Over the past two decades, the human ability to pick up on patterns in sensory input has been investigated in a large number of studies. Many of those were inspired by Saffran, Aslin, and Newport’s (1996) seminal work on statistical learning (SL), which revealed that infants can extract syllabic patterns presented in a continuous stream based solely on transitional probabilities between elements (for a description of the tasks, see Box 1).

Research on the learning of regularities was prevalent in the implicit-learning literature before the introduction of SL as a theoretical construct (for discussions, see Christiansen, 2019; Perruchet & Pacton, 2006). This research predominantly employed the task of artificial grammar learning (AGL) offered by Reber (1967). In this task, participants are presented with sequences of stimuli generated by a miniature grammar and are then asked to decide whether novel sequences are grammatical (e.g., Altmann, Dienes, & Goode, 1995; see Box 1). Although AGL performance was originally taken to tap the implicit learning of abstract grammar rules, current theories maintain that it can also be explained by judgments of statistically related surface similarity between test items and the items—or fragments of the items—presented during familiarization (e.g., Conway & Christiansen, 2005; Knowlton & Squire, 1996; for a review, see Pothos, 2007). A second task commonly used to tap implicit learning of regularities was the serial-reaction-time (SRT) task (Nissen & Bullemer, 1987; see Box 1). In the SRT task, participants’ response times evidence their implicit learning of repeated perceptuomotor sequences (e.g., Robertson, 2007; Squire & Zola, 1996).

Experimental work in the domain of implicit learning has primarily focused on the mechanisms that underlie acquiring knowledge without awareness and the role of consciousness in the learning process and its outcomes (e.g., French & Cleeremans, 2002).
Statistical Regularities in Spoken and Written Language

The main promise of SL theory was to provide an alternative to domain-specific and innate theories of language acquisition (Chomsky, 1965; Fodor, 1983). SL-based accounts of language focus on experienced-based learning mechanisms that enable the acquisition of linguistic regularities (e.g., Adriaans & Kager, 2010; Erickson & Thiessen, 2015; Safran & Thiessen, 2007). Natural languages can indeed be described in terms of a rich array of statistical regularities and quasiregularities. For example, in any given language, spoken words are characterized by language-specific phonotactic constraints, which result in characteristic phonemic and phonetic sequences. Native speakers can implicitly assimilate the patterns of conditional probabilities of phonemes in their language, perceive word boundaries in the absence of pauses between words, and develop expectations regarding incoming speech segments (e.g., Adriaans & Kager, 2010; Safran & Thiessen, 2007). Likewise, the acquisition of syntax can be thought of as learning a set of regularities regarding how words are aligned in propositions (e.g., Safran & Wilson, 2003; Thompson & Newport, 2007).

Considering written language, orthographies can be characterized by a set of correlations that determine the possible co-occurrences of graphemes or letter sequences, such as bigrams, trigrams, prefixes and suffixes, double letters, and so on. For example, beginner readers of English have been shown to judge baff as more word-like than bbaf, demonstrating sensitivity to the legal position of double consonants in their written language (Cassar & Treiman, 1997; see also Chetail, 2017; Chetail, Balota, Treiman, & Content, 2015; Pacton, Perruchet, Fayol, & Cleeremans, 2001). In addition, writing systems are characterized by high or low correlations in grapheme-to-phoneme mappings (e.g., Apfelbaum, Hazeltine, & McMurray, 2013; Ziegler et al., 2010), so that proficient reading involves the acquisition

Box 1. Glossary of the experimental paradigms

**Statistical learning embedded-pattern task (SL-EPT).** This task involves a continuous visual or auditory familiarization stream that comprises embedded patterns (with high transitional probabilities between the patterns’ constituents and lower transitional probabilities between patterns’ boundaries). Familiarization is followed by a test phase in which the preference for the embedded patterns over foil patterns is assessed.

**Artificial-grammar-learning (AGL) task.** In a typical AGL experiment, participants are exposed to sequences generated by a miniature grammar. After the exposure phase, participants’ learning of the underlying structure of the grammar is measured by asking them to classify the grammaticality of a new set of sequences, some of which follow the grammar and some of which do not.

**Serial-reaction-time (SRT) task.** In this task participants are presented with sequences of visual stimuli appearing sequentially in several locations on the screen and are required to press a key corresponding to the relevant location as quickly as possible. In structured blocks, the serial order of locations is fully or probabilistically determined and can be learned implicitly by participants, as indicated by increasingly faster response times. Typically, response times for structured sequences are compared with response times for random sequences, reflecting the extent of learning.
of probabilistic cues regarding how a given grapheme is pronounced given its location within the word and its neighboring letters (e.g., Ziegler & Goswami, 2005). Finally, languages exhibit systematic relations between form and semantic meaning through morphological structure, allowing the fast processing and decomposition of morphologically complex words (e.g., Rastle, Davis, & New, 2004). All these regularities can, in principle, be thought of as the target of continuous SL in the process of literacy acquisition (see also Bogaerts, Christiansen, & Frost, 2020; Ellemans, Steacy, & Compton, 2019; Sawi & Rueckl, 2019).

### Individual Variation in SL and Linguistic Abilities

The theoretical approach outlined above makes a clear prediction regarding individual differences: If language learning involves the assimilation of regularities, then individuals with better SL capacities should also show better linguistic skills (for a discussion, see Siegelman, 2020). A number of recent studies have indeed reported evidence confirming this prediction. For example, performance in an auditory SL task was found to correlate with predictors of early literacy skills (Spencer, Kaschak, Jones, & Lonigan, 2015), overall reading performance (Qi, Sanchez Araujo, Georgan, Gabrieli, & Arciuli, 2019), and children's lexical abilities (J. L. Evans, Saffran, & Robe-Torres, 2009; Mainela-Arnold & Evans, 2014). In the same vein, significant (albeit small) correlations were reported between the ability of both school-age children and adults to detect dependencies in a sequence of visual stimuli and their reading abilities (Arciuli & Simpson, 2012; Torkildsen, Arciuli, & Wie, 2019; but for contrasting results, see Schmalz, Moll, Mulatti, & Schulte-Körne, 2019). Testing adult second-language learners, individuals with better visual SL performance were found to better assimilate literacy in Hebrew (Frost, Siegelman, Narkiss, & Afek, 2013). Similar correlations have been reported using the AGL task (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2012; but see Pavlidou & Bogaerts, 2019).

