo, & Lavenex, 2015; Carroll et al., 2010; Endress & Potter, 2014; Endress & Szabó, 2017; Shipstead & Engle, 2013), and if WM is at least partially domain-specific, then these underlying processes must be domain-specific as well, and there should

domain-specific attentional or executive functions.

6.2. The case of inhibition

Inhibition is a second Executive Function that does not appear to be fully domain-general. The term inhibition can refer to at least three potentially separable aspects: prepotent response inhibition, resistance to distractor interference, and resistance to proactive interference (Diamond, 2013; Friedman & Miyake, 2004). However, these aspects are not necessarily fully independent (see Friedman & Miyake, 2004, and below). Inhibition can have wide-ranging consequences in different domains. For example, resistance to temptation in preschoolers (e.g., in the marshmallow test) famously predicts outcomes in variety of domains, from social functioning to academic performance (Mischel et al., 1989). More generally, Diamond (2013) reviews data suggesting that executive control functions are impaired in mental disorders from addictions to schizophrenia and predict outcomes from physical health to school and job success to marital harmony. At first sight such results seem difficult to explain if executive function were *not* domain-general.

However, in other cases, inhibitory abilities seem much more linked to specific domains, suggesting that inhibitory abilities are at least partially domain-bound and that independent inhibitory abilities exist in different domains. In this section, I will suggest that the research on the ability to resist distractors echoes that on Statistical Learning: Different measures of distractor resistance show little correlation with each other, suggesting that, as in the case of Statistical Learning, there are several independent mechanisms for distractor resistance. Following this, I show how such specialized inhibitory abilities can evolve, albeit using

resistance to prepotent responses as a case study. I then argue that recent neurobiological models of inhibition support a domain-bound view, and end by discussing how the far-ranging consequences of inhibitory abilities can be explained under a domain-bound view.

6.2.1. Correlations across distractor resistance tasks

As in the case of Statistical Learning, evidence from cross-tasks correlations suggests that inhibitory abilities are at least partially independent. For example, Shilling, Chetwynd and Rabbitt (2002) used four Stroop-like tasks, and measured correlations in performance.¹²

While performance in each task was consistent across blocks in a given participant, performance across tasks was uncorrelated (see also Bjorklund & Kipp, 1996, for a review of studies suggesting that inhibitory abilities are at least partially independent, though cross-task correlations might be observed when the tasks are extremely similar; Shilling, Chetwynd and Rabbitt, 2002).

Likewise, Cipolotti et al. (2016) tested frontal patients on a Stroop task, and on the Hayling Sentence Completion test, where participants have to complete a sentence fragment with an *unrelated* word. For example, a correct

¹² During the first task, participants saw a digit made of other digits (e.g., a big 2 made of individual 2 characters). If participants have to report the small digit, they show Stroop-like interference when the global digit mismatches the local one (Navon, 1977). During the second task, participants saw one to four identical digits on a screen, and had to report their number. Stroop-like interference ensues if the identity of the digits does not match their number (e.g., reporting that there are four digits in 3 3 3 3; Windes, 1968). In the third task, participants saw a large arrow pointing in one of the four cardinal directions, and with a word corresponding to a direction printed inside the arrow; a mismatch between the printed direction and the arrow direction (e.g., the word LEFT printed inside an arrow pointing to the right) leads to interference (Shor, 1970). The fourth task was the classic Stroop task, producing interference between color labels and their font color (Stroop, 1935).

completion of the fragment “London is a very busy...” would be an *unrelated* word such as “...banana”; in contrast, a continuation with a related word such as “... city” would be an incorrect response. Cipolotti et al. (2016) found that, after controlling for fluid intelligence, performance on the two tasks was uncorrelated. Further, patients with left-lateralized lesions were impaired on the Stroop task, but much less impaired on the sentence completion task, while patients with right-lateralized lesions showed the opposite pattern.

Similar results have been found with dogs (Brucks, Marshall-Pescini, Wallis, Huber, & Range, 2017): Performance on tests of inhibitory control do not seem to correlate with each other, suggesting that the underlying mechanisms are at least partially dissociable.

Together, results such as Shilling et al.'s (2002), Cipolotti et al.'s (2016) and Brucks et al.'s (2017) thus echo the statistical learning results by Siegelman and Frost (2015), who showed that statistical learning abilities are largely uncorrelated across domains. As mentioned above, such results are difficult to reconcile with a purely domain-general view. After all, if inhibition (or statistical learning) were domain-general, individuals who are good at one inhibitory task should also be good at another inhibitory task. As a result, there appear to be at least partially independent inhibitory abilities.

6.2.2. How do domain-bound inhibitory mechanisms evolve?

The discussion so far suggests that there are independent inhibitory mechanisms even for rather similar tasks. This raises the question of how such mechanisms might evolve. It turns out that, for some animals, domain-bound inhibitory mechanisms are important for their survival.

Cleaner wrasse are a case in point because, in the wild, they face a marshmallow test of sorts in feeding contexts. They are fish who feed on ectoparasites carried by “client” fish. However, they prefer to eat the client fish’s mucus (Grutter & Bshary, 2003). As client fish punish cleaners who feed on their mucus (Bshary & Grutter, 2002; Bshary & Schäffer, 2002), cleaner wrasse have to inhibit their tendency to feed on client mucus to avoid punishment.

Danisman, Bshary, and Bergmüller (2010) asked whether this led to better domain-general inhibition in a reverse reward contingency task. In this task, subjects have to choose the smaller of two rewards to obtain the larger one. This task has been used as a test of inhibitory function in many species and is typically difficult even for primates (Addessi & Rossi, 2011; Boysen & Berntson, 1995; Murray, Kralik, & Wise, 2005; Vlamings, Uher, & Call, 2006; see Shifferman, 2009, for a review). If the requirement for inhibition during foraging led to an improvement in domain-general inhibition, cleaner wrasse should perform well on the reverse-contingency task.

