UNIVERSITY OF PAVIA – IUSS SCHOOL FOR ADVANCED STUDIES PAVIA

Department of Brain and Behavioral Sciences (DBBS)
MSc in Psychology, Neuroscience and Human Sciences

THESIS TITLE

Predictive authority in Statistical Learning: acquisition, availability, and deployment in a Global–Local visual paradigm

Supervisor:
Prof. Davide Crepaldi

Thesis written by
Eleonora Lippi

Academic year 2025/2026

Table of Contents

ABSTRACT

Statistical learning is commonly defined as the ability to extract regularities from structured input. However, behavioural success does not transparently reveal what has been learned, whether the acquired structure is concurrently available, or whether it can guide behaviour when competing predictions coexist. This dissertation addresses these issues through a novel visual statistical-learning paradigm in which participants were exposed to two probabilistic dependencies embedded in the same stream: a local dependency and a temporally more extended global contingency, matched in predictive strength and later placed in direct conflict. Two experiments examined complementary levels of the learning-to-use trajectory, distinguishing accuracy-based sensitivity to each dependency from behavioural deployment, defined here as the selective use of one predictive structure to guide choice when competing regularities supported incompatible responses. Experiment 1 was conducted in person with counterbalanced stimulus lists and assessed whether the paradigm elicited interpretable group-level behavioural signatures across accuracy-based, deployment-related, and response-dynamic indices. Experiment 2 was administered online with a standardized list and evaluated whether the same index architecture could support participant-level characterization through diagnostic analyses of reliability, distributional adequacy, and within-family redundancy. Across experiments, the paradigm did not provide robust evidence for dual accuracy-based acquisition. Crucially, however, accuracy and deployment did not collapse into the same behavioural dimension. The in-person sample showed a group-level tendency to deploy the globally consistent alternative under conflict, despite the absence of a corresponding global-specific acquisition advantage. Conversely, the online sample did not reproduce this shared deployment bias but showed stronger participant-level alignment between relative learning balance and conflict choice. These findings suggest that statistical learning in probabilistic multi-regularity environments cannot be reduced to above-baseline accuracy. Rather, behaviour reflects the conditional interaction among sensitivity to structure, availability of predictive alternatives, decisional authority under conflict, and measurement stability. Although exploratory, the paradigm offers a principled framework for studying when statistical regularities become under uncertainty.

Keywords: statistical learning; visual statistical learning; probabilistic learning; predictive authority; predictive processing global–local dependencies; behavioural deployment; individual variability.

.

INTRODUCTION

Chapter 1. Beyond a phenomenological account of statistical learning: the representational ontology of a structural plurality

When explanatory adequacy is concealed behind empirical robustness

Statistical learning has increasingly been posited as a reliable component of human cognition, underlying a pervasive sensitivity to the distributional spatial and temporal properties of structured environments. The salience of such a computational capacity lies in its organizational potential, enabling learners to abstract cohesive and predictive regularities from an otherwise chaotic and informatively rich stream, emerging in the absence of explicit goals, purposeful strategies or reinforcement contingencies. Spanning from early demonstrations of implicit acquisition of artificial grammars by mere exposure, (Reber, 1967) to seminal findings highlighting the usually downplayed extent of experience-dependence in learning mechanisms of preverbal infants (Saffran, Aslin, & Newport, 1996), converging evidence accumulated over several decades supporting the robustness of the phenomenon. Furthermore, the legitimacy of this construct as an empirically solid target of inquiry, rather than a mere epiphenomenon, has been substantiated by the consistency of documented prevalence of statistical learning across modalities, age groups, and task contexts.

At the same time, however, the very reproducibility of these paradigms has fostered a debatable tendency to treat statistical learning as explanatorily self-sufficient, licensing the conflation of experimental consistency with descriptive adequacy: while empirical robustness corroborates the occurrence of learning, it does not clarify the features of the acquired representational structure or the processes underlying its behavioural expression.

This practice is already apparent in the foundational logic through which implicit learning is operationalized: since Reber's (1967) paradigmatic demonstration that participants could acquire structured knowledge without conscious access, indirect behavioural measures have been legitimized as indicative of learning. A substantial ambiguity seems consequential: if learning is detected independently of awareness, then the access to the authentic nature of the cognitive model is presumably inferred and, as such, possibly subject to post-hoc epistemological distortions. Indeed, similar performance patterns may arise from qualitatively different internal processes and behavioural data alone don't allow to disambiguate outcomes of sensitivity to either local statistics or chunk-based representations, task-specific response heuristics from successful acquisition of abstract relational knowledge. Hence, while behavioural inference gains legitimacy as a viable and reliable empirical tool, it can be conceived as a necessary but not sufficient epistemic probe. It might be tempting to harmonize this criticality by resorting to the implicit nature of learning itself, as the absence of conscious access can be naïvely mistaken for a warranty of the immediate and straightforward mapping between behaviour and acquired structure. However, an assumption that presumes contiguity of processes as a valid proxy for causal relationship is logically untenable. Comparable to how the attribute of implicitness concerns access to representations rather than authenticity of format or scope (Turk-Browne, Jungé, & Scholl, 2005; Monaghan, Schoetensack, & Rebuschat, 2019), the absence of explicit awareness does not suffice to imply a shallow kind of learning, purely tied to surface features and ill-suited to transcend individual exemplars. Such a reductionist view is, indeed, easy to falsify once we fully acknowledge the documented wealth of empirical evidence concerning generalization phenomena across contexts or rule extension to novel settings. Notwithstanding that awareness may modulate learning trajectories and behavioural expression, it is certainly not sufficiently ubiquitous, nor efficient enough, to specify the epistemic unit of learning; therefore, a principled characterization of the structure that has been acquired is required to supplement measures merely indexing the occurrence of pattern acquisition.

Without detracting from the meaningfulness in establishing evolutionary salience and age-invariance of statistical learning abilities, a parallel level of under specification is implicit in the canonical demonstrations of developmental statistical learning by Saffran et al. (1996, 1997). In the study, compelling evidence is provided that infants can incidentally extract unitary structures from continuous streams of nonsensical speech by tracking transitional probabilities between adjacent elements. The very success of this paradigm, however, contributed to progressively narrow the theoretical scope of the construct: by operationalizing learning primarily through sensitivity to adjacent transitional probabilities, a specific structural dimension was privileged and epitomized as the primary representational form of an otherwise constitutively heterogeneous phenomenon. Despite being beneficial to experimental tractability, this uniform conceptualization represented an a priori constraint of the space of the admissible structures, potentially reducing the full appreciation of its breadth.

This historical trajectory underscores the relevance of a deeper issue: regularities differ in their epistemic status, idiosyncratically shaping the learning problem they instantiate. Deterministic regularities define one-to-one mappings that, once discovered, support categorical predictions whose correctness is invariably guaranteed by the successful application of rule-based schemas, which, if and when adequate, afford the possibility to eliminate uncertainty. Conversely, probabilistic regularities, even after extensive exposure, preserve competition among alternatives, with multiple choices remaining viable based on their relative, although differentially graded, compatibility with the same evidence. Here learning aims not to resolve uncertainty, but to continuously reshape expectations through weighting and updating processes, yielding, rather than discrete rules, graduated predictions. This distinction is methodologically critical, as it determines whether learning culminates and can, thus, be quantified, as certainty, or whether it continues to operate under conditions of residual uncertainty.

Notably, transitional probabilities as the canonical cue of statistical learning, do not in themselves determine which of the above outlines learning regimes is instantiated. In the canonical work by Saffran and colleagues, the relational status of the units embedded in the information stream was deterministically arranged: transitional probabilities reliably cued a rule-based solution of the learning problem, supporting chunking and segmentation into stable, identifiable units and effectively eliminating uncertainty. The same formal cue, however, can be deployed to construct learning contexts in which uncertainty is preserved: here subjectivity will determine the informativeness of each option, affording a fundamentally different kind of learning, detectable in a flexible bias in expectation rather than in accuracy of predictions.