### SL and Language Impairments

This brings us to the main thrust of the present article. The hypothesized link between SL and linguistic abilities has prompted a parallel investigation of the theoretical link between SL and language difficulties, mainly developmental dyslexia (DD) and specific language impairment (SLI; see Arciuli & Simpson, 2012; Lammertink, Boersma, Wijnen, & Rispens, 2017; Schmalz, Altoè, & Mulatti, 2017). The logic seems straightforward: If individual capacities in SL underlie language learning because language presents an array of statistical regularities that must be learned for proficient language use, then atypical SL capacities would lead to atypical language learning. In the abstract of a recent review, Saffran (2018) wrote,

> While a full picture is not yet available, results [of studies using SL tasks with different groups of children with developmental disabilities] to date suggest that studies of [statistical] learning are both feasible and informative about learning processes that may differ across diagnostic groups, particularly as they relate to language acquisition. (para. 2)

In essence, Saffran is mirroring the argument above: If SL computations are part of a theory of language learning, they should also be part of a theory that explains developmental deficits in language learning or processing. Although this assertion can hardly be contested, in practice, a range of different rationales for investigating SL in these special populations has been offered. These rationales involve several theories regarding underlying deficits, linking them to SL via a range of theoretical constructs in the domain of learning.

In the DD and SLI literature, researchers have often associated the theoretical construct of SL with deficits in some procedural or implicit learning tasks, or both (e.g., SRT task, AGL task; for description of tasks, see Box 1). For example, Gabay, Thiessen, and Holt’s (2015) starting point is that a general procedural-learning deficit underlies DD (see Ullman, 2004). In the view of Gabay et al., SL abilities similarly “draw upon procedural learning systems” (p. 934). This leads to the hypothesis that people with DD should show deficits in SL tasks. Kahta and Schiff (2019) viewed SL to be synonymous with implicit sequential learning, which is considered to be impaired in DD: “the core deficit in DD is poor implicit sequential learning processes, also known as poor statistical learning processes” (p. 143). They therefore hypothesized that DD populations should be impaired in SL tasks such as the auditory embedded-pattern task. Van Witteloostuijn, Boersma, Wijnen, and Rispens (2019) stated, “it has been suggested that dyslexia is associated with a domain-general learning deficit rather than a deficit that is specific to the processing of phonological material. This domain-general learning mechanism is often referred to as statistical learning [emphasis in original]” (para. 3). They therefore tested children with DD across a range of SL tasks and argued that “the hypothesized SL deficit has been claimed to be independent of the domain and modality in which SL is tested” (para. 4). A very different theoretical approach was provided by Sigurdardottir, Fridriksdottir, Gudjonsdottir, and Kristjánsson (2018). They argued that readers with DD show hypoaivation in ventral visual stream, which supports word, object,
and face recognition. Because these brain regions are shaped by visual SL, dyslexics should show deficits in visual SL.

Considering SLI, J. L. Evans et al.'s (2009) starting point is that SLI is related to deficits in implicit learning abilities. They defined SL as an instantiation of implicit learning: “a paradigmatic measure of implicit learning during infancy and childhood is statistical learning—the tracking patterns of regularities over input such as syllables, tones, or shapes” (p. 323). This leads to the hypothesis that children with SLI will show deficits in SL tasks such as the embedded-pattern task. Haebig, Saffran, and Ellis Weismer (2017) derived their prediction of impaired SL performance in children with SLI from the procedural-learning-deficit hypothesis of SLI (Ullman & Pierpont, 2005). They wrote, “If atypical language development in ASD [autism spectrum disorder] and SLI is derived from underlying impairments in statistical learning as proposed by the PDH [procedural-learning-deficit hypothesis], children in both diagnostic groups should demonstrate poor statistical learning” (p. 1253). These authors related procedural learning to SL by mentioning that “both fit under the umbrella of implicit learning” (p. 1252). As a final example, in a recent article, Lamertink and colleagues (2020) directly connected the diagnosis of developmental language disorder and SLI is derived from underlying impairments in statistical learning as proposed by the PDH; procedural-learning-deficit hypothesis, children in both diagnostic groups should demonstrate poor statistical learning” (p. 1253). These authors related procedural learning to SL by mentioning that “both fit under the umbrella of implicit learning” (p. 1252).

Our aim in this brief review is not to offer criticism of one rationale or another but rather to critically discuss how very different theoretical accounts and sequences of arguments end up with identical predictions about a difference in SL performance between special as opposed to control populations. This should be a cause for concern because with this state of affairs, empirical findings demonstrating the hypothesized group effect do not lead to favoring one theory over another. Taking the examples above, for any given sample of participants showing reduced SL performance, what should be concluded regarding the underlying deficit that has led to this result? Does it reflect problems in procedural learning? Implicit learning? A general deficit in the ventral visual stream? Let us be clear, we do not take issue here with the fact that there are different theories regarding the origin of DD or SLI. This state of affairs has been discussed in length in the DD literature (e.g., for discussion, see Ramus & Ahissar, 2012). The problem we see here is that a range of different theories are offered, but they all propose identical predictions (i.e., impaired SL performance). This leaves us with no way to empirically distinguish the different theoretical explanations.

The Problem of Underspecification

Our main argument in this article is that present studies tying SL to reading disabilities or SLI are (a) vague regarding the demarcation lines between the different theoretical constructs in the domain of learning (e.g., “motor procedural learning,” “nonmotor procedural learning,” “implicit learning,” “sequence learning,” “statistical learning”), (b) vague about the mapping of the experimental tasks they use to the theoretical constructs that these tasks are supposed to tap, and (c) vague about how these theoretical constructs are linked to language difficulties.