In contrast to this prediction, most cleaner fish simply failed on this task. While there are probably many differences between inhibition in a feeding context and the task demands of a reverse contingency task, these results are at least compatible with the interpretation that cleaner fish did not develop better domain-general inhibitory abilities, and that some animals might have domain-specific copies of inhibitory control mechanisms due to their feeding ecology.¹³

¹³ Domain-specific inhibitory abilities have been proposed in humans as well. For example, Bjorklund and Kipp (1996) proposed that females had better inhibitory control in social contexts, notably for controlling how they displayed emotions, and for inhibiting behaviors,

6.2.3. How are inhibitory mechanisms related?

The results discussed so far do not imply that different inhibitory tasks have nothing in common. For example, and as mentioned above, Friedman and Miyake (2004; see also Miyake et al., 2000) separated inhibitory mechanisms into three distinct complexes – prepotent response inhibition, resistance to distractor interference, and resistance to proactive interference. Participants then completed three tasks tapping into each of these complexes, as well as a number of other tasks.¹⁴ As Shilling et al. (2002), Friedman and Miyake (2004) found that correlations across tasks were fairly low. However, using structural equation modeling, they found that the tasks testing prepotent response inhibition and resistance to distractor interferences correlated with a common latent variable. Further, the tasks testing resistance to proactive interference correlated with a different latent variable. A very similar conclusion has been reached in dogs (Brucks et al., 2017): As mentioned above, performance on tests of inhibitory control did not seem to correlate with each other. However, the tasks mapped onto

notably for resisting temptation and self-regulation and that this inhibitory female advantage did not translate to an overall inhibitory advantage outside of the social domains. However, Bjorklund and Kipp's (1996) review was based on experiments with children whose focus were not necessarily gender-differences. In young adults, evidence for such a female advantage is mixed (e.g., Cross, Copping, & Campbell, 2011; Silverman, 2003a, 2003b), and might depend on the stage of the menstrual cycle (Hosseini-Kamkar & Morton, 2014). As a result, the evidence for such gender-differences appears strongest in populations that do not yet reproduce sexually, and might also be due to maturational differences between boys and girls.

¹⁴ Prepotent response inhibition was measured through an antisaccade task, a stop signal, where participants had to withhold a trained response on a subset of the trials, and by a Stroop task. Resistance to distractor interference was measured using the Eriksen flanker task, a word naming task in the presence of distractor words, as well as a shape matching task in the presence of distractor shapes. Resistance to proactive interference was measured in a Brown-Peterson task, a paired associates task and a cued recall task, where participants learned both a target and a distractor list.

three common factors (albeit different ones from those found in humans), namely persistency, compulsivity and decision speed.

Friedman and Miyake (2017) thus suggested that the common prepotent response and distractor inhibition variable might reflect goal maintenance. However, although both of Cicolotti et al.'s (2016) inhibition tasks might well map onto this variable, they involved dissociable brain mechanisms, suggesting that the question of how different inhibitory tasks are related is still an open one.

6.2.4. Neurobiological models of inhibition

The results reviewed so far suggest that inhibitory abilities are partially independent and dissociable. I thus suggest that they are domain-bound and might reflect duplicates of an inhibitory circuit. This possibility is consistent with some recent neurobiological models of inhibition. It has long been observed that frontal regions are activated by a variety of tasks, and might thus implement domain-general mechanisms (e.g., Duncan & Owen, 2000; Fedorenko, Duncan, & Kanwisher, 2013), leading some investigators to call these regions the “multi-demand cortex” (MDC; e.g., Duncan, 2010). However, this common activation in the MDC might reflect the requirement to dynamically track task- and goal-relevant aspects of the task (e.g., Duncan, 2010). Critically, recent models suggest that this might occur through a distributed set of duplicated (or at least similar) circuit mechanisms. Specifically, if the MDC tracks task- and goal-relevant representations, it might provide top-down amplification of these representations. In contrast, (inhibitory) control might occur in local circuits through local competition among representations. Due to their top-down amplification, task-relevant representations are more likely to win this competition against irrelevant

representations that do not receive such amplification (e.g., Egner & Hirsch, 2005; Erika-Florence, Leech, & Hampshire, 2014; Hampshire & Sharp, 2015; see also Banich & Depue, 2015; Munakata et al., 2011, for related proposals). If so, even cognitive control might be implemented through potentially duplicated mechanisms that might be as simple as lateral inhibition.¹⁵

This view also provides a rather straightforward account of apparently domain-specific control abilities. Species and domains with domain-specific control abilities might simply have an additional way to maintain goal-relevant representations. For example, in cleaner fish, detecting a live client fish might lead to amplification of the relevant behavioral responses (feeding on ectoparasites rather than mucus), making these responses more likely to win in *local* competitive interactions among possible behavioral responses. Further, even though inhibition has the connotation of being effortful and related to self-control, this amplification might be fairly automatic and stimulus-driven: salient stimuli might simply activate (and thus amplify) salient goals. If this speculation is correct, inhibition might fundamentally rely on duplicated, inhibitory circuits across the brain.

This view would also be consistent with proposals that the mind contains multiple “allocation mechanisms” that route information to the appropriate processing mechanisms (Barrett, 2005; Samuels, 2000, 2012). For example, Barrett (2005) proposed that information gets routed in analogy to the lock-and-

¹⁵ It should be noted that the view that inhibition often consists of goal monitoring and maintenance is not universally accepted (e.g., Anderson & Hanslmayr, 2014, but see e.g., Chatham et al., 2012; Jonker, Seli, & MacLeod, 2015).

key mechanism of enzymatic reactions. Enzymes are molecules that chemically modify other molecules (their substrates). When an enzyme and a substrate molecule encounter each other (through random diffusion of the molecules), they can “recognize” each other because the substrate has certain features that allow it to attach to the enzyme, similarly to how a key fits only into specific locks.

Mutatis mutandis, the mind might analyze information by an array of feature detectors; if a piece of information matches the features, it then gets routed to the appropriate processing mechanisms. If such a processing mechanism happens to exert top-down control, it might provide a domain-bound control mechanism.

6.3. Are Executive Control functions fully domain-bound? The possible role of hierarchical structure

The data reviewed so far suggests that at least some inhibitory functions are domain-bound, with independent inhibitory mechanisms in different domains. These mechanisms might thus have evolved through duplication. However, inhibitory functions also have wide-ranging consequences in our mental and physical lives (Diamond, 2013; Mischel et al., 1989). How can we reconcile these results?

As in the case of cross-domain effects in the acquisition of identity relations, I suggest that the answer to this question is the hierarchical nature of processing. If goals are maintained (and thus amplified) by mechanisms at the top of the processing hierarchy, these top-level control mechanisms are expected to influence processing in a wide variety of domains, even though these top-level control mechanisms might be just one of many duplicate goal-maintenance mechanisms that exist in the brain. In line with this view, there is evidence that

certain prefrontal regions can exert widespread suppression over broad brain regions (Banich & Depue, 2015; Munakata et al., 2011).

In other words, I propose that, among the (domain-bound and duplicated) goal maintenance mechanisms we possess, some might exist at the top level of the processing hierarchy, which, in turn, might enable them to exert far ranging influences.