From an informational perspective, this focus on probabilistic structure is not incidental. Developmental evidence indicates that learners preferentially allocate attention to stimuli of intermediate predictability, where uncertainty is neither eliminated nor overwhelming (Kidd, Piantadosi, & Aslin, 2012): such features optimally maximize informational value, simultaneously providing sufficient structure to support learning while affording enough contextual variability to remain instructive. Evolutionarily designating the value of learning systems, their efficacy and usefulness lie in the development of a computational architecture mirroring the requirements of an environment that not collapse into determinism; whereby proficiency in extracting coherence does not preclude tolerance to a degree of ambivalence. Accordingly, quasi-regularity and residual uncertainty can be conceived not as a failure of statistical learning, but as a defining feature of its operating modality.

The assumption of a kind of learning that unfolds under persistent uncertainty cast further doubt in the practice of treating behavioural outcomes as transparent outputs of acquired structure. Considering that performance reflects, beyond what has been learned, how information has been selected, prioritized, and deployed in context, an interpretative stance is required that foresees responses as contextually relevant, flexibly adjusted behavioural resolution of competition among partially informative sources of evidence. Consequently, error does not straightforwardly signal learning failure, nor does accuracy unambiguously index learning strength, with direct implications for how statistical learning is probed. Indeed, in probabilistic environments, the experimental task becomes a constitutive component of its behavioural outcomes; by imposing relevance criteria and response demands, a specific paradigm selectively enhances the salience of certain dimensions of structure and concurrently down-weight other ones.

The inferential consequences of this task dependence are substantial: tasks articulated as to privilege specific structural dimensions, such as adjacency, temporal proximity, or local predictability, constitutionally establish the type of statistical learning preferentially elicited and the aspects of learned structure that will become behaviourally manifest. Since observed effects cannot be uniquely attributed to changes in learning per se, the absence of an effect may reveal a mismatch between learned structure and task demands or measures, whose scope is suboptimal in specifically addressing the feature of interest. Hence, the assessment of construct validity demands a broadening of the concept space of statistical learning as encompassing a plurality of learning problems that, even when nominally grouped under the same construct, are defined by different structural constraints. Assuming this perspective, the well-documented variability in statistical learning outcomes, yielding weak correlations across tasks, inconsistent group-level effects, and dissociations across modalities, becomes an expected outcome of sampling among different subsets of a broader construct scope, rather than noise obscuring a unitary ability.

Crucially, this pluralism is not merely functional but structural: evidence increasingly converges in endorsing the articulate nature of statistical learning, which engages multiple interacting processes ranging from attentional selection to decision policies, through evidence accumulation and updating dynamics, each component specifically sensitive to different aspects of the input. Attentional mechanisms hierarchically gate the cognitive availability of sensory inputs, critically being shaped by the relevance of prior expectations and by evolutionarily determined pressures toward informational efficiency; updating mechanisms balance sensitivity to new evidence against stability, determining the overtime evolution of subjective expectations; decision policies arbitrate among competing predictions in light of task goals and niche demands, translating graded expectations into categorical, quantifiable response. From an inferential standpoint, statistical learning cannot be reduced to a homogeneous mechanism exclusively driven by the strength of a single representation, it rather appears as a composite, dynamic and systemic organization inseparable from its ecological context, where learning outcomes reflect the coordination of an apparatus operating under informational and computational constraints, directly reflected in the plethora of statistical learning manifestations. This multi-processual theoretical perspective logically motivates why attempts to characterize statistical learning as a unitary capacity have repeatedly encountered empirical and conceptual obstacles: the phenomenon of statistical learning is robust precisely because it reflects a flexible system. tuned to structured, yet uncertain, environments; its explanatory complexity is the price of that flexibility.

None of these components can be however cleanly isolated through standard experimental manipulations, whose nature is inherently “fat-handed” (Eronen & Bringman, 2021) owing to the fact that changing the structure of the input simultaneously affects multiple stages of the inferential pipeline. Therefore, the attempt to trace a one-to-one relationship between the acquisition of structured expectations and the specifical deployment of such expectations to guide behaviour in a certain context is epistemologically frail. In probabilistic environments learning and usage not necessarily align, making it virtually impossible to distinguish learners who have acquired reliable expectations, just not yet sufficiently specified or too faint to be expressed in overt responses, from subjects relying instead on heuristic strategies that bypass learned structure altogether. Successful learning doesn’t necessarily manifest as monotonic improvements in performance; while such behavioural patterns appear sustainable in deterministic environments where learning converges on certainty, probabilistic contexts may instead produce counterintuitive learning trajectories, with the enhanced inferential conflict potentially leading to unstable performance, increased variability, or even apparent interference. That being the case, it appears critical to disentangle learning measures and performance metrics.

The seemingly delineating challenge, therefore, is not merely to demonstrate learning, nor to refine measures in search of a more reliable proxy for a putative latent ability; rather, to advance toward principled explanations aiming to characterize which dimensions of structure can legitimately be targeted by an empirical paradigm that accounts for the distinctive nature of interacting inferential processes in the interpretation of behavioural outcomes. Further, performance as a whole demands to be reconceived as depending on the reciprocal interdependence between input, representation, and use developing under conditions of environmental uncertainty.

A conceptual roadmap to a multidimensional structure

Once a plurality of phenomena legitimately identifiable as occurrences of statistical learning has reasonably been proposed, a tempting and seemingly consequential effort is to advance a purely dimensional account, one descriptively enumerating type, format, complexity of statistics. However, the dynamic features of statistical learning require us to understand how structural units are functionally organized, which principles determine their differential relevance, and under which conditions they become behaviourally expressed. Abandoning a scalar handling of the phenomenon, experimental designs must aim not to assess a generic gradient of learning strength, rather to identify the member of the statistical learning family that is being targeted, the learning criticalities currently instantiated and the structural interpretative dimension that is situationally advantageous to exploit. A device able to allocate learning priorities justifies the adaptive need to confront, in naturalistic input, the presence of simultaneously available regularities that, if all equally weighted, render environmental informativeness null and impose significant expenditure of already limited computational resources.

A principled, systematic attempt to map this composite space must ideally apply both within and across modalities and each atomic dimension should be qualitatively characterized in the representational and computational demands it comprises. A valuable compass in this topic is the work of Growns et al. (2020), that, even within the single domain of visual statistical learning, uncover a reasonable level of analysis of the resulting performance along orthogonal axes: the type of statistic defining predictability and the format in which that statistic is instantiated.

First, the nature of the statistical information underlying predictions is critical: conditional structure can arise when a relational metric instantiates predictivity ("*since A happened, then B is likely*"); conversely, distributional structure emerges from sensitivity to cardinal properties such as frequency of occurrence or relative prevalence. Capitalizing on correlational strength as a proxy to infer similarity of engaged processes, it appears that domain-general principles are operationally shared across tasks distinct in this first axis; however, the degree of ability in distributional and conditional tasks is not mutually predictive, with enhanced sensitivity to conditional structure not necessarily entailing superior capabilities in the distributionally organized dimension. As clearly argued by Growns et al (2020), what is commonly grouped under a single label or a global ability, encompasses instead a set of partially dissociable learning processes that differ in the type of structure they target, the representations they support, and the behavioural signatures they produce.

Furthermore, since statistical structure is not a property detached from the processing limitations of the percipient system, its unique temporal, spatial and modality-dependent properties become epistemologically relevant, licensing to consider the format of the representation as a second orienting axis. Accordingly, Zhou et al. (2024) show that even within the domain of distributional learning, presumably entailing homogeneous and comparable demands, individual performance show dissociations that are systematically modulated by whether structure is instantiated in auditory or visual modality, with temporal or spatial layouts. Ordin, Polyanskaya, and Samuel (2021) further hone the interpretation of such effects: inter-modality differences arguably embody the result of ecological pressure in determining a stratification of which kinds of regularities are more functionally useful in a specific domain. Such hierarchization will circumscribe the defining features of the learning problem and the computational *niche* that a specific system is optimized to handle; accordingly, the distinctive temporal acuity of the acoustic domain and the idiosyncratic spatial sensitivity of the visual apparatus can be leveraged to motivate distinctive modality-specific learning masteries.