Figure 1 illustrates our take on the different theoretical constructs of cognitive faculties in the literature, their interrelations, and the mapping between experimental tasks and the faculties they are supposed to tap. Admittedly, the exact architecture reflects one possible theoretical analysis of the literature, and others may draw other architectures (see Krishnan, Watkins, & Bishop, 2016). Nevertheless, a constructive discussion of how SL is related to language deficits must outline at the outset an initial set of hypotheses regarding the faculties that underlie the relevant deficits.

Our review of the studies above shows that there is little agreement regarding how SL is related to the cognitive faculties in the domain of learning. To some, SL is entirely captured in the construct of implicit learning; to others, it is contained in the overlap between implicit and sequence learning; and to still others, it is a subset of procedural learning. This exemplifies the current ambiguity on the level of the theoretical constructs: The demarcation lines between them are not well specified. In addition, regardless of the exact architecture of cognitive faculties, the mapping between experimental tasks and the theoretical constructs they are supposed to tap shows that different tasks are used as proxies of the same construct and that a given task is taken to be a proxy of more than one theoretical construct. In the following sections, we expand on these issues and exemplify how they hinder scientific advances in understanding whether and precisely how SL deficits contribute to specific difficulties with reading or spelling and more general language disabilities. We start with a historical overview in which we aim to pinpoint the sources of ambiguities and vagueness at all these levels.

A brief history: What is impaired in developmental dyslexia and specific language impairment?

DD is commonly defined as a learning disorder characterized by persistent difficulties with reading or spelling, or both, in the presence of normal intelligence and typical educational opportunities (e.g., American Psychiatric
The impairments of individuals with DD are specific to or at minimum most pronounced in the domain of literacy. This stands in contrast to the profile of individuals diagnosed with SLI, who also have—despite normal development in all other areas—problems in learning to talk and display general semantic and syntactic deficits (e.g., Bishop, 2006; Bishop & Snowling, 2004).

The underpinning factors of DD and SLI have been the focus of extensive and heated debates (for reviews, see Bishop, 2006; Ramus, 2003; Ramus & Ahissar, 2012). Most research on DD has focused on language-specific factors (e.g., impaired phonological representations, Snowling, 2000; problematic phonological access and retrieval, Boets et al., 2013; Ramus & Szenkovits, 2008; or impairments in a specific sensory domain such as audition, e.g., Goswami, 2011, or vision, Bosse, Tainturier, & Valdois, 2007). Yet other research efforts, increasingly more in recent years, have focused on domain-general factors related to the classical taxonomy of cognitive faculties such as attention (Hari & Renvall, 2001), memory (Perez, Majerus, Mahot, & Poncelet, 2012; Smith-Spark & Fisk, 2007), and—perhaps of greatest interest for the current article—procedural learning (Nicolson & Fawcett, 1990, 2007, 2011; Ullman, 2004) define the deficit more broadly as a deficit in skill and habit learning by the procedural memory system (see Box 2). Regarding the underlying causes of SLI, language-specific theories have largely focused on grammatical impairments within a generative framework (e.g., Clahsen, Bartke, & Göllner, 1997; van der Lely & Battell, 2003), whereas domain-general accounts have focused on impairments such as working memory (for a review, see Henry & Botting, 2017,) or, again, procedural learning (Ullman & Pierpont, 2005).

In support of the procedural-learning-deficit hypothesis, individuals with DD and children with SLI have been shown to be impaired also in the automatization of balance (Hill, 2001; Nicolson & Fawcett, 1990), in numerous motor tasks including limb coordination and mirror drawing (Fawcett et al., 1996; Hill, 2001; Vicari et al., 2005), and in the SRT task that involves motor responses (for meta-analyses, see Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Lum, Ullman, & Conti-Ramsden, 2013).

This is where the aforementioned vagueness regarding the demarcation lines between the different theoretical constructs and the mapping between tasks and constructs comes into play: The SRT task is also
prevalent in the field of implicit learning, where it is
taken to tap implicit forms of learning. This has led to
a shift in focus from a *motor procedural-learning deficit*
in DD to an *implicit sequence-learning impairment*
(e.g., Jiménez-Fernández, Vaquero, Jiménez, & Defior,
2011; Vicari, Marotta, Menghini, Molinari, & Petrosini,
2003). Indeed, by now authors of a large body of stud-
ies have used the SRT task to tap both procedural and
implicit-learning abilities in DD and SLI populations,
arguing for a procedural-learning deficit or an implicit-
learning deficit (e.g., Jiménez-Fernández et al., 2011;
Menghini, Hagberg, Caltagirone, Petrosini, & Vicari,
2006; Stoodley, Harrison, & Stein, 2006; Stoodley, Ray,
Jack, & Stein, 2008; Tomblin, Mainela-Arnold, & Zhang,
2007; Vicari et al., 2005). To complicate things further,
numerous researchers have employed the AGL task to
compare performance of DD or SLI samples with that
of control participants to test the hypothesis of an
implicit-learning deficit (e.g., Kahta & Schiff, 2016; Nigro,
Jiménez-Fernández, Simpson, & Defior, 2016; Pavlidou,
Kelly, & Williams, 2010; Zwart, Vissers, Kessels, & Maes,
2018; for a review, see van Witteloostuijn, Boersma,
Wijnen, & Rispens, 2017). The same AGL task, however,
has also been taken to reflect *nonmotor* procedural-
learning abilities (e.g., Finn et al., 2016; Gabay, Schiff,
& Vakil, 2012) and *sequential* procedural-learning abili-
ties (e.g., Krishnan et al., 2016).