7. Alternative explanations

I have reviewed a variety of results that seem to call for duplicated cognitive mechanisms. These results are of four types. First, some mechanisms are available in many domains, but not in others. Second, such mechanisms perform the same computations in multiple domains simultaneously. Third, performance on tasks that appear to reflect very similar computations is uncorrelated across domains. Fourth, there are biologically plausible candidate implementations that are fairly simple, and might exist widely in the brain. I will now discuss two alternative solutions to the first two observations (though they do not address the remaining observations). Specifically, the putatively duplicated mechanisms might really reflect single central mechanisms. However, these central mechanisms might not be “connected” to some domains, and might support multiplexing so that they can deal with multiple regularities simultaneously.

7.1. Are domain-general mechanisms selectively connected to different domains?

A possible alternative to duplicated mechanisms are central mechanisms that are selectively connected to some domains, but not others. For example, a single, central identity-detector might be connected to a variety of domains, but not to

those domains dealing with consonants, syntactic categories and so on.

Developmental considerations make this view particularly plausible. In particular, 4-month-olds might learn certain identity-patterns over musical material that 7-month-olds cannot learn (Dawson & Gerken, 2009). This observation raises the possibility that apparently domain-bound mechanisms are initially present in *all* domains, but then get pruned. This view would be analogous to how certain perceptual abilities are lost within the first year of life (Hannon, Soley, & Levine, 2011; Hannon & Trehub, 2005; Kelly et al., 2007; Pascalis, de Haan, & Nelson, 2002; Polka & Werker, 1994; Sugita, 2008; Weikum et al., 2007; Werker & Tees, 1984). For example, while infants can initially discriminate all phonemes from all human languages, they lose those discrimination abilities that are not used in their native language within their first year of life (Polka & Werker, 1994). Similar results have been found in face perception (Kelly et al., 2007; Pascalis et al., 2002).

This pruning might take at least two routes. First, there might be central mechanisms (e.g., an identity-detector), and the connections of these mechanisms to different domains get cut. However, existing data are problematic for this and other accounts, at least when pruning is assumed to be complete. For example, and as mentioned above, infants older than seven months have difficulty learning identity-relations with musical stimuli. However, they perform better when exposed to identity-relations with speech stimuli before (Dawson & Gerken, 2009; Marcus et al., 2007). Under a domain-general view with a single, central identity-detector, it is not clear why training with one set of stimuli should improve detection with a different set of stimuli; after all, the musical stimuli

should either be connected to the identity-detector, or they should not.

In line with this view, once perceptual narrowing took place, exposure to the relevant stimuli can reverse it, both during infancy (Pascalis et al., 2005; Sugita, 2008) and also later on during childhood (Sangrigoli, Pallier, Argenti, Ventureyra, & Schonen, 2005). As a result, it does not seem to be the case that the connections between a central processing mechanisms and the different domains simply get cut.

A possible explanation of Dawson and Gerken's (2009) results is that both speech items and musical stimuli are connected to this central identity-detector. However, the connections between the musical stimuli and the repetition-pattern might be weak; as a result, the identity-detector needs to be primed with speech items so that musical stimuli can drive it. While this view would obviate the need for duplicated mechanisms, more research is needed to establish it.

Crucially, pruning might also occur through a second route that is compatible with the duplication view. Specifically, each domain might initially have its own copies of the duplicate mechanisms, and these copies are lost over development. Further, if entire brain pathways can become duplicated (Chakraborty & Jarvis, 2015), their constituent computational mechanisms can be lost individually in each duplicate, either over evolution or over development through pruning.

In addition to pruning, there is, at least theoretically, a second route by which duplicate mechanisms might appear only in certain domains but not others: Each domain might have the *potential* for certain computations (that might be inherited from the brain region from which the corresponding brain region is

copied), but these computations might require appropriate environmental stimulation for realizing this potential. For example, and as mentioned above, the visual word form area in the left lateral occipitotemporal sulcus might be particularly suitable for word processing due to its connectivity with other brain areas, but requires specific stimulations (e.g., reading instruction) for becoming selective for reading visual words (Dehaene et al., 2010; Hannagan et al., 2015; Saygin et al., 2016). If so, it is at least theoretically possible that some duplicate computational mechanisms might remain dormant in most domains, and are realized only in certain domains due to relevant environmental input.

Be that as it might, a central, domain-general mechanism (whose connections to certain domains get pruned) does not explain why performance on tasks such as the Stroop task or statistical learning is uncorrelated across domains (e.g., Cipolotti et al., 2016; Shilling et al., 2002; Siegelman & Frost, 2015; Thompson et al., 2019), nor how the same mechanism can be used simultaneously on multiple stimuli if there is only a single copy of the mechanism. I will discuss the latter issue in the next section.

7.2. Can domain-general mechanisms be multiplexed?

As mentioned above, there is theoretical and empirical evidence suggesting that the same computations can be performed independently and simultaneously on multiple stimuli. Examples include the restrictions on consonant repetitions in Semitic languages (Berent & Shimron, 1997; Frisch et al., 2004; McCarthy, 1979), infants' ability to learn multi-level repetition-patterns (Kovács & Endress, 2014), and the requirement for multiple independent edges in alignment theories.

One possible solution to make such results consistent with single copies of central, domain-general mechanisms is to postulate that these central mechanisms are capable of “multiplexing:” They might be able to process stimuli from different domains by switching back and forth between the stimuli, or possibly by keeping the stimuli separate using some other mechanism, similar to how a radio receiver can separate the signals of different radio stations although they are all carried by different frequencies in the electromagnetic spectrum. However, such multiplexing capabilities come at the cost of a rather complex, and possibly implausible, architecture, compared to the fairly simple circuits that might support local computations. I will now outline these requirements in more detail.

In the case of edges, alignment theories require the detection of simultaneously occurring edges from different constituents, which likely rules out time-sharing-based multiplexing, and likely requires multiple edge detectors within the central edge detector. Moreover, it requires the central mechanism to keep track of the sources of the different edges. For example, when, during a conversation, a phone starts ringing, the ringtone certainly provides auditory edges, but these edges are unlikely to have any effect whatsoever on grammatical processing of edges during the conversation, among many other reasons because the ringtone is perceived in a different “stream” from speech (Bregman, 1990). As a result, it is not sufficient to simply detect edges, but the central mechanism would also need to keep track of the *kinds of edges* it detects, and process them accordingly. A central, domain-general edge-detector would thus likely postulate *central* copies of the edge-detector *within* the central edge-detector instead of local distributed copies, and would additionally require considerable processing

capabilities to keep track of the kinds of edges it is processing.