Postulating type of statistic and format as organizing principles allow to examine whether apparent inconsistencies in statistical learning literature can be recast as dimension mismatches, due to improper nominal assimilation under the same label of different empirical aims, in fact diverging in the kind of statistic made informative or in the modality-specific formats used to implement them. Besides, the extent to which strategic prioritisation among the postulated pool of dimensions might differ across different subjects, is a further challenge to the notion of a default operating modality of statistical learning mechanisms. Namely, traditional narratives often assume an underlying complexity gradient embodied in representational "depth", whereby transitional probabilities are the authentic unit of statistical learning, with segmentation or chunking processes gradually emerging. Newport and Aslin's (2004) contribution on nonadjacent dependencies refines this theory: although evidence of non-adjacent acquisition encompassing only phonetic segments but not syllables seem to corroborate the constraint on the specific unit instantiating learning; learning at a distance cannot be oversimplified as a kind of noisier adjacent learning. They proposed a suitable explanation of this finding as depending on the two-fold effort imposed by the processing of syllables, which are not atomic elements as phonetic segments. Being syllables a compound of constituents, a chunking process, working in parallel to the baseline statistical learning computation, is required to combine disjointed elements into a larger unit. The enhanced difficulties in the acquisition of non-adjacency relative to transitional probabilities can thus be reformulated as stemming from not comparably taxing demands imposed by the degree of complexity of the material instantiating statistical structure.

Emerson and Conway (2023) advance the debate by leveraging the inferential potential intrinsic the use of differential reaction times as a metric to distinguish the degree of developing sensitivity to the predictiveness of positional information, in turn determining the forming responsiveness to statistical structure. Their modelling approach isolate the kind of predictions instantiated by transitional-probability and chunking accounts about the different stabilization and generalization profiles of units and subunits. Refuting a gradient of complexity, orderly progressing from the acquisition of simpler adjacent occurrences toward more abstract patterns, they demonstrate different dynamics of representational stabilizations, outlining a system that simultaneously maintains multiple descriptions of structure, selectively privileging which one will be strengthened and, eventually, behaviourally exhibited. Notably, when chunk-based, unitary representations become stable and contextually relevant, performance can give the impression that conditional learning about constitutive subunits has weakened; however, sensitivity to local statistics may persist and even emerge by employing appropriate probes. Therefore, the competitive dynamics underlying information processing don't seem to remove less currently relevant cognitive units, which remain, although transiently weakened, latent organizational principles, potentially yielding interference or disturbance for the predominant one.

A convergent line of evidence is provided by developmental neuroscience: Benjamin et al. (2023) show that neural entrainment dynamics can reveal sensitivity to different level of statistical structure, with signals associated with the frequency of occurrence of transitional probability remaining functionally and temporally distinguishable from those diagnostics of larger units. Crucially, the emergence of segmentation-like signatures does not imply that lower-order statistics are not tracked at all; rather, which representations will gain dominance seems established through continuous reweighting of relative functional relevance.

The purpose of this reclassification of statistical learning is not to unnecessarily multiplicate epistemic categories, as typical of logical fallacies, it rather aims to recognize the hermeneutic value in avoiding to arbitrarily seclude the nature of a percept from the unique representational system of its perceiver. Nevertheless, a mere mapping effort wouldn't be sufficient, unless it proves to be an asset in accounting for robust empirical phenomena historically treated as anomalies or limitations of statistical learning paradigms. As for inconsistencies and dissociations in individual performance across tasks, phenomena potentially undermining the validity of a construct, might instead be treated as tools to theoretically probe and circumscribe the extent of this system's computational capabilities.

Exemplars are Gómez's (2002) findings on variability, factor that they found to operate as a structural manipulation redistributing informativeness across dimensions. Since statistical structures spanning over different distances are governed by partly dissimilar operating logics, contextual modulations produce substantially divergent detrimental or beneficial effects. For instance, input variability actively destabilizes locally consistent patterns, suppressing default computations reliant on adjacent probabilities; hence inducing a strategical transition to higher order "gestaltic" computations depending on similarities across non-contiguous items. Sensitivity to the relational status of non-adjacent elements consequentially arise from unreliability of proximity information, suggesting a role for variability in actively orchestrating shifts in the balance between confidence in parallel and competing structural hypotheses, each evaluated in its relative proficiency in handling a flexibly changing learning problem.

Equally valuable are paradigms providing theoretical stress tests by inquiring boundaries of abstraction and representational scope of statistical learning phenomena. Marcus et al. (1999) provide influential evidence of the effective extraction of algebraic-like rule in infants as early as at 7 months old: through implicit measures of preferential looking, habituation was assessed across experiments where reliance on local conditional statistics wouldn't have been sufficient to access the rule underlying the degree of "word-resemblance" of each sequence. Complementary to the assessment of impressive abstractive potential of statistical learning mechanisms, Lengyel et al. (2019) showed that exposure to a tactile unimodal statistical input can give rise to predictive accuracy in determining visual features of the same object. Here extraction of statistics flexibly transfers across domains, capitalizing on the inferential power intrinsic in regularly experienced, trans-modality co-occurrences. Remarkably, the dimension of concrete, object-like representational predictability (Lengyel et al, 2019) seems to share statistical computational basis with a more abstract, relational realm (Marcus et al, 1999).

In this perspective, the main value of the above mapping effort is to delineate the potential of a perspective transitioning from framing the process of statistical learning as a passively receptive one to a proactive interaction of the subjects' dynamic inferential architecture with an overwhelmingly uncertain habitat.

Structural architectures yield ontological commitments

The taxonomy previously outlined equips us with interpretative tools helpful to harmonize a plethora of phenotypes of statistical regularities, differing in structural scope, temporal span and informational targets. A mere structural layer is however insufficient to address a deeper issue: which is the cognitive nature of representational entities resulting from statistical learning processes? This subtle, yet quintessential transition shifts the level of speculation from what information is in principle available and coexisting in the input to the featural qualities of distinct representational objects, their different constraints and predictive affordances: precisely, from descriptive specifications to ontological commitments.

A classical approach to characterize the representational distinctiveness of outcomes of incidental extraction of structure occurring in unsupervised learning situations exploits the "algorithmic" level of analysis (Marr, 1982). The dominant interpretation of acquisition processes held by traditional implicit learning theories leverages progressive chunk formation, statistical learning (SL) paradigms instead customarily privilege segmentation mechanisms driven by transitional probabilities, particularly among contiguous elements. However, this long-standing dichotomy proves far less diagnostic than what is assumed (Perruchet and Pacton, 2006): statistical computation and chunk-based accounts can indeed converge on identical behavioural predictions and equally be challenged by identical empirical phenomena, most notably the acquisition of non-adjacent dependencies. Such evidence deters from an unambiguous recognition of the representational commitment currently made by the system from behavioural outcomes, as apparent statistical segmentation may as well reflect the formation of predictive chunks, whereas chunking-like effects may emerge as downstream corollaries of probabilistic sensitivity.