**Box 2. Procedural versus declarative learning systems**

Initial evidence for memory not being a single entity came from patient H.M., who became densely amnesic
following a bilateral resection of the medial temporal lobe (MTL). Although he was unable to acquire new
information such as facts, he could learn new motor skills (Corkin, 1968; Milner, 1962). A decade later came
evidence that amnesic patients could also acquire perceptual skills such as reading mirror-reversed words
(Cohen & Squire, 1980). Given these findings, a distinction was made between *procedural* and *declarative*
memory systems—the former responsible for skill-based learning that is implicitly expressed through perfor-
man and the latter responsible for learning about facts and events.

Declarative learning has been shown to depend on the MTL system, and hence the umbrella term *nonde-
clarative* learning is used to describe learning that can be accomplished without the MTL system. It includes
procedural learning of skills and habits but also perceptual learning, priming, and classical conditioning (see
e.g., Squire & Dede, 2015; Squire & Zola, 1996). Procedural learning is thus considered one among several
types of learning that appear to be largely implicit.

This is the approach adopted in some definitions of
procedural learning, which are essentially based on
neuroanatomical demarcations. By this view, procedural
learning encompasses learning supported by the cere-
bellum and basal ganglia network. The advantage of
this definition is that exclusion and inclusion are clear
and unequivocal. Its drawback is in its explanatory
adequacy. If the definition of a faculty is not related to
its cognitive operations, then any function supported
by its corresponding neurocircuitry would be included
(e.g., eye-movement control would fall under proce-
dural learning because it recruits the basal ganglia
network).

A different approach to definition focuses on the
nature of knowledge that is acquired. Returning to our
early example of procedural learning, it can be defined as
any learning related to assimilating procedures (see Box
2), whether motor or not. Note that under this approach,
the underlying neurocircuitry is irrelevant to inclusion/
exclusion. The advantage of this approach is in its
greater transparency with regard to the cognitive opera-
tions of the theoretical construct. The drawback, how-
ever, is in the flexibility of inclusion/exclusion criteria.
Arguments regarding whether a learning situation is
procedural, implicit, or statistical are often based on
presuppositions or generated post hoc given the obtained
results.

A third approach to definition is to focus on the
computations underlying learning. This approach does
not consider the outcome knowledge but rather the
computational principles of the learning process that
result in the acquired knowledge. This approach is
exemplified by a common early view of SL that consid-
ers it as “the tracking of transitional probabilities” (e.g.,
Aslin, Saffran, & Newport, 1998; Thiessen & Saffran,
2003). Another example is the definition by Frost et al.
(2015): “Our approach construes SL as involving a set
of domain-general neurobiological mechanisms for

**Vagueness of the demarcation lines
between theoretical constructs
in the domain of learning**

Why do we not see agreement about the boundaries
of various cognitive faculties? Typically, this has to do
with different strategies to set the inclusion/exclusion
criteria that define them. One strategy to define a cog-
nitive construct is through its underlying neurocircuitry.

Why do we not see agreement about the boundaries
of various cognitive faculties? Typically, this has to do
with different strategies to set the inclusion/exclusion
criteria that define them. One strategy to define a cog-
nitive construct is through its underlying neurocircuitry.
learning, representation, and processing that detect and encode a range of distributional properties within different modalities or types of input” (p. 119). The advantage of this approach is that it requires explicit theoretical justification and is again clear about inclusion/exclusion criteria. A complication is that in many cases, different computations can lead to the same behavioral outcome. Thus, in the SL literature, extraction of patterns can be explained by tracking transitional probabilities, but it can also be explained by the continuous binding of elements into chunks (e.g., Perruchet, 2018). Moreover, some authors have argued that SL involves additional computations (e.g., learning of distributional statistics; Thiessen & Erickson, 2013). Hence, even the same approach to definition (i.e., one on the basis of computational principles) can lead to different definitions and demarcation lines.

The issue of definition becomes critical when the experimental tasks enter the game. If a brain-based approach is adopted, the selection of any task tapping the construct requires brain localization. In other words, the task relevance or irrelevance must be demonstrated through imaging or patient work. For example, any task that is not shown to implicate the cerebellum/basal ganglia system would not be labeled procedural learning (e.g., Conway, Arciuli, Lum, & Ullman, 2019; T. M. Evans & Ullman, 2016). If a knowledge-outcome approach is taken, the task relevance should be based on the nature of the acquired knowledge. For example, testing whether the SRT task leads to procedural knowledge requires excluding the possibility of explicit, declarative knowledge (e.g., Esser & Haider, 2017). Finally, computation-based definitions require a precise theory regarding the underlying computations involved that should be supported by computational work.

To be clear, there might not be a single right path to take in terms of definitions, and all three options discussed have merits. However, keeping in mind that constructive discussions regarding the relevant theoretical constructs and their interrelations require that definitions and inclusion criteria be made explicit, the following discussions will center on the possible computations involved in SL and their relevance to the language impairments in DD and SLI.

Let us return to our initial question, asking why different theoretical accounts converge on the identical prediction that participants with language impairments should show impaired performance in SL. In itself, the fact that different theoretical accounts lead to identical predictions is not necessarily a cause for concern if in principle the theories generate at least one contrasting prediction. For this, the theories need to generate precise predictions.

SL, however, is a vaguely defined construct. To begin with, there is no agreement regarding what constitutes SL: Interpretations range from a narrow definition (e.g., tracking transitional probabilities) to “all learning is SL” (see Arciuli & Conway, 2018; Frost et al., 2019). Different theories relate to the construct of SL from very different perspectives. Returning to our examples above, some approaches viewed SL as included within/overlapping with implicit learning (e.g., J. L. Evans et al., 2009; Gabay et al., 2015; Kahta & Schiff, 2019), whereas another centered on SL being a mechanism shaping ventral visual stream regions that support object and word recognition (e.g., Sigurdardottir et al., 2018). These different theoretical approaches lead to identical predictions because they focus on a vague commonality between the targeted cognitive construct (e.g., procedural learning, visual recognition, etc.) and SL. Precise predictions can be obtained through well-specified assumptions regarding the underlying computations that are shared by theoretical constructs. At present, however, SL research often reverts to abstract verbal theorizing regarding the commonalities of a range of regularity learning situations without specifying what regularities are the object of perception and learning, how they are represented in memory, and what the precise learning mechanisms are.