Such a view is also in direct contradiction to the hierarchical organization of perception and cognition. After all, to detect edges in a variety of domains, and for different linguistic constituents, such a mechanism would need access to fairly low-level information from a variety of domains and process it.

For these reasons, I believe that a central, domain-general edge-detector is implausible at best. In contrast, local (copies of) edge-detectors can rely on fairly simple and possibly widespread circuits, and detecting edge-alignment would require little more than coincidence detection of the output of these local circuits, even though it is an open question whether these simple circuits would support more abstract edges as well.

A central, domain-general mechanism for repetition-detection is implausible for related reasons, but would additionally require a working-memory component. Specifically, a central repetition-detector might process hierarchical repetition-patterns of repetition-patterns (Kovács & Endress, 2014), either through some form of time-sharing-based multiplexing or through some other route. If it uses time-sharing, it needs to retain the results of the lower-level repetition-detection in some kind of memory; these memories then need to be compared in order to detect the higher-level repetition-pattern. If multiplexing occurs through some other route, the central repetition-pattern would likely need multiple internal copies of a repetition-detector that also keep track of the source of the repetition. Again, local copies of the mechanisms would simply be replaced with central copies.

In contrast to this complex model, repetition-patterns of repetition-patterns

can be detected in a straightforward way by local, but hierarchically organized copies of repetition-detectors: repetitions might be detected at the lower level, and the output of these detectors might be compared at the next level. In fact, there is evidence that the output of an identity detector can be used for further processing. For example, five-month-olds can form associations between identity patterns and visual symbols (Kabdebon & Dehaene-Lambertz, 2019), while at least bilingual 12-month-olds can form similar associations between identity patterns and spatial locations (Kovács & Mehler, 2009). Hence, local, hierarchically embedded identity detectors seem a more plausible architecture than a set of central copies of identity detectors with some kind of memory component.

8. Elemental computations vs. domains

The discussion so far focused on elemental computations such as identity-detectors and statistical learning. However, discussions of domain-specificity and domain-generality typically focus on larger and complex entities, such as a “language module” or a “face processing module,” whether they exist at all, and whether they are specific to humans. However, as these “modules” are defined as the sets of mechanisms that deal with a certain type of information (e.g., language), they are likely complex and diverse, and likely use many different elemental computational mechanisms. For example, the language “module” includes identity-relations (for reduplication), pitch perception (for prosody), the syntactic computations involved in c-command (whatever they are), and many others. Some of these elemental computations are likely available in other domains and species as well; others might be specific to one domain. Discussions of domain- and species-specificity thus become more falsifiable if they are

situated at the level of computations as well as their implementations that might or might not be domain- and species-specific.

This conclusion echoes the discussion of a viable notion of a “domain” from the introduction. Definitions that are purely based on informational content (e.g., “faces”, “language”, and so forth) are unlikely to account for what is specific about a domain (and for the hierarchical organization of cognition). That is, while the arguments that distinguish the function of a mechanism from its implementation are as valid and important as they always were to understand any mechanism (Marr & Nishihara, 1992), elucidating the specificity of mechanisms requires integrating both their functions and implementations.

9. Conclusions

The case studies reviewed above suggest that some cognitive mechanisms might have become duplicated for the course of evolution. There are three types of evidence for this view. First, some computations can be performed in many domains, but not in others. Second, some tasks that tap into apparently similar computations are largely uncorrelated across domains; further, tasks such as statistical learning seem to have different properties depending on which can of stimuli it operates over. Third, some computations can be performed simultaneously in multiple domains or at different hierarchical levels. All of these observations are explained naturally if each domain has its own instantiations of the underlying mechanisms.

There is a fourth but more speculative argument for the existence of duplicated cognitive mechanisms. At least the case studies reviewed here might reflect fairly simple circuit motifs, some of which (e.g., lateral inhibition) are

known to be ubiquitous in the brain. However, it is entirely unknown whether such circuits would support more computations on abstract representations as well.

Considering the computations as well as the instantiations of the mechanisms performing these computations will thus reveal a much more fine-grained picture of domain-generalty. Some mechanisms might be truly domain-specific, others truly domain-general, and yet others domain-bound. As a result, what might make a domain special relative to other domains, however, is the set of the available elemental computational operations, and how they can be combined (Barrett, 2012; Endress, Nespors, et al., 2009). Individual elemental computational mechanism can be specific to a single domain as well, but need not.

References

- Addessi, E., & Rossi, S. (2011). Tokens improve capuchin performance in the reverse-reward contingency task. *Proceedings. Biological Sciences*, 278, 849–854.
- Alberts, J. R., & Gubernick, D. J. (1984). Early learning as ontogenetic adaptation for ingestion by rats. *Learning and Motivation*, 15, 334–359.
- Alhama, R. G., & Zuidema, W. (2019). A review of computational models of basic rule learning: The neural-symbolic debate and beyond. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-019-01602-z>
- Allman, J. M., & Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (*Aotus trivirgatus*). *Brain Research*, 31, 85–105.
- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *J Exp Child Psychol*, 106, 20–29.
- Altmann, G. T. (2002). Learning and development in neural networks – the importance of prior experience. *Cognition*, 85, B43–50.
- Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends Cogn Sci*, 18, 279–292.
- Antell, S. E., Caron, A. J., & Myers, R. S. (1985). Perception of relational invariants by newborns. *Developmental Psychology*, 21, 942–948.
- Arena, P., Patané, L., Stornanti, V., Termini, P. S., Zäpf, B., & Strauss, R. (2013). Modeling the insect mushroom bodies: Application to a delayed match-to-sample task. *Neural Networks*, 41, 202–211.