A critical test bench to disprove, as classically endorsed by dualistic accounts of learning trajectories, presumed inadequacies of chunking models in reproducing behaviourally documented efficacy in non-adjacent acquisition, is provided by Isbilen et al. (2022)'s work. Employing a chunking-based serial recall task, they infer, from the comparable degree of recalling accuracy mastered for both adjacent and non-adjacent dependencies, the presence of parallel computations and further show that learning non-adjacent structure involves the formation of specific predictive units that go beyond positional information, allowing learning to encompass both item-level and internal dependencies' knowledge. Particularly notable is the demonstration that the acquisition of non-adjacent frame structure is sufficiently stable to scaffold generalization to items differing in the intervening element. Taken together, these considerations rule out that sensitivity to transitional probabilities and chunk formation may be treated as mutually exclusive explanatory levels, engaging computational mechanisms uniquely dedicated to distinguishable tokens of learning. Therefore, we can go further the mere acknowledgment that multiple structural dimensions can be processed within a unique input stream: even when a single structural dimension, as in the case of non-adjacent dependencies, is cleanly isolated, the designation of the computational architecture producing the specific nature of its learning outcomes is not unequivocal. While non-equivalent in their internal organization and in their adequacy to handle different representational objects, chunking processes and segmentation mechanisms may partially overlap in their work. What instantiates learning outcomes is not the specific cognitive modality the system preferentially resorts to; rather, it becomes fundamental the granularity of the unit of prediction.

Considering that the specific representational commitment of the cognitive system is a necessary premise that determines prediction, generalization and behavioural control in a given context, its characterization effort can benefit from a set of neural signatures leveraging architecture constraints, functional consequences and temporal persistence.

Exemplarily, Turk-Browne et al. (2009) attempted a better characterization of the phenomenon of statistical learning employing the informativeness of neural constraints, not merely in the effort to delineate an anatomically reliable localizationist picture; rather, they treated findings that sensitivity to statistical regularities was reflected in neural activity across multiple regions, including the medial temporal lobe, to derive representationally diagnostic implications. This area is indeed found to be reliably involved in processes of relational binding, licensing the interpretative stance that statistical learning phenomena operate in a relationally driven logic, being peculiarly apt in forming stimuli-specific association from streams not explicitly segmented (as typical instead of apparently similar paradigms, such as Artificial Grammar Learning). Moreover, these early emerging neural signatures can be dissociated from behavioural data, remaining observable even without being manifestly expressed in subsequent explicit familiarity judgments. This documented divergence is theoretically refining: since the presence of learning-related neural signal does not guarantee that the corresponding representation will manifestly drive behaviour, then behaviour itself must already reflect a process of selection among coexisting representational products.

Further contributing to the exploration of the architectural constraints of statistical learning computations, Schapiro et al. (2013) indicated latent temporal structure of events as an organizing principle of internal representational space. Contexts characterized by community structure are those in which the temporal association schema between single event is patterned in a way that favours the emergence of a clustered cognitive landscape, in which similarity is shared more among elements within than across clusters. Crucially, MVPA analyses of acute learning-induced changes prove that temporal community structure elicits a reorganization of representational similarity that is detectable at a neural level and mirrors the topological proximity of events based on their temporal structure. Therefore, learning doesn't purely adjust relationships among elements, but promote the formation of new and relatively autonomous representational objects altogether. Latent relational structure has a range of possible instantiations, spanning from probabilistic transitions among contiguous elements to higher order entities such as temporal community ones; at this juncture the magnitude of scope reached from the dependency status is critical in regulating the representational form of composing elementary items in terms of the similarity of predictions they can instantiate about the future. Schapiro et al (2017) found that both the analysis of patterns of hippocampal activation/deactivation and functional connectivity become increasingly similar over learning time in the processing of events that can be traced back to the same temporal context, as diagnostic of an ongoing structuring of a temporally organized internal space. Such inferences concerning the significance of relational binding and integrative qualities for statistical learning computations, however, only hinge on correlational evidence, thus can particularly benefit from a more causal validation. Providential in that sense is the clinical case report of Schapiro et al. (2014), that suggest that middle temporal lobe integrity is necessary to provide the cognitive system with the ability of configural, relational processing needed to extract temporal regularities, despite not being essential to discrimination occurring at single item level.

While binding operations are ubiquitous operating feature of the phenomenon of statistical learning, acknowledging representational plurality imply that relational structure may differ in the extent of its combinatorial scope, with integration of non-adjacent elements presupposing operations that exceed the boundaries of purely local associative mechanisms, particularly leveraging memory systems specialized for relational integration. Evidence by de Diego-Balaguer et al. (2016) outline different developmental trajectories leading to acquisition of word segmentation and non-adjacent rule formation. This ontogenetical pattern, although seeming to imply hierarchical complexity of these units of learning, reflects instead an optimal compliance of the maturing temporal attention and cognitive control systems to the currently relevant adaptive request. Under a linguistic acquisition framework, it is a paramount initial demand for the system to learn automatically without limitations. In this sense an earlier-developing exogenous attentional system driven by saliency is particularly suited to rapidly capitalize on conspicuous co-occurrences of cues within a narrow range of space or time. A kind of processing capturing similarity across intervening items and allowing to extract more global gist-like representations hinges conversely on a later forming endogenous attentional system. The representational diversity of the objects resulting from learning is therefore suggestive of dissimilar mechanisms, respectively encoding different levels of input invariance, rather than a progressive acquisition path with the aim to gradually converge on a single enriched representation.

Temporal persistence provides an additional, stringent, diagnostic dimension, that justify using comparability in consolidation and forgetting profiles as a proxy of the identity or divergence of the underlying cognitive representation rather than mere differences in accessibility. Liu et al. (2023) report a dissociation in the pattern of learning traces employing metrics implicitly or explicitly assessing statistical learning outcomes. While the expression of implicit statistical knowledge can strengthen over time, explicit knowledge of the same material tends to decay; furthermore, the effect of repeated testing is selectively beneficial for explicit representations and minimally impacting on implicit learning. Those observed different trajectories, yet again suggestive of a plausible plurality of underlying memory representations, are complemented with evidence concerning the distinctiveness of the qualitative decay pattern of implicit and explicit mnestic traces, an index of their idiosyncratic pathways of statistical structure's prioritization. Indeed, consciously accessed representations show the sharpest decline overtime in sensitivity to detailed features in favour of the consolidation of an abstract gist, while precision of details is strengthened in unconscious, implicit traces.

Although a consistent corpus of literature coalesces in recognizing non-equivalence and non-uniqueness in the cognitive representation resulting from statistical learning, a consequential and especially remarkable question arises. Insofar as functional compatibility is not warranted for representations that encode different regularities, an identical context may realistically support incompatible predictions, leaving unresolved how predictive commitments are formed overtime when multiple options are simultaneously available, how strategical prioritization may differ inter-individually and how insidious could be the mapping from learning to actual behavioural expression.

Chapter 2. A hierarchical organization of temporal statistical learning: predictive horizons, commitment policies and regime governance

Epistemic non-closure and temporal sequentiality: the viability of a predictive processing account of temporal statistical learning

A characterization of the patterns of extraction and stabilization of regularities from probabilistic serial inputs demands specification of additional constraints beyond the hitherto established composite representational nature of statistical learning or the absence of transparency in interpreting behavioural signatures attributable to the unique epistemic affordances of the task. Above all, temporally articulated informational streams introduce an epistemic asymmetry that directly depends on the feature of sequentiality, hallmark attribute not being elicited by spatial configurations, which can in principle be apprehended holistically. The informational status in temporal embeddings, indeed, is progressively disclosed as evidence accumulates and, while behavioural relevance is inherently prospective, only retrospective demonstrations are available to the inferential system. Furthermore, time is not a neutral container of regularities, as temporal parameters shape what can be learned and timing rate of the input governs the access to evidence, determining whether the contingent condition allow learners to acquire sufficiently stable prediction anchors.

Emberson, Conway, and Christiansen (2011) empirically demonstrated that identical arrangements of statistical structure can yield divergent learning outcomes depending on presentation rate, that interacts with domain-specific processing architectures, determining how circumstantial manipulations of temporal rate selectively enhance or impair visual and auditory sensitivities. The significance of their findings, therefore, doesn't merely lie in the observation that timing affects performance, but in how the inferential scope and efficiency is differentially constrained by the integration window within which the process of evidence accrual and stabilization occurs. As a consequence, the same statistical relations do not equally survive across temporal regimes, revealing a constitutive interconnectedness between perception and learning, where modality-specific temporal acuity and overall processing constraints are shared across the perceptual-cognitive continuum, with resultant sharedness of biases and idiosyncrasies.