Abstract sketches do not provide a precise language for scientific discourse. Our brief review shows that researchers use the term SL to mean different things. They have different assumptions and intuitions regarding the computations underlying it and precisely how these relate to language impairments such as DD or SLI. From this perspective, defining SL as “a set of domain-general neurobiological mechanisms for learning, representation, and processing that detect and encode a range of distributional properties within different modalities or types of input” (Frost et al., 2015, p. 119) points to potential differences of computations (albeit not specifying them) in learning regularities given different modalities and types of inputs and is, therefore, a constructive step forward (for other computational approaches, see Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Thiessen, 2017; Thiessen, Kronstein, & Hufnagle, 2013).

Vagueness in the mapping between tasks and constructs

In the section above, we discussed the vagueness at the level of the theoretical constructs. In this section, we focus on the mapping between the constructs and the tasks taken to measure them. A straightforward situation is one in which there is a one-to-one mapping between tasks and constructs: Each task measures one construct, and each construct is measured by one task. This is, however, clearly not the state of affairs depicted in Figure 1. The historical trajectory of the field of
learning at large (Squire & Dede, 2015), and research on learning in DD in particular, led to mappings of one task to many constructs as well as mappings of many tasks to one construct. This situation is by itself not necessarily problematic. However, because the aim of investigations is to identify the underlying deficit—or deficits—of language impairment, explicit discussions of the relations between tasks and constructs are required for targeted experimentation.

**One same task is taken to measure different constructs.** One-task-to-many-constructs mappings are depicted in Figure 1 by multiple upward arrows departing from a single experimental task. Consider, for example, the SRT task. As per our review above, the task was developed to investigate the attentional requirements of sequence learning, the relation between learning and awareness, and the separation of memory systems (Nissen & Bullemer, 1987). Considering performance to reflect skilled-based knowledge, some researchers have taken the SRT task as a proxy of procedural learning (e.g., Hsu & Bishop, 2014; Lum et al., 2014; Willingham, Nissen, & Bullemer, 1989). However, focusing on performance as reflecting learning without awareness, other authors have taken the SRT task as a measure of implicit learning (e.g., Norman, 2015; Sævland & Norman, 2016). More recently, given the statistical structure present in the sequence or sequences, participants’ performance on the task has also been taken to measure SL ability (e.g., Christiansen, 2019; Perruchet & Pacton, 2006; Schmalz et al., 2017). The SRT task potentially taps all three constructs; there is no inherent contradiction here. One could view the SRT task as a task that relies on multiple nonoverlapping constructs, or one could adopt a theoretical view that places SL within procedural learning, which in turn is a subset of the bigger construct of implicit learning. An important question, however, is how to interpret impaired performance of a certain population (e.g., adults with DD) on the SRT task in light of this one-task-to-many-constructs mapping. Does it reflect problems with procedural learning? Implicit learning? Or SL? As demonstrated by Figure 1, this situation is not unique to the SRT task (for a discussion of AGL, see Christiansen, 2019). Here we argue that without more theoretical discussion and experimental work disentangling the different contributions to task performance, demonstrations of impaired performance in special populations are not readily interpretable.

**Different tasks are taken to measure the same construct and are taken to be interchangeable.** Many-tasks-to-one-construct mappings are depicted in Figure 1 by multiple arrows pointing to a single theoretical construct (e.g., “statistical learning”). Table 1 exemplifies this many-tasks-to-one-construct mapping by presenting a list of studies connecting DD and SLI to the theoretical construct of SL and specifies the range of experimental tasks employed to test this connection.

The choice of a specific task (rather than another) for a given study is rarely motivated by considering what computations are relevant to DD or SLI according to the specific aspects of language that these special populations have difficulties with (for an exception, see Tong, Leung, & Tong, 2019). Rather, tasks are often taken as interchangeable under the assumption that they all tap SL in the broad sense. This is nicely illustrated by the choice of tasks by van Witteloostuijn et al. (2019). These authors administered three different tasks testing children with DD and control children: a visual embedded-pattern learning task, a SRT task, and an auditory task of nonadjacent dependency learning. They hypothesized that children with dyslexia should experience difficulties across tasks tapping into SL abilities. Therefore, we assess children’s SL performance in a range of SL tasks that have previously been shown to be sensitive to learning in (typical) child populations and that span a number of methodological variations of SL tasks (e.g., modality, the type of statistical structure to be learned, online and offline measures). (para. 4)

However, if SL is a componential rather than unified ability (for an extensive discussion, see Frost et al., 2019; Siegelman, Bogaerts, Christiansen, & Frost, 2017; see also Arciuli, 2017), then these different tasks tap different dimensions of SL (e.g., in terms of modality, nature of the material, type of statistical dependencies) and most probably implicate quite different computations. Table 2 exemplifies this issue by dissecting the different tasks used to tap into SL in studies of special populations along a selection of potential processing dimensions, which suggests that different tasks involve some nonoverlapping computations.

In the section below, we discuss how such a fine-grained approach to SL tasks might help in offering precise theoretical accounts of the link between SL and language difficulties by considering their shared computations.

**Vagueness regarding how theoretical constructs are linked to language difficulties**

Accounts proposing an SL deficit in DD have been vague about explaining how and why this learning deficit leads to a core difficulty specifically in the domain of reading and spelling (for a similar argument
regarding domain-general theories of DD, see Ramus & Ahissar, 2012). In principle, SL as a theoretical construct appears to offer a potential learning mechanism that, on the one hand, is domain-general yet, on the other, is particularly relevant to the domain of language and literacy. However, the theory ought to be much more specific about the types of SL computations that are relevant to DD given what reading and spelling imply (see also Schmalz et al., 2017). In this vein, we outline two outstanding questions.