- Aslin, R. N., & Newport, E. L. (2012). Statistical Learning. *Current Directions in Psychological Science*, 21, 170–176.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychol Sci*, 9, 321–324.
- Baddeley, A. D. (1996). The fractionation of working memory. *Proc Natl Acad Sci U S A*, 93, 13468–13472.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nat Rev Neurosci*, 4, 829–839.
- Banich, M. T., & Depue, B. E. (2015). Recent advances in understanding neural systems that support inhibitory control. *Current Opinion in Behavioral Sciences*, 1, 17–22.
- Banta Lavenex, P., Boujon, V., Ndarugendamwo, A., & Lavenex, P. (2015). Human short-term spatial memory: Precision predicts capacity. *Cognitive Psychology*, 77, 1–19.
- Barrett, H. C. (2005). Enzymatic Computation and Cognitive Modularity. *Mind & Language*, 20, 259–287.
- Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proc Natl Acad Sci USA*, 109 Suppl 1, 10733–10740.
- Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, 132, 137–150.
- Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience and Biobehavioral Reviews*, 77, 177–193.
- Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence

- from the obligatory contour principle. *Cognition*, 64, 39–72.
- Bjorklund, D. F., & Kipp, K. (1996). Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychological Bulletin*, 120, 163–188.
- Bonatti, L. L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cogn Psychol*, 44, 388–426.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychol Sci*, 16, 451–459.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychol Sci*, 18, 924–925.
- Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology. Animal Behavior Processes*, 21, 82–86.
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16, 106–113.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.
- Briscoe, S. D., & Ragsdale, C. W. (2018). Homology, neocortex, and the evolution of developmental mechanisms. *Science*, 362, 190–193.
- Brucks, D., Marshall-Pescini, S., Wallis, L. J., Huber, L., & Range, F. (2017). Measures of Dogs' Inhibitory Control Abilities Do Not Correlate across Tasks. *Frontiers in Psychology*, 8, 849.

- Brusatte, S. L., O'Connor, J. K., & Jarvis, E. D. (2015). The Origin and Diversification of Birds. *Current Biology: CB*, 25, R888-898.
- Bshary, R., & Grutter, A. S. (2002). Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, 63, 547–555.
- Bshary, R., & Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, 63, 557–564.
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54–115.
- Carroll, L. M., Jalbert, A., Penney, A. M., Neath, I., Surprenant, A. M., & Tehan, G. (2010). Evidence for proactive interference in the focus of attention of working memory. *Canadian Journal of Experimental Psychology*, 64, 208–214.
- Caruana, R. (1997). Multitask Learning. *Machine Learning*, 28, 41–75.
- Chakraborty, M., & Jarvis, E. D. (2015). Brain evolution by brain pathway duplication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370. <https://doi.org/10.1098/rstb.2015.0056>
- Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E. E., Dabelsteen, T., Pakkenberg, B., ... Jarvis, E. D. (2015). Core and Shell Song Systems Unique to the Parrot Brain. *PloS One*, 10, e0118496.
- Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M. T., & Munakata, Y. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. *PloS One*, 7, e31546.

- Chen, J., Jansen, N., & Ten Cate, C. (2016). Zebra finches are able to learn affixation-like patterns. *Animal Cognition*, *19*, 65–73.
- Chen, J., & Ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes*, *117*, 29–34.
- Chevalier, G., & Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in Neurosciences*, *13*, 277–280.
- Chomsky, N. (1980). *Rules and Representations*. Oxford: Basil Blackwell.
- Christiansen, M. H., & Curtin, S. (1999). Transfer of learning: Rule acquisition or statistical learning? *Trends Cogn Sci*, *3*, 289–290.
- Cipolotti, L., Spanò, B., Healy, C., Tudor-Sfetea, C., Chan, E., White, M., ... Bozzali, M. (2016). Inhibition processes are dissociable and lateralized in human prefrontal cortex. *Neuropsychologia*, *93*, 1–12.
- Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cogn Sci*, *7*, 547–552.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 24–39.
- Cope, A. J., Vasilaki, E., Minors, D., Sabo, C., Marshall, J. A. R., & Barron, A. B. (2018). Abstract concept learning in a simple neural network inspired by the insect brain. *PLoS Computational Biology*, *14*, e1006435.
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford, UK: Oxford University Press.
- Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral

- components of working memory storage. *Journal of Experimental Psychology. General*, 143, 1806–1836.
- Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, 137, 97–130.
- Danisman, E., Bshary, R., & Bergmüller, R. (2010). Do cleaner fish learn to feed against their preference in a reverse reward contingency task? *Anim Cogn*, 13, 41–49.
- Dawson, C., & Gerken, L. (2009). From domain-general to domain-sensitive: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, 111, 378–382.
- de la Mora, D. M., & Toro, J. M. (2013). Rule learning over consonants and vowels in a non-human animal. *Cognition*, 126, 307–312.
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56, 384–398.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., ... Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330, 1359–1364.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proc Natl Acad Sci U S A*, 105, 5909–5914.
- Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory.

- Proc Natl Acad Sci U S A*, 105, 5915–5920.
- Domjan, M. (1983). Biological Constraints on Instrumental and Classical Conditioning: Implications for General Process Theory. *Psychology of Learning and Motivation*, 17, 215–277.
- Domjan, M. (2015). The Garcia-Koelling Selective Association Effect: A Historical and Personal Perspective. *International Journal of Comparative Psychology*, 28.
- Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *Proceedings of the International Conference on Machine Learning (ICML)*. Presented at the Beijing, China. Retrieved from <http://arxiv.org/abs/1310.1531>
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14, 172–179.
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23, 475–483.
- Dunlap, A. S., & Stephens, D. W. (2014). Experimental evolution of prepared learning. *Proceedings of the National Academy of Sciences*, 111, 11750–11755.
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, 8, 1784–1790.

- Endress, A. D. (under review). *A simple, biologically plausible feature detector for language acquisition.*
- Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Act Psychol*, 135, 182–190.
- Endress, A. D. (2013). Bayesian learning and the psychology of rule induction. *Cognition*, 127, 159–176.
- Endress, A. D. (2014). How are Bayesian models really used? *Cognition*, 130, 81–84.
- Endress, A. D., Cahill, D., Block, S., Watumull, J., & Hauser, M. D. (2009). Evidence of an evolutionary precursor to human language affixation in a nonhuman primate. *Biol Lett*, 5, 749–751.
- Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition*, 13, 483–495.
- Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105, 577–614.
- Endress, A. D., & Hauser, M. D. (2009). Syntax-induced pattern deafness. *Proc Natl Acad Sci USA*, 106, 21001–21006.
- Endress, A. D., & Hauser, M. D. (2011). The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *J Exp Psychol Learn Mem Cogn*, 37, 77–95.
- Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *Q J Exp Psychol*, 62, 2187–2209.

- Endress, A. D., & Mehler, J. (2010). Perceptual Constraints in Phonotactic Learning. *J Exp Psychol Hum Percept Perform*, 36, 235–250.
- Endress, A. D., Nespors, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends Cogn Sci*, 13, 348–353.
- Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *J Exp Psychol Gen*, 143, 548–65.
- Endress, A. D., & Szabó, S. (2017). Interference and memory capacity limitations. *Psychol Rev*, 124, 551–571.
- Endress, A. D., & Wood, J. N. (2011). From movements to actions: Two mechanisms for learning action sequences. *Cognit Psychol*, 63, 141–171.
- Engel, T. A., & Wang, X.-J. (2011). Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *Journal of Neuroscience*, 31, 6982–6996.
- Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Curr Dir Psychol Sci*, 11, 19–23.
- Erika-Florence, M., Leech, R., & Hampshire, A. (2014). A functional network perspective on response inhibition and attentional control. *Nature Communications*, 5, 4073.
- Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 16616–16621.
- Fischer-Baum, S., Charny, J., & McCloskey, M. (2011). Both-edges representation of letter position in reading. *Psychonomic Bulletin & Review*, 18, 1083–1089.