Moreover, temporal embeddings instantiate a constitutively incomplete kind of knowledge, since the role of each element in a sequence can be fully determined only by subsequent context, which becomes accessible exclusively when the inferential power it endows is no longer cognitively or behaviourally contingent. Non-closure is therefore a distinctive trait of temporal statistical learning: as perception and action cannot wait for the future to become accessible, subjects must commit to strategies of attention allocation, interpretation and response preparation under conditions of informational partiality. This regime determines the relevance of anticipatorily constraining the space of possible futures, prompting a kind of real-time cognition that doesn't construct meaningfulness only at the completion of the sequence, but that exploits calibrated expectations under indeterminacy. Thus, the critical issue becomes how stored dependencies pre-structure the processing of what has not yet occurred, by prospectively narrowing the hypothesis space of forthcoming events before the validity of the extracted regularities guiding processing is retrospectively confirmed. Accordingly, introducing a predictive vocabulary is not an unnecessary pretentiousness, but an explanatory valuable assumption about the medium through which learning constructs operational efficiency in time. Such premise, however, demands a disciplined specification, as the use of the term "prediction" in cognitive literature has ranged from postulations of simple covariance-based anticipation to strong Bayesian commitments to a "Grand Unified Theory", framing overall cognition as a process of hierarchical generative inference.

Particularly instructive on that issue is the work of Clark (2013), as it complements the formal attempt to define the framework of predictive processing with a cautionary boundary. Conceptually, the author assumes a Helmholtzian posture in designating perception as a process of probabilistic, knowledge-driven inference, conceivable as an inverse model of analysis-by-synthesis. He postulates that the expectations generated by higher-level hypotheses are mapped to lower-level sensory input: what is expected is attenuated and only what resists expectations is treated as informative, as it signals significant residual mismatches between top-down anticipations and bottom-up evidence. Accordingly, only those warnings will be allowed to effectively propagate to superior hierarchical levels in the form of prediction errors. Notwithstanding his clear adherence to the framework, Clark doesn't conceal the debatable ambiguity of the term "prediction", the blurred boundaries of its scope and the potential danger of unfalsifiable elasticity it brings. Recognizing such ambiguities, the present argument does not specifically commit to a certain algorithmic implementation or claim an empirically settled doctrine of a Bayesian-optimal brain. Rather, I will adopt a predictive vocabulary and conceptual framing insofar as it is structurally legitimised by the intrinsic nature of temporal statistical learning.

The "black box" analogy (Clark, 2013) proves particularly useful in illuminating the outlined perspective: due to the virtual inaccessibility of the external world, the only attainable knowledge for our cognitive system entails its own internal states, leveraged by our neural architecture to formulate a guess on the next state of our subjective internal economy. In the attempt to optimally match the statistical properties of the signal source, each learner's internal generative model is progressively refined through gradient descent learning, aiming to ultimately increase its posterior probability. Notably, in this framework, neural representations need to reflect the feature of uncertainty: this is achieved by a system that, in the process of negotiation of multiple neural representations' relative plausibility, encodes each of them as a density function. Since in probabilistic environments deviations from expectation are inevitable and uncertainty is never eliminated, a refinement of what is meant by "anticipatory constraint" is mandatory. Contexts dominated by stochasticity, where evidence is disclosed incrementally in a noisy regime make the mere presence of an error not sufficient to determine a reorganization of the current predictive asset: the crucial question shifts from whether a mismatch occurs, to whether it should matter. Further, the learner should reach an optimal balance policy between stability and flexibility: a cognitive system constantly updated risks to collapse structure into volatility, one updated not often enough is adaptively inert.

Yon and Frith (2021) clarify this point by formalizing predictive precision as a process of weighting prediction errors occurring during belief updating and hinging on subjective contextual assessments. Specifically, if the environment is deemed as stable, even small deviations may signal meaningful change, while, if the context is considered volatile, the same deviations may be treated as irrelevant noise. Accordingly, divergent beliefs about the relative precision of bottom-up evidence or top-down predictions determine different updating strategies that depends on the relative allocation of reliability, thus confidence, to the two contradictory informational channels. Noteworthy, precision itself is estimated and can be potentially unveridical due to subjective miscalibration, with individuals that disproportionately rely on prior structure generating a condition of self-fulfilling prediction and systematic downweigh of meaningful deviations, whereas excessive sensitivity to error can induce behavioural immobilism due to perceived volatility in belief states.

A clarifying expansion on the feature of "constructivism" in generative formulation of perception is provided by Clark (2013), whose work underlies how a system operating under temporal uncertainty must implicitly gauge not only the content of prediction and the reliability of the "prognostic" informational channels, but also the temporal scale to commit to. The hierarchical character of prediction mirrors how the environment affords multiple layered temporal statistical structures, with each level diverging in the temporal horizon over which prediction must operate. The typically longer integration window characterizing global regularities determines the need of more extensive evidence accumulation before they can achieve representational solidity, while local regularities can stabilize over shorter spans. Besides acknowledging distinctiveness of evidential threshold, a cognitive mapping of this global/local dichotomy is proposed. Lower perceptual levels are postulated to operate over finer, detailed scales and higher predictive levels are predicted to construct coarse-grained internal models summarizing regularities over broader temporal scales. The author further highlights how the proposed hierarchical brain organization blurs the boundary between perception and cognition: if the priors generated by higher-level models constrain lower-level processing and if precision weighting reflects the subjective balance between estimates of environmental stability and top-down prediction reliability, then individuals will display broad variability both in their perceptual and learning commitments. Specifically, overweighting higher-level priors can lead to inflexible reliance on global structure, preferential gist-like extraction of the material to be learned and cognitive tolerance for transient ambiguity; conversely, heavily leaning toward lower-level error signals can produce disproportionate sensitivity to local representations and resistance to interference from higher-level bias.

Attempts such as the one advanced by Feldman and Friston (2010) validate the increasing theorical relevance of the concept of precision weighting, by complementing the phenomenon with a physiological mechanistic perspective. Specifically, attention is conceived as an inferential system performing gain-control over prediction-error signals (Clark, 2013; Feldman & Friston, 2010) through modulation of the synaptic gain of sensory evidence. The delineated inferential landscape appears therefore to be second-order (Heilbron and Chait, 2018), with the cognitive system generating predictions both concerning "the content" (i.e. represented statistical regularities) and its reliability, resulting in continuous and relatively arbitrary processes of calibration of confidence.

Despite controversies in acknowledging predictive coding as a comprehensive theory, specific tenets of this account are clearly supported by influential empirical evidence. Exemplar is the assumption of prediction not occurring in a post-perceptual layer but being intrinsic to temporal perception itself. Heilbron and Chait (2018), reviewing evidence from auditory cortex, illustrate multiple documented data legitimising the assumption that the system does not passively register isolated elements, but processes each item against a dynamically maintained model that strives to account for the likelihood of what will happen next. Bolstering the assumption of an anticipatory nature constitutive of perception and confuting a perspective of inference as a subsequent cognitive layer are reports substantiating how the long-term input history influence response to single items or paradigms where omission of an expected stimulus elicits neural activity as though the predicted event had occurred.

At a higher level of abstraction, the postulation of the free-energy principle (Friston, 2010) corroborates accounts claiming the structural necessity of anticipatory regulation for adaptive systems, a mechanism that is proposed to transversally operate across action, perception and learning domains. The author posits that self-organizing organisms, when faced with constantly changing environments, are subject to the homeostatic imperative to minimize the long-term average entropy through a short-term effort to avoid surprisal, hence, free energy. Perception and action can be framed as complementary routes to this end: either updating internal models to better account for sensory input or sampling the environment in ways that confirm prior expectations.