How does the proposed SL deficit lead to the particular profile of difficulties observed in DD? Here we ask, what are the types of statistics that dyslexics have difficulties learning, which then lead to their difficulties with literacy? Successful acquisition and processing of a written language is considered to rely on multiple types of statistical regularities, such as frequency of letter co-occurrences (e.g., bigrams, trigrams), orthography-to-phonology correspondences, morphological regularities (e.g., prefixes, suffixes, etc.), conditional probabilities of letters/words in context, and so on. If a weakness in the learning of regularities is hypothesized to underlie dyslexia, one should specify the profile of reading impairment from a regularity learning perspective.

Why would an SL deficit lead to specific problems with literacy rather than problems with perceiving regularities in both written and spoken language, as in the case of SLI? Figure 2 outlines various linguistic abilities and how they relate to SLI and DD. As Bishop and Snowling (2004) pointed out, the linguistic difficulties of individuals with SLI and DD overlap, and SLI and DD are frequently comorbid. Yet they represent two distinct disorders: Whereas people diagnosed with DD have a core deficit in literacy, SLI diagnosis requires deficits in production and comprehension of spoken language as well (see also Catts, Adlof, Hogan, & Weismer, 2005). A

Table 1. Overview of Studies Connecting DD and SLI to the Construct of Statistical Learning Using the Range of Experimental Tasks

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample (age)</th>
<th>Experimental task(s)</th>
<th>Modality</th>
<th>Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD versus control participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabay et al. (2015)</td>
<td>Adults (18–35 years)</td>
<td>SL-EPT</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Sigurdardottir et al. (2018)</td>
<td>Adults (18–60 years)</td>
<td>SL-EPT</td>
<td>A</td>
<td>Tones</td>
</tr>
<tr>
<td>Singh et al. (2018)</td>
<td>Children (8–12 years)</td>
<td>Target detection SL-EPT</td>
<td>V</td>
<td>Abstract shapes</td>
</tr>
<tr>
<td>Tong et al. (2019)</td>
<td>Children (7–8 years)</td>
<td>SL-EPT</td>
<td>V</td>
<td>Abstract shapes</td>
</tr>
<tr>
<td>van Witteloostuijn et al. (2019)</td>
<td>Children (7–11 years)</td>
<td>SL-EPT</td>
<td>V</td>
<td>Alien figures</td>
</tr>
<tr>
<td>He &amp; Tong (2017)</td>
<td>Children (8–11 years)</td>
<td>SRT</td>
<td>VM</td>
<td>Visual stimulus, 4 spatial locations</td>
</tr>
<tr>
<td>Kahta &amp; Schiff (2016)</td>
<td>Adults (18–33 years)</td>
<td>AGL</td>
<td>V</td>
<td>Consonant letters</td>
</tr>
<tr>
<td>Kahta &amp; Schiff (2019)</td>
<td>Adults (19–35 years)</td>
<td>AGL</td>
<td>A</td>
<td>Musical tones</td>
</tr>
<tr>
<td>Schiff et al. (2017)</td>
<td>7th grade ($M = 12$ years)</td>
<td>AGL</td>
<td>V</td>
<td>Shapes</td>
</tr>
<tr>
<td>Vandermosten et al. (2019)</td>
<td>3rd grade ($M = 9$ years)</td>
<td>Distributional learning</td>
<td>A</td>
<td>Speech sounds</td>
</tr>
<tr>
<td>SLI versus control participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. L. Evans et al. (2009)</td>
<td>Children (6–14 years)</td>
<td>SL-EPT</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Mainela-Arnold &amp; Evans (2014)</td>
<td>Children (8–12 years)</td>
<td>SL-EPT</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Haebig et al. (2017)</td>
<td>Children (8–12 years)</td>
<td>SL-EPT</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Plante et al. (2017)</td>
<td>Adults ($M = 20$ years)</td>
<td>SL-EPT (unfamiliar natural language)</td>
<td>A</td>
<td>Speech syllables</td>
</tr>
<tr>
<td>Lammertink et al. (2020)</td>
<td>Children (5–8 years)</td>
<td>NA-dep learning</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Hsu et al. (2014)</td>
<td>Adolescents (13–15 years)</td>
<td>NA-dep learning</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Iao et al. (2017)</td>
<td>Children (8–10 years)</td>
<td>NA-dep learning</td>
<td>A</td>
<td>Speech syllables (CV)</td>
</tr>
<tr>
<td>Sengottuvel &amp; Rao (2013)</td>
<td>Children (8–13 years)</td>
<td>SRT</td>
<td>VM</td>
<td>Visual stimulus, spatial locations</td>
</tr>
<tr>
<td>Hall et al. (2017); Hall et al. (2019)</td>
<td>Children (7–9 years)</td>
<td>Artificial grammatical category learning</td>
<td>A</td>
<td>Pseudowords</td>
</tr>
</tbody>
</table>

Note: DD = developmental dyslexia; SLI = specific language impairment; CV = consonant-vowel; CVC = consonant-vowel-consonant; A = auditory; V = visual; VM = visuomotor; AGL = artificial grammar learning; SRT = serial reaction time; NA-dep = nonadjacent dependency; SL-EPT = statistical learning embedded-pattern task.
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Embedded-pattern task</th>
<th>Syllables</th>
<th>Tones</th>
<th>Presentation mode</th>
<th>Nature material</th>
<th>Modality</th>
<th>Nature statistical dependencies</th>
<th>Type of transitional dependencies</th>
<th>Artificial grammar learning</th>
<th>Serial-reaction-time task</th>
<th>Note</th>
</tr>
</thead>
</table>
domain-general deficit in perceiving or learning statistical regularities (as per the common definition of SL) does not explain why a pronounced impairment with printed information is the state of affairs in DD. Moreover, several published studies reported correlations between SL performance and individual differences in linguistic skills that are considered to fall outside of the core difficulties of people diagnosed with DD, such as vocabulary acquisition (J. L. Evans et al., 2009; Mainela-Arnold & Evans, 2014) and comprehension of syntax in spoken language (Kidd & Arciuli, 2015). A theoretical account that links SL specifically to DD would have to address these findings, explaining the core deficit in the written domain.