- Fischer-Baum, S., & McCloskey, M. (2015). Representation of item position in immediate serial recall: Evidence from intrusion errors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*, 1426–1446.
- Fischer-Baum, S., McCloskey, M., & Rapp, B. (2010). Representation of letter position in spelling: Evidence from acquired dysgraphia. *Cognition*, *115*, 466–490.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn*, *28*, 458–467.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *J Exp Psychol Gen*, *144*, 30–47.
- Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Dev Sci*, *12*, 504–509.
- Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition*, *120*, 360–371.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology. General*, *133*, 101–135.
- Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204.

- Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity Avoidance and the OCP. *Nat Lang Ling Theory*, 22, 179–228.
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19, 117–125.
- Fu, Y., Kaneko, M., Tang, Y., Alvarez-Buylla, A., & Stryker, M. P. (2015). A cortical disinhibitory circuit for enhancing adult plasticity. *ELife*, 4, e05558.
- Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., ... Stryker, M. P. (2014). A cortical circuit for gain control by behavioral state. *Cell*, 156, 1139–1152.
- Gallistel, C. R. (2017). The Coding Question. *Trends in Cognitive Sciences*, 21, 498–508.
- Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1974). Behavioral regulation of the milieu interne in man and rat. *Science*, 185, 824–831.
- Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1976). Flavor aversion studies. *Science*, 192, 265–267.
- Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4, 123–124.
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18, 1–16.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for

- cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3, 191–197.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nat Neurosci*, 2, 568–573.
- Gemberling, G. A., & Domjan, M. (1982). Selective associations in one-day-old rats: Taste-toxicosis and texture-shock aversion learning. *Journal of Comparative and Physiological Psychology*, 96, 105–113.
- Gemberling, G. A., Domjan, M., & Amsel, A. (1980). Aversion learning in 5-day-old rats: Taste-toxicosis and texture-shock associations. *Journal of Comparative and Physiological Psychology*, 94, 734–745.
- Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, 1–11.
- Gervain, J., Macagno, F., Cogoï, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proc Natl Acad Sci U S A*, 105, 14222–14227.
- Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of “sameness” and “difference” in an insect. *Nature*, 410, 930–933.
- Goddard, C. A., Mysore, S. P., Bryant, A. S., Huguenard, J. R., & Knudsen, E. I. (2014). Spatially reciprocal inhibition of inhibition within a stimulus selection network in the avian midbrain. *PloS One*, 9, e85865.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of

- form. *Paleobiology*, 8, 4–15.
- Grothe, B. (2018). How the Barn Owl Computes Auditory Space. *Trends in Neurosciences*, 41, 115–117.
- Grutter, A. S., & Bshary, R. (2003). Cleaner wrasse prefer client mucus: Support for partner control mechanisms in cleaning interactions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270, S242–S244.
- Gubernick, D. J., & Alberts, J. R. (1984). A specialization of taste aversion learning during suckling and its weaning-associated transformation. *Developmental Psychobiology*, 17, 613–628.
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455, 665–668.
- Hale, M. E., Kheirbek, M. A., Schriefer, J. E., & Prince, V. E. (2004). Hox Gene Misexpression and Cell-Specific Lesions Reveal Functionality of Homeotically Transformed Neurons. *Journal of Neuroscience*, 24, 3070–3076.
- Hampshire, A., & Sharp, D. J. (2015). Contrasting network and modular perspectives on inhibitory control. *Trends in Cognitive Sciences*, 19, 445–452.
- Hangya, B., Pi, H.-J., Kvitsiani, D., Ranade, S. P., & Kepecs, A. (2014). From circuit motifs to computations: Mapping the behavioral repertoire of cortical interneurons. *Current Opinion in Neurobiology*, 26, 117–124.
- Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S.

- (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends in Cognitive Sciences*.
<https://doi.org/10.1016/j.tics.2015.05.006>
- Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science*, *14*, 865–872.
- Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proc Natl Acad Sci U S A*, *102*, 12639–12643.
- Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, *121*, 400–415.
- Hasselmo, M. E., & Wyble, B. P. (1997). Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behavioural Brain Research*, *89*, 1–34.
- Hauser, M. D., & Glynn, D. (2009). Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? *J Comp Psychol*, *123*, 161–167.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53-64.
- He, J. (2001). On and off pathways segregated at the auditory thalamus of the guinea pig. *J Neurosci*, *21*, 8672–8679.
- Henson, R. N. (1998). Short-term memory for serial order: The Start-End Model.

Cognit Psychol, 36, 73–137.

Hirschfeld, L. A., & Gelman, S. A. (1994). Toward a topography of mind: An introduction to domain specificity. In L. A. Hirschfeld & S. A. E. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 3–36). Cambridge University Press.

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Dev Sci*, 14, 1445–1458.

Hosseini-Kamkar, N., & Morton, J. B. (2014). Sex differences in self-regulation: An evolutionary perspective. *Frontiers in Neuroscience*, 8, 233.

Hupé, J.-M. (2017). Comment on “Ducklings imprint on the relational concept of ‘same or different.’” *Science*, 355, 806.

Hurley, I., Hale, M. E., & Prince, V. E. (2005). Duplication events and the evolution of segmental identity. *Evol Dev*, 7, 556–567.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology*, 387, 167–178.

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: Classifying and distinguishing between models. *Nature Reviews Genetics*, 11, 97–108.