Altogether, the pool of notions hitherto outlined instruct a conceptual shift from "learning as acquiring" to "learning as anticipatorily deploying". Temporal statistical learning supplies structured regularities, but the epistemic non-closure afforded by the feature of sequentiality preclude conceiving learning solely as the stabilization of that content. Prediction is constitutively required to regulate individual commitment, thus epitomizing the operational bridge between stabilization and deployment. Once this is acknowledged, some ensuing matters require further elaboration: first and foremost, the extent and principles that constrain the horizon over which prospective inference is exerted by prior structure.

Hierarchical predictive horizons: error propagation geometries, falsifiability profiles and density revision constraints

Once posited the role of prediction as an operational bridge that allows learned regularities to exert causal influence under temporal non-closure, the consequent, non-trivial step is to consider the temporal scope of their extension. Indeed, prediction is not a pointwise operation, immediately attached to the subsequent element; it is necessarily indexed to a span: anticipatory expectations wouldn't be conceptually distinguishable from a sort of hindsight bias, expressed in a pervasive, subjectively partial interpretative stance, if their processing did not operate over a definable portion of forthcoming processing. Prediction in temporally articulated environments demands the specification of its prognostic horizon, as the span of propagation of the anticipatory commitment brings along consequences in terms of diagnostic delay, degrees of freedom in the attribution of causality, susceptibility to error and extent of functional relevance of the input.

In order to designate this predictive radius, Lerner et al (2011) define a notion conceptually paralleling the spatial domain: temporal receptive windows (TRW) quantify the portion of elapsed time during which sensitivity to sensory input retain sufficient power to affect the forming response. The authors employ linguistic material progressively scrambled at different narrative scales to probe the regionally specific temporal dependence in neural responses as inferred from the reliability of each area's responsiveness to distinct units of meaning and their relative violation. The effort to cortically map the specific boundaries of the pre-structuring power over future processing revealed a hierarchical neural arrangement, integrating contextual information over systematically increasing timescales: from early auditory areas not affected at all by the coherence of the input's temporal history and higher-order regions exhibiting increasing dependability on temporally extended stimulus' consistency. Under an epistemological perspective of propagation span, statistical adjacencies might be conceived as encompassing a short predictive horizon characterized by a narrow propagation radius and entailing a locally confined scope of constraint; non-adjacent dependencies, conversely, populate an extensive horizon yielding a broad propagation radius that imposes a constraint field distributed across intervening material.

Hasson et al. (2015) dovetail those neurally sketched principles with their computational features, proposing to abolish an outdated differentiation between information storage and information processing and meaningfully shifting the explanandum from "systems of memory" to "memory of systems". They propose a "process memory" framework that formalizes the claim that information processing accumulates evidence overtime virtually ubiquitously at the cortical level, actively shaping real world cognition, and perception, under multiple receptive windows, rather than being segregated in some sort of mnestic buffer.

Once prediction is conceptualized as occupying a field in time, one connotated by a specific spatial extent, a first structural consequence follows: even falsifiability acquires a latency. Short-horizon commitments entailed by adjacent contingencies, operate under conditions of rapid accountability, whereby the presence of confirmatory or disconfirmatory evidence lies in the temporally proximal context; on the other hand, engaging to non-adjacent dependencies constitutively imply to defer confirmation due to intervening elements that temporally insulate hypotheses from immediate adjudication. The rhythm at which the space of hypotheses contracts and is updated depends directly on the nature of the statistical structure present in the input, as well as accountability depends upon the depth of temporal dependencies. When statistical regularities are locally adjacent, the cognitive system operate its predictive commitment in a regime incrementally dissipating uncertainty: here tests of possible mismatches between prediction and reality of the input swiftly occur due to the inherent proximity of cues scaffolding structural information and exposure to disconfirmation is so frequent that the inferential system can revise with minimal temporal separation between hypothesis' formulation and adjudicating evidence. Nonadjacent dependencies, by contrast, induce structurally underspecified, yet hardly alterable commitments, formulated during an interval where the system must continue to process input in the absence of feedback to evaluate its accuracy. Since relevant evidence accumulate over structurally interfering items, the state of epistemic suspension induce potentially more volatility in learning but paradoxically greater resistance to updating once a regularity is seemingly discovered, due to the necessity to (sub-optimally) crystallize belief in an unreliable context. Notably, the delay of falsifiability instantiated by statistical non-adjacencies impacts the attribution of the meaning of error itself, hindering immediacy and reliability in the causal ascription of a mismatch. In short-horizon regimes, failures in prediction can be swiftly localized, as test-points of revision occur at a consistently dense rate, while in long-horizon regimes, the longevity of interpretative distortions cannot be precisely estimated, and multiple explanatory loci are compatible with remote fallacies. Accordingly, the system cannot definitely resolve the source of the error and keeps oscillating between hypotheses of misestimation of the underlying statistical relation, misidentification of the relevant unit of meaning, or stochastic fluctuation that is inherent to a probabilistic input and doesn't concretely undermine the reliability of the extracted dependencies. The geometric horizon of commitment and its depth might therefore introduce differential effects on "causal dispersion" and divergent learning patterns, owing to how the range of possible statistical structures afford distinct evidential availability and temporal frame between prediction and violation. The main concern for the system shifts from the fact that a prediction failed, to the definition of a plausible topology of failure, raising the issue of what and at which level the extended causal chain should be revised. The present argument, however, needs some parsimonious specifications. As already stated, long-horizon commitments are not assumed to be inherently more complex or abstract: they are supposed to instantiate a distinct epistemic regime, whose structural consequences ensue from temporal distance alone and whose degree of adaptive fitness cannot be presumed.

Differential predictive constraints forming under different extent of commitment horizon don't merely affect the interpretation of already occurred evidence but govern a range of typical inferential trajectories extending into the future. Locally recurrent diagnostic checkpoints typical of short-horizon regimes allow to disconfirm predictively inadequate interpretations before they can accumulate substantial inferential inertia, exerting enduring and maladaptive systematic distortions by lingering upon subsequent perception. Belief update proceeds in tightly coupled cycles of prediction and revision, minimizing the temporal separation between commitment and potential violation detection. On the other hand, the temporally extended insulation from disconfirmation distinctive of non-adjacent statistical structure, confronts the system with some critical computational challenges. Under conditions of diagnostic postponement, the space of hypotheses potentially viable can't be promptly pruned, compelling the system to retain competing relational interpretations that proliferate in an unresolved state. As a result, elimination is no longer continuous and continuously assessed; it becomes punctuated, assuming either features of update resistance where hypotheses acquire temporal "immunity" simply by virtue of span or a characteristic constitutive volatility. Both such outcomes ensue from a computational condition of combinatorial proliferation, whereby some boundaries must be set to limit the operation of a computational device that would otherwise face an unmanageable explosion of the learning problem (Newport and Aslin, 2004). Distal commitments afford a level of abstraction over intervening material, with the latter, while not being locally diagnostic of the relation occurring among predictive and predicted elements, introducing, however, a set of degrees of freedom depending on its extension. As the distal endpoint disambiguating prediction moves further away and the predictive horizon increases, so does the number of configurations, all partially consistent with the accumulating evidence, that the system extracts from the intervening material, determining an expansion of the pool of actively maintained relational possibilities. Newport and Aslin's (2004) work on non-adjacent dependencies illuminate this point demonstrating that learners can acquire distal relations, yet such acquisition is firmly constrained by the very nature of the representational unit of learning. Exemplarily, they show how statistical computations over nonadjacent patterned regularities particularly suffer from being enveloped in hierarchically superior structure: demonstrations of statistical learning don't extend from isolated phonetic segments to syllables, whose compound nature hinders the immediate recognition of constituent elements as belonging to the same informative pool. As representational units in non-adjacent dependencies require themselves a process of internal composition, the system's inferential burden exceeds its computational capabilities.