To move forward, researchers could aim to explicitly draw theoretical links between a specific SL task and specific linguistic skills (and/or impairment) given the hypothesized computations shared by the task and the skill (and/or impairment). A recent article by Hall, Van Horne, and Farmer (2019) provides a rare example of a study that does exactly this. Focusing on children and adults with developmental language disorder, they selected verb bias sensitivity as the linguistic skill of interest and artificial category learning as the SL task because of their hypothesized shared underlying components. Although no robust evidence was found for the predicted relationship, the approach is promising, and extending it to different linguistic skills and special populations generates a range of testable predictions. For example, if proficient reading involves the registering of letter co-occurrence statistics (e.g., Cassar & Treiman, 1997; Gingras & Sénéchal, 2019), individual abilities in learning transitional probabilities (or embedded patterns) in the visual modality are expected to predict individual abilities in reading and spelling skill, whereas individuals with DD would show impaired performance on this task.

In the same vein, the learning of nonadjacent dependencies between spoken syllables in an auditory stream (e.g., Gómez, 2002) could be taken as a predictor of syntactic processing of spoken language input given that tracking syntactic structure requires identifying nonadjacent relations (e.g., Grunow, Spaulding, Gómez, & Plante, 2006; Misyak, Christiansen, & Tomblin, 2010). In this case, however, impaired performance on this specific task should be predicted for individuals with SLI rather than individuals with DD. We do not voice here a specific theory regarding what is implicated in reading or syntactic processing. We simply argue that such a theory should be made explicit when selecting a specific SL task for a given linguistic competence and hence for a given special population with a given language impairment. Note that in testing theoretical links, it would be advantageous to consider not only group differences but also individual differences (within a diagnostic category) for the specific language skill of interest to explain relative strengths and weaknesses (an approach taken in the modeling work on DD by Ziegler et al., 2008, and the recent study by Hall et al., 2019, on developmental language disorder mentioned above).

Another possible strategy would be to take a data-driven approach. Rich, large-scale data sets, including performance on a multitude of SL tasks outlined in Table 2 and others, from large samples of neurotypical participants as well as participants with DD and SLI would allow researchers to employ techniques such as structural equation modeling to extract latent variables and to compare different diagnostic groups on each latent variable. These latent variables might correspond to some dimensions outlined in Table 2 (e.g., the type of regularity or modality), but, of course, different theoretical dimensions may be also uncovered.

Independent of whether one’s preferred strategy is testing specific theoretical links or launching data-driven analyses, a fine-grained approach to SL tasks has the promise of making sense of the mixed findings regarding an SL deficit in DD in the literature (for reviews and meta-analysis, see Lum et al., 2013; Schmalz et al., 2017; van Witteloostuijn et al., 2017). It can help clarify what specific aspects of SL are important for oral language acquisition, syntactic abilities, or literacy skills. With respect to the aspects of SL that are important for literacy, it is worth noting that learning the mappings between orthography to phonology and vice
versa was not captured by any of the empirical studies in Table 1. Indeed, none of the studies taking a “statistical” approach to DD have looked at the learning of cross-modal regularities (despite extensive work on cross-modal SL in other domains, e.g., Mitchel & Weiss, 2011; Weiss, Poepsel, & Gerfen, 2015). This is surprising given that phonological decoding ability (i.e., the process of converting orthographic sequences to their spoken forms) is widely accepted to be one of the fundamental skills underlying proficient reading and that deficits in phonological decoding have been consistently documented in populations with DD (for a review, see Vellutino, Fletcher, Snowling, & Scanlon, 2004).

Finally, in parallel to taking a fine-grained approach to SL and linguistic difficulties, future researchers might want to attempt to better understand their interrelation by considering development itself as a key aspect in shaping phenotypical outcomes. A neuroconstructivist perspective (e.g., Karmiloff-Smith, 1998) would postulate that subtle differences in the extent or nature of an SL impairment present early in development might change the course of developmental pathways significantly. This contrasts with the notion (seemingly often—at least implicitly—assumed in research on developmental language disorders) that an entire cognitive construct, here SL, is initially impaired and necessarily remains impaired to the same extent throughout development. Whereas studying SL across development poses practical and psychometric challenges (Arnon, 2020; see also West, Vadillo, Shanks, & Hulme, 2018), such empirical evidence could allow a leap forward in our understanding of various profiles of language difficulties.

Other special populations as another piece of the puzzle. We have so far considered only SLI (or developmental language disorder) and DD, but the conundrum grows even further given the fact that SL deficits are postulated to give rise not just to those two disorders but also to language difficulties in other special populations, in particular to those observed in individuals with autism spectrum disorder (e.g., Jeste et al., 2015; Scott-Van Zeeland et al., 2010). If an SL deficit underlies language difficulties in all of these special populations, why do they present so differently? Worth mentioning in this context is the recent study by Lieder et al. (2019). They postulated a difference in the relative weighting of recent versus older stimuli for the tracking of stimuli statistics. Compared with control participants, individuals with DD would rely more on information about the immediate past (i.e., fast forgetting), whereas individuals with autism spectrum disorder would rely on longer-term statistics (i.e., slow updating). This work exemplifies a well-specified account that aims to jointly explain DD and autism spectrum symptoms, as well as the differences between them, in terms of SL computations.