Jacobs, G. H. (2009). Evolution of colour vision in mammals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364, 2957–2967.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic

- neural field model of visual working memory and change detection. *Psychological Science*, 20, 568–577.
- Jonker, T. R., Seli, P., & MacLeod, C. M. (2015). Retrieval-Induced Forgetting and Context. *Current Directions in Psychological Science*, 24, 273–278.
- Jovanic, T., Schneider-Mizell, C. M., Shao, M., Masson, J.-B., Denisov, G., Fetter, R. D., ... Zlatić, M. (2016). Competitive Disinhibition Mediates Behavioral Choice and Sequences in *Drosophila*. *Cell*, 167, 858–870.e19.
- Kaas, J. H. (1989). The evolution of complex sensory systems in mammals. *The Journal of Experimental Biology*, 146, 165–176.
- Kabdebon, C., & Dehaene-Lambertz, G. (2019). Symbolic labeling in 5-month-old human infants. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 5805–5810.
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *J Exp Psychol Gen*, 130, 169–183.
- Kane, Michael J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18, 614–621.
- Kane, Michael J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *J Exp Psychol Gen*, 132, 47–70.
- Kane, Michael J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working

- memory capacity and the top-down control of visual search: Exploring the boundaries of “executive attention”. *J Exp Psychol Learn Mem Cogn*, 32, 749–777.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17, 4302–4311.
- Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does grammar constrain statistical learning? Commentary on Bonatti, Peña, Nespor, and Mehler (2005). *Psychol Sci*, 18, 922–923.
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, 18, 1084–1089.
- Kovács, Á. M., & Endress, A. D. (2014). Hierarchical processing in seven-month-old infants. *Infancy*, 9, 409–425.
- Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, 325, 611–612.
- Koyama, M., Minale, F., Shum, J., Nishimura, N., Schaffer, C. B., & Fetcho, J. R. (2016). A circuit motif in the zebrafish hindbrain for a two alternative behavioral choice to turn left or right. *ELife*, 5. <https://doi.org/10.7554/eLife.16808>
- Koyama, M., & Pujala, A. (2018). Mutual inhibition of lateral inhibition: A network motif for an elementary computation in the brain. *Current Opinion in Neurobiology*, 49, 69–74.
- Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2010). Target

- detection by opponent coding in monkey prefrontal cortex. *Journal of Cognitive Neuroscience*, 22, 751–760.
- Langbein, J., & Puppe, B. (2017). Comment on “Ducklings imprint on the relational concept of ‘same or different.’” *Science*, 355, 806.
- Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., & Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nature Neuroscience*, 16, 1662–1670.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley and Sons.
- Letzkus, J. J., Wolff, S. B. E., Meyer, E. M. M., Tovote, P., Courtin, J., Herry, C., & Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature*, 480, 331–335.
- Liu, K. S., Gray, M., Otto, S. J., Fetcho, J. R., & Beattie, C. E. (2003). Mutations in deadly seven/notch1a Reveal Developmental Plasticity in the Escape Response Circuit. *Journal of Neuroscience*, 23, 8159–8166.
- Longe, O., Randall, B., Stamatakis, E. A., & Tyler, L. K. (2007). Grammatical categories in the brain: The role of morphological structure. *Cereb Cortex*, 17, 1812–1820.
- Ludueña, G. A., & Gros, C. (2013). A self-organized neural comparator. *Neural Computation*, 25, 1006–1028.
- Lynch, M. A. (2004). Long-term potentiation and memory. *Physiological Reviews*, 84, 87–136.
- Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154, 459–473.

- Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science*, 307, 1121–1124.
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, 92, 155–161.
- Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: An embarrassment of riches. *Neuron*, 44, 5–21.
- Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, 67, 130–150.
- Marcus, G. F. (2004). *The birth of the mind: How a tiny number of genes creates the complexities of human thought*. Basic Books.
- Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychol Sci*, 18, 387–391.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80.
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews. Neuroscience*, 5, 793–807.
- Marlin, B. J., Mitre, M., D'amour, J. A., Chao, M. V., & Froemke, R. C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature*, 520, 499–504.
- Marr, D., & Nishihara, H. K. (1992). Visual information processing: Artificial intelligence and the sensorium of sight. In S. M. Kosslyn & R. A.

- Andersen (Eds.), *Frontiers in Cognitive Neuroscience* (pp. 165–186).
Cambridge, MA: MIT Press.
- Martin, L. T., & Alberts, J. R. (1979). Taste aversions to mother's milk: The age-related role of nursing in acquisition and expression of a learned association. *Journal of Comparative and Physiological Psychology*, 93, 430–445.
- Martinho, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept of “same or different.” *Science*, 353, 286–288.
- McCarthy, J. J. (1979). *Formal Problems in Semitic Phonology and Morphology* (Doctoral dissertation, MIT). MIT, Cambridge, MA.
- McCarthy, J. J. (1982). Prosodic structure and expletive infixation. *Language*, 48, 574–590.
- McCarthy, J. J., & Prince, A. (1993). Generalized alignment. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1993* (pp. 79–153). Boston, MA: Kluwer.
- McCarthy, J. J., & Prince, A. (1999). Faithfulness and identity in prosodic morphology. In R. Kager, H. van der Hulst, & W. Zonneveld (Eds.), *The prosody morphology interface* (pp. 218–309). Cambridge: Cambridge University Press.
- McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends Cogn Sci*, 3, 166–168.
- McClelland, J. L., Rumelhart, D. E., & Group, T. P. R. (Eds.). (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- Miller, P., & Wang, X.-J. (2006). Inhibitory control by an integral feedback signal

- in prefrontal cortex: A model of discrimination between sequential stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 201–206.
- Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, 244, 933–938.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *J Child Lang*, 37, 545–564.
- Moravcsik, E. (1978). Reduplicative constructions. In J. H. Greenberg (Ed.), *Universals of human language: Word structure* (Vol. 3, pp. 297–334). Stanford, CA: Stanford University Press.
- Moro, A. (2008). *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. Cambridge, MA: MIT Press.
- Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, 15, 453–459.
- Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule Learning by Rats. *Science*, 319, 1849–1851.
- Murray, E. A., Kralik, J. D., & Wise, S. P. (2005). Learning to inhibit prepotent responses: Successful performance by rhesus macaques, *Macaca mulatta*,

- on the reversed-contingency task. *Animal Behaviour*, 69, 991–998.
- Mysore, S. P., & Knudsen, E. I. (2012). Reciprocal inhibition of inhibition: A circuit motif for flexible categorization in stimulus selection. *Neuron*, 73, 193–205.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Neiworth, J. J. (2013). Chasing sounds. *Behav Processes*, 93, 111–115.
- Nelson, R., & Kolb, H. (2003). ON and OFF pathways in the vertebrate retina and visual system. In L. M. Chalupa & J. S. Werner (Eds.), *The visual Neurosciences* (Vol. 1, pp. 260–278). Cambridge, MA: MIT Press.
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue & Linguaggio*, 203–231.
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Foris: Dordrecht.
- Nothwang, H. G. (2016). Evolution of mammalian sound localization circuits: A developmental perspective. *Progress in Neurobiology*, 141, 1–24.
- Nozari, N., & Novick, J. (2017). Monitoring and Control in Language Production. *Current Directions in Psychological Science*, 26, 403–410.
- Ohno, S. (1970). *Evolution by Gene Duplication*. New York, NY: Springer.
- Owen, S. F., Tuncdemir, S. N., Bader, P. L., Tirko, N. N., Fishell, G., & Tsien, R. W. (2013). Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature*, 500, 458–462.
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296, 1321–1323.

- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proc Natl Acad Sci U S A*, *102*, 5297–5300.
- Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): Learning with respect to categories of color, shape, and material. *Anim Learn Behav*, *15*, 421–32.
- Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. (2013). Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nature Neuroscience*, *16*, 1068–1076.
- Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., & Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, *503*, 521–524.
- Pinker, S. (1998). *How the mind works*. London, UK: Allen Lane. Penguin.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology. Human Perception and Performance*, *20*, 421–435.
- Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, *116*, 361–367.
- Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The Human Fetus Preferentially Engages with Face-like Visual Stimuli. *Current Biology: CB*, *27*, 1825-1828.e3.
- Rubino, C. (2013). Reduplication. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*, 669–680.
- Samuels, R. (2000). Massively modular minds: Evolutionary psychology and cognitive architecture. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Modularity, Language and Meta-Cognition* (pp. 13–46). Cambridge University Press.
- Samuels, R. (2012). Massive Modularity. In E. Margolis, R. Samuels, & S. Stich (Eds.), *Oxford Handbook of Philosophy of Cognitive Science*. Oxford, UK: Oxford University Press.
- Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & Schonen, S. de. (2005). Reversibility of the Other-Race Effect in Face Recognition During Childhood. *Psychological Science*, *16*, 440–444.
- Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., ... Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nature Neuroscience*, *19*, 1250–1255.
- Scholl, B., Gao, X., & Wehr, M. (2010). Nonoverlapping Sets of Synapses Drive On Responses and Off Responses in Auditory Cortex. *Neuron*, *65*, 412–

421.

- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599–1603.
- Seidenberg, M. S., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science*, 284, 433.
- Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, 9, 565–573.
- Seidl, A., & Johnson, E. K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *J Child Lang*, 35, 1–24.
- Seligman, M. E. (1970). On the generality of the laws of learning. *Psychol Rev*, 77, 406–418.
- Shifferman, E. M. (2009). Its own reward: Lessons to be drawn from the reversed-reward contingency paradigm. *Animal Cognition*, 12, 547–558.
- Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, 40, 605–619.
- Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 39, 277–289.
- Shor, R. E. (1970). The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, 32, 346–365.

- Shukla, M. (2005). Language from a biological perspective. *J Biosci*, 30, 119–127.
- Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognit Psychol*, 54, 1–32.
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc Natl Acad Sci U S A*, 108, 6038–6043.
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, 177, 198–213.
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105–120.
- Silverman, I. W. (2003a). Gender Differences in Delay of Gratification: A Meta-Analysis. *Sex Roles*, 49, 451–463.
- Silverman, I. W. (2003b). Gender differences in resistance to temptation: Theories and evidence. *Developmental Review*, 23, 219–259.
- Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, 25, 256–260.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 39–67). Cambridge University Press.

- Spierings, M. J., & ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings of the National Academy of Sciences*, *113*, E3977–E3984.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol*, *18*, 643–662.
- Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 394–398.
- Surridge, A. K., Osorio, D., & Mundy, N. I. (2003). Evolution and selection of trichromatic vision in primates. *Trends in Ecology & Evolution*, *18*, 198–205.
- Taylor, J. S., & Raes, J. (2004). Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics*, *38*, 615–643.
- Tompson, S. H., Kahn, A. E., Falk, E. B., Vettel, J. M., & Bassett, D. S. (2019). Individual differences in learning social and nonsocial network structures. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*, 253–271.
- Toro, J. M., Bonatti, L. L., Nespor, M., & Mehler, J. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychol Sci*, *19*, 137–144.
- Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Percept Psychophys*, *70*, 1515–1525.

- Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept Psychophys*, 67, 867–875.
- Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *J Exp Psychol Gen*, 134, 552–564.
- Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *J Exp Psychol: Hum Perc Perf*, 35, 195–202.
- van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E4501–E4510.
- van Der Velde, F., & de Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience*, 13, 479–491.
- van Heijningen, C. A. A., Chen, J., van Laatum, I., van der Hulst, B., & Ten Cate, C. (2013). Rule learning by zebra finches in an artificial grammar learning task: Which rule? *Anim Cogn*, 16, 165–175.
- Versace, E., Spierings, M. J., Caffini, M., Ten Cate, C., & Vallortigara, G. (2017). Spontaneous generalization of abstract multimodal patterns in young domestic chicks. *Animal Cognition*, 20, 521–529.
- Vlamings, P. H. J. M., Uher, J., & Call, J. (2006). How the great apes (Pan troglodytes, Pongo pygmaeus, Pan paniscus, and Gorilla gorilla) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology. Animal Behavior*

Processes, 32, 60–70.

- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316, 1159.
- Wen, S., Ulloa, A., Husain, F., Horwitz, B., & Contreras-Vidal, J. L. (2008). Simulated neural dynamics of decision-making in an auditory delayed match-to-sample task. *Biological Cybernetics*, 99, 15–27.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.
- Wetzels, W. L., & Mascaró, J. (2001). The Typology of Voicing and Devoicing. *Language*, 77, 207–244.
- Windes, J. D. (1968). Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology*, 78, 318–322.
- Wong, J. H., Peterson, M. S., & Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition*, 108, 719–731.
- Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, 108, 522–532.
- Wood, J. N. (2009). Distinct visual working memory systems for view-dependent and view-invariant representation. *PLoS One*, 4, e6601.
- Xu, H., Jeong, H.-Y., Tremblay, R., & Rudy, B. (2013). Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron*, 77, 155–167.

- Yamazaki, Y., Suzuki, K., Inada, M., Iriki, A., & Okanoya, K. (2012). Sequential learning and rule abstraction in Bengalese finches. *Animal Cognition*, *15*, 369–377.
- Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology & Evolution*, *18*, 292–298.
- Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., ... Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science*, *345*, 660–665.
- Zhao, W., Zhou, P., Gong, C., Ouyang, Z., Wang, J., Zheng, N., & Gong, Z. (2019). A disinhibitory mechanism biases *Drosophila* innate light preference. *Nature Communications*, *10*, 124.

Acknowledgements

I thank M. D. Hauser, Á. Kovács, M. Ramscar, M. Shukla and three anonymous reviewers for helpful comments on an earlier draft of this manuscript.