The pressure to actively retain multiple viable predictive hypotheses in the absence of adjudicating evidence poses an algorithmic ordeal that constitutes a discriminating factor in subjective trajectories of learning outcomes. Not only the individual attentional span and the extent of operational efficacy of statistical learning mechanisms, but also tolerance of ambiguity, resistance to interference, individual flexibility and degree of acceptance of failure are called into question and determine an enhancement of variability in the range of subjective learning outcomes of non-adjacent structure. Extended predictive horizons seem therefore to potentially jeopardize our representational economy through exponential proliferation, unless some mechanism of structural gating is leveraged to restrict the extent to which relational encodings are allowed to propagate.

A fruitful avenue is to instantiate strategies of dimensionality reduction by systematically discarding interposing details irrelevant to the distal dependency, while recoding invariant relational structure into meaningful units. Evidence from chunking-based serial recall paradigms (Isbilen et al., 2022) suggests that non-adjacent learning fundamentally hinges on the formation of predictive units. To disambiguate between chunking-like processes that induce compressed, wholistic word-like representations from a mere positional encoding, the authors probed sensitivity to multiple levels of regularities in a task explicitly targeting non-adjacencies and interpreted findings of generalization beyond specific encoded sequences as efficient extrapolations of consistent relational frames, flexible enough to transcend over specific exemplars. Once surface variability is collapsed into a compressed schema, the system operates a hierarchical prioritization that consists in purposefully loosening the computational effort over the full combinatorial detail of the not predictively diagnostic intermediate states, while allowing the system to successfully propagate a distal commitment. Under this perspective, compression becomes a feasibility condition that adheres to the need to counteract uncontrollable proliferation imposed by depth.

Crucially, how the temporal horizon of adjacent and non-adjacent dependencies differentially modulate error attribution latencies and combinatorial expansion don't determine a fixed, preferential viability of one regularity over the other. Instead, environmental structure dynamically redistributes the relative inferential advantage of statistics differing in temporal depth, making short and long-range commitments comparatively stable according to the adequacy they warrant in facing contextual requirements. The temporal horizon of statistical structure is not an intrinsic property of the input or the learner alone, with its specific suitability conceivable as an emergent property, contingently arising from the relational dynamics between the principles of the subjective inferential architecture and environmental structure.

Particularly instructive about the logics through which epistemic stabilization of acquired temporal relations emerges as a function of contextual features is the work of Gómez (2002). As already shown, their manipulations of the breadth of variability in the pool of intervening items amidst non-adjacent dependencies resulted in observable shifts in the default perceptual dominance, with increases in middle elements' surface variability selectively enhancing non-adjacent dependencies, while systematically impairing learning grounded in adjacent contingencies. A particularly fruitful approach to such empirical evidence addresses how scaling manipulations of intervening variability magnitude induce a reorganization in the reliability of diagnostic density policies to optimize the relative behavioural dominance of competing evidential evaluation policies leveraging temporally shorter- or longer-range statistical structure. The cognitive architecture, faced with a relatively stable environment, conveniently support locally adjacent structures that are continuously refined through dense feedback rather than temporally deferred adjudications: under contextual low variability the signal-to-noise ratio of local mappings is high, supporting quick and reliable processes of hypothesis contraction, refinement and elimination, not penalized by protracted latencies of indeterminacy. Increasing variability reorganizes this landscape: due to the broader range of fluctuations in intervening elements, the signal-to-noise ratio of proximal contingencies deteriorates, owing to the increasing frequency of comparatively less informative mismatches: the density of diagnostic checkpoints increases arbitrariness and unpredictability due to the proliferation of error signals devoid of disambiguating potential. Distal commitments, conversely, become insulated from this local instability, insofar as they constitutively adopt a policy endorsing sparser check events tied to invariant endpoints in striking contrast with the surface contextual fluctuation. The simplistic assumption of variability directly strengthening long-span learning can thus be better reformulated as the inferential equilibrium's compliance with contextual requirements, with flexible redistributions of predictive power determined by how distinct span-determined topologies optimize signal-to-noise ratios contingent on situational attributes.

Equally valuable to the sensitivity of a cognitive architecture to environmental structure system is the faculty to contingently adjudicate differential degrees of adaptiveness between an inferential posture, privileging immediacy in commitments and recurrency in its verification rather than a diagnostically delayed predictive stance relying on distributed invariants. Indeed, each temporal horizon affords specific strengths and distinct vulnerability profiles, with different spans being particularly resilient against certain forms of noise while amplifying exposure to others.

Once environmental structure is postulated capable of inverting the relative viability of distinct spans depending on the contextual significance of distinct features of prediction and to do so with flexible dynamism, then there is no principled reason to assume that only one temporal span operates at a time. Multiple predictive timescales may coexist and be concurrently processed within the same informative stream: within a single predictive regime, structurally distinct commitments across more than one temporal receptive field might be sustained simultaneously. Such coexistence is substantiated at an architectural neural level by Ding et al.'s (2016), neurophysiological findings that demonstrated concurrent entrainment of neural activity to hierarchical linguistic structures unfolding at distinct temporal scales, from syllabic rhythms to larger phrasal and sentential units. A carefully crafted manipulation was devised by the authors to experimentally dissociate the transitional probabilities of lower units from higher level structure, in order to exclude that responses taken as indices of processing occurring at a hierarchically superior level (i.e. sentential level) were a mere epiphenomenon of enhanced predictability of smaller units, developing under conditions of protracted exposure afforded by larger linguistic structures. Crucially, entrainment at slower scales was found to be not reducible to cumulative sensitivity to adjacent transitional probabilities; reflecting instead an internally constructed sensitivity to a kind of structure whose temporal scope differs from the one issued by immediate co-occurrence statistics. Not only the system adheres to contextual requirements of multilevel responsiveness to temporally articulated structure by forming representational plurality; neural phenomena suggest simultaneous span-indexed tracking, with a degree of overlap not precluding computational independence.

Behavioural evidence (Deocampo, Conway, & Smith, 2019) converges in supporting the assumption of a system capable of maintaining multiple relational description, instantiated at distinct predictive timescales and concurrently populating the same informative stream. However, predictive commitments that fostered distinct temporal propagation radius revealed a robustness asymmetry, with interfering perturbations yielding differential effects in the retention of adjacent and non-adjacent dependencies. Specifically, the authors proposed, without perceptually highlighting the different nature of coexisting input, experimental material consisting of visuo-spatial and visuo-verbal sequences where adjacent pairs were embedded within nonadjacent ones and each regularity implemented independent deterministic associations. Notably, this manipulation seems to be located at the conceptual opposite pole of the one adopted by Gomez: rather than favouring acquisition of distal dependencies by broadening intervening items' variability, here the two kinds of statistical structure are acquired in a complementary fashion, with the degree of invariance embodied in the transitional probability of the embedded unit favouring the relative segmentation of the two streams. Following the initial phase of exposure, an additional experimental session was proposed, during which instances of sequences arranged in an ungrammatical fashion were introduced, along with sentences adhering to the "grammatical mapping" previously acquired. This operation, therefore, could afford the potential of endangering the deterministic nature of the original mapping, if the repetitive presentation of "ungrammatical" counter-examples reveals to be sufficiently effective in stabilizing novel units as meaningful instances of learning. Adjacencies are shown to be comparatively more impervious to disruption by anomalous exemplars, relative to the correspondent fragility of non-adjacent dependencies, suggesting, once again, that commitments at different spans predict differential vulnerability profiles to interference and contextual manipulation. For this specific case, a set of principled justification of this empirical evidence is proposed by the authors. First, a difference in the time frame of stabilization dynamics is proposed to advantage adjacent dependencies, whose rapid dynamics of consolidation make them impermeable to manipulations subsequently occurring. Furthermore, chunking accounts and competitive logics are explanatorily combined to interpret the inferior resistance of non-adjacencies: as their acquisition proceeds within protracted timeframes, units of learning maintain flexibility, remaining provisionally unresolved in order to accumulate greater evidential mass. Furthermore, non-adjacent representations progressively become more order-sensitive, instantiating a competitive regime whereby encountering individual stimuli in sequences different from those previously experienced undermines the maintenance strength of the latter.