Moving Beyond Studies With Singular Confirmatory Predictions

How can we reach a constructive state of affairs in which different theories regarding the link of DD and SLI to SL generate at least one contrasting prediction? A necessary condition is that the theory will specify a range of predictions to be tested in parallel: predictions regarding what should be impaired and, importantly, predictions regarding what should not. However, our review of the literature shows that studies explicating the link of DD and SLI to SL offer but one single prediction and then proceeded to test it via a confirmatory strategy. The logic of inference is of the following kind: “The well-documented DD deficit in implicit learning should imply a deficit in SL (as measured by one or multiple experimental tasks tapping SL) because SL is part of implicit learning.” What is missing is a parallel test of exclusion. For example, “If SL is impaired because of the implicit nature of the learning process, then DD will not be impaired in SL tasks that implicate explicit learning conditions.” Or, “If an SL deficit for DD is related to regions of visual object recognition, then DD will not show SL deficit in other modalities.” In the studies exemplified above (see Table 1), this was not done. For example, Sigurdardottir et al. (2018), who focused on problems with the ventral visual stream, did not show that DD are performing normally on auditory SL. Gabay et al. (2015), who focused on SL abilities as drawing on procedural learning, demonstrated impaired performance on an SL embedded-pattern-learning task with linguistic and nonlinguistic auditory materials but did not show how the DD group performed on an explicit learning task with the same materials.

This issue was well formulated by Ramus and Ahissar (2012), who argued that group studies of this kind should always demonstrate normal performance in a condition that does not involve the specific theoretical construct under investigation because if not, the observed poor performance cannot be tied to the specific hypothesized deficit. Stressing the importance of demonstrating specificity even further, note that many of the current findings of impaired SL performance could, in principle, be attributed to abilities not related to learning—most notably, low-level sensory or working memory deficits. This is because control tasks that assess the ability to encode the materials used in the learning task, or their short-term storage, are typically not administered. It seems likely, for example, that a deficit in encoding auditory stimuli would result in poor auditory SL performance and difficulties with language
acquisition, so it would be much more informative to observe poor SL performance in the presence of normal, unimpaired encoding performance. Note that the issue of testing singular predictions is not specific to SL as a field. However, the investigation of the role of SL in language impairments is a relatively new area of research, and yet it seems not to escape this critical pitfall.

Concluding Remarks and Guidelines for Future Research

This final section outlines possible directions for moving a research program forward that connects SL to language impairments. Such a research program holds the promise of better understanding how SL abilities determine linguistic abilities and disabilities. First and foremost, we argue that researchers should be explicit regarding their approach to definition and inclusion/exclusion criteria while discussing the theoretical constructs that are relevant to the specific targeted deficits. One should be particularly explicit regarding what SL is and its relations to related cognitive faculties.

Second, a theoretical approach that regards SL to underlie a given language impairment should explicate how and why the SL deficit leads to the specific symptoms that characterize the impairment and not to others.

Third, constructive advances require that studies should be designed from the outset to contrast different theories regarding the role of SL in a language impairment rather than confirm a singular prediction. Practically, this could be achieved by including control conditions in the study design, for which normal performance is anticipated.

Finally, a fine-grained approach to different measures of SL abilities (in contrast to considering experimental SL tasks as interchangeable) can generate testable predictions regarding the specific SL computations that are relevant to a given impairment.

Transparency

Action Editor: Aina Puce
Editor: Laura A. King
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding
This work was supported by Marie Skłodowska-Curie Grant Agreement 745258 (IF-EF, European Union’s Horizon 2020 Research and Innovation Programme) awarded to L. Bogaerts and European Research Council Advanced Grant Project 692502-L2STAT awarded to R. Frost.

ORCID iD
Louisa Bogaerts https://orcid.org/0000-0001-6145-8662

Acknowledgments
The work on this article was completed at National Taiwan Normal University.

Notes
1. Other lines of research have investigated SL in the context of autism and populations with cochlear implants (e.g., Deocampo, Smith, Kronenberger, Pisoni, & Conway, 2018; Jeste et al., 2015; Scott-Van Zeeland et al., 2010). However, we chose to focus on DD and SLI given the hypothesized link between SL and linguistic abilities. For both of these developmental disabilities, language difficulties are at the core of their symptoms, and by definition, all diagnosed individuals experience significant difficulties with one or several aspects of language. This is not the case with, for example, autism, for which just a subpopulation displays language difficulties.
2. Note that some researchers and clinicians have moved away from the diagnostic label specific language impairment toward the more inclusive label developmental language disorder (e.g., Bishop, 2017; Hall, Van Horne, McGregor, & Farmer, 2017; Plante, Patterson, Sandoval, Vance, & Asbjørnsen, 2017). SLI assumes difficulties only with language and thus a discrepancy between verbal and nonverbal abilities. Developmental language disorder is less restrictive, so individuals with lower nonverbal abilities can also receive this diagnosis, and it can co-occur with other neurodevelopmental disorders. Whereas a more inclusive diagnostic label might be helpful and constructive for clinical purposes, this diagnostic change might hold additional difficulties for research, primarily because a shared diagnostic label might give the illusion of reflecting a unified profile of difficulties. There is a clear advantage in specifying precisely what aspect of language difficulty one is trying to comprehend or predict, and highly inclusive diagnostic labels do not provide insights regarding the specific difficulties that are under investigation.
3. This mapping situation is related to what are known as jingle-fallacies (e.g., Gonzalez, MacKinnon, & Muniz, 2020). Jangle-fallacies occur when different labels or different measures actually refer to the same construct (Kelley, 1927, pp. 62–65), jingle-fallacies occur when a same term or measure refers to different constructs (Thurstone, 1903).
4. Whereas most accounts of SLI do not exclude difficulties with literacy, cases of SLI without DD have been reported (Ramus, Marshall, Rosen, & van der Lely, 2013), so the converse question may be raised as well: If an SL deficit gives rise to SLI, why would it not, at the same time, give rise to DD?
5. Some paradigms, such as the SRT task and Hebb repetition paradigm, have a control condition by design because they define learning in terms of a difference between performance on random (unique) sequences and structured (repeated) sequences. For tasks such as the embedded-pattern task and AGL task, this is not the case.
References