Within the present framework, coexistence acquires a precise inferential meaning, with each timescale-determined predictive horizon embodying a unique configuration of falsifiability latency, policy of constraint density, combinatorial proliferation, and robustness/vulnerability profiles. When multiple horizons are active within the same informative stream, the system doesn't merely represent statistical structure at different "levels"; it rather operates under multiple simultaneous epistemic regimes, which are not reducible to one another, nor define sequential stages of a single process, but constitute instead qualitatively distinct modes of predictive control. Coexistence, however, introduces tension, insofar as multiple predictive timescales may differ in how and when they are falsified, can generate anticipatory constraints that are respectively compatible, orthogonal, or divergent or, even when predictions do not directly conflict, they may vary in confidence depending on contextual variability and interindividual differences in cognitive dispositions. Statistical learning, under this perspective, can be conceived as the configuration of span-differentiated temporal commitments whose relative viability is context-dependent and where the inferential advantage is dynamically redistributed across spans. Having outlined both the computational characteristics afforded by different hierarchical levels of temporally articulated statistical structure and their distinct adherence to specific environmental requirements, it remains underspecified which internal principle regulate the arbitration of weighting, selection and behavioural deployment amongst the representational plurality of predictive horizon.

From predictive regimes as latent organizations of the stream to behavioural governance

Having established that temporally structured environments can simultaneously scaffold a plurality of viable interpretations and that such multiplicity can coexist within individuals' cognitive architectures doesn't however solve the inferential problem that the learning agent must confront: he is called to determine, among representational candidates, the one that is advantageous to adopt as currently operative and to continuously track its momentary adequacy in order to adjudicate the relative convenience of policy maintenance versus strategy revision. The level of structural relationship presently entrusted as explanatory would conformably establish the scope of informativeness of featural configurations of the stimulus, accordingly tune and allocate individuals' sensitivity, configure the parsing process of the informational stream and sculpt the perceptual and interpretative framework. The inferential regime does something with deeper implications than purely detecting the level of statistical dependencies currently structuring the environment: it establishes the degree of coherence of the system's assumption with the informational stream, the extent of evidence-bending allowed to reconcile discrepancies that are still considered tolerable and the threshold that, once crossed, forces to reappraise deviations as meaningful violations of expectation.

Starting from the process of evidence accumulation, learning trajectories emerge as history-dependent estimations of the sensory stream, whose latent organizational principles are extrapolated in a not neutral fashion, with the significance of new observations being constitutively constrained by the ongoing interpretative framework. Gebhart, Newport, and Aslin (2009) proposed a set of principled determinants of such temporally asymmetric learning dynamics, starting from individual's stationarity biases, which entail their proclivity to conceive structures as rarely undergoing rapid and frequent changes. In natural environments where structure coexist with noise and only a fraction of the available evidence is predictively instructive, the ideal observer's approach would be to compute variance over a sample of exemplars large enough to determine the minimal subset yielding asymptotically low fluctuations (Gebhart et al., 2009). However this solution is generally disregarded by learning agents that draw often provisional and tentative conclusions from small evidential pools, typically leading to garden-path errors whereby initial commitments induce to show a remarkable resistance to update early-formed representations, only overcome when a consistent amount of countervailing evidence is gathered. This cognitively robust and pervasive phenomenon is probed by the authors in the statistical learning domain through a series of seminal experiments, exposing learners to artificial streams, whose underlying statistical structure was generated by non-uniform distributions and where midstream transitions in the informational nature of the stimulus were not signalled by low level perceptual cues. Crucially, successful extraction and commitment to the early portion of the sequence exerted a "primacy effect" resulting in the blocked acquisition of the patterned relationship instantiated by the second informative stream. Nonetheless, under conditions of perceptually marked shift between the two distributions, or when the time of exposure to countervailing evidence was sufficiently protracted (tripled compared to the earlier portion's duration), garden-path phenomena were shown to be reversible. Error-recovery processes were indeed observed and led to an above-chance performance for testing items originating from both the streams: unveiling how implicit error monitoring processes were already taking place to evaluate the fit between the current data sample and the ongoing representational structure. A plurality of representations was therefore maintained but required a different evidential threshold to be overtly displayed in conscious perception, with the primacy bias constituting the optimal solution to prevent inefficiencies due to delayed commitments in computing structure, while preserving processes that allowed to adjust ongoing learning. Under this perspective, garden-path effects result from a predictive regime filtering incoming evidence through its explanatory framework, with early observations playing a pivotal role in determining which regime becomes active and constituting an interpretive scaffold where stabilization of predictions occurs by constraining how new information is incorporated.

Path-dependence is not unique to artificial grammar paradigms and the extent to which perceptual processing can be conceived as shaped by the currently privileged predictive framework is provided by studies belonging to the line of research of Event Segmentation Theory (EST). Zacks et al. (2007) advanced the notion of "event models" as working memory representations neurally implemented in transient activation's changes, where the convergence of multimodal input from semantic memory representations renders the currently adopted internal model a sort of schema, recapitulating previously learned information about the sequential structure of similarly occurred activities. Perceptual constancy is achieved through the compound action of currently re-instantiated long-term scripts and gating mechanisms particularly robust to variability that ensue from the sensory pathway, which is allowed to exert a meaningful influence on the event model only when prediction mismatches exceed the threshold dictated by the error detection system. A cognitive architecture whose essential purpose is to preserve resources by maintaining representational stability to volatile environmental fluctuations shifts to the more energetically taxing process of updating only when the persistence of prediction errors is considerably suggestive of environmental unreliability of the interpretative framework. The detection of boundaries between events represents the epiphenomenal resultant of a predictive organization of perception, where reset and update processes elicit the subjective experience of episode segmentation. The domain of statistical learning emerges as an instance of the cognitively spontaneous tendency to segment the informational stream into meaningful units, or internal event models, hinging on inferential hypotheses about the latent generative organization of experience.

While moment-by-moment adherence to a predictive regime pivots on the dominating inferential prior, sequential streams scaffold a plurality of potential segmentation logics that unfold in a nested fashion, with the temporal dynamism in the coexistence of different levels of abstraction demonstrated by the neural signatures of multi-timescale learning (Maheu et al.; 2019). To determine the most likely timescale of brain perceptual inference, they used a reversed-engineering approach: they compared individuals' levels of surprise, as inferred from MEG-recorded auditory responses to violations at different temporal scales, to theoretical levels of unpredictability computed from a family of learning models that differed in the temporal span of previous' evidence integration used for inference. The operationalization of prediction errors leveraged the long-term average quantity of Shannon's surprise emerging from discrepancies between predictions and actual observations, thus allowing to assess the relationship between latency of the neural response and timescale of integration. They found a correlation between long latencies in neural responses and processing units spanning in their aggregation capacity up to a maximum of 5 items (typical working memory's range), displaying an unusual reversal of the classically emerging association between long-timescales of integration and late phases of responses. This conundrum was explained by the authors by appealing to the dynamicity of coexistent, temporally hierarchical levels of inferential operations allowed by the sensory sequential stream. Since traditional paradigms typically exposed participants to pre-segmented chunks, an already structured level of abstraction was provided and could be leveraged to scaffold research of patterned statistical information over longer ranges; instead, the continuously unfolding stream employed by the authors precluded the immediate access to hierarchically superior computations and subjective inference operated over a smaller unit of analysis, specifically, in a span entailing up to 5 preceding items.

Mirroring the dynamism exhibited by the integration windows of latent temporal organizational principles, a degree of fluidity is manifested also by the processes of evidential accumulation, that may differentially govern the dynamics of how subjects enter on a predictive regime and exit from it. Neural evidence from Barascud et al. (2016) converges with the behaviourally assessed unevenness between early structural commitment and policy revision of such adherence, as assessed by the observed tendency of learners to require only modest evidence to install coherent latent predictive hinges, yet substantially stronger, perceptually enhanced, or temporally protracted evidence to relinquish it. The experimental material employed by the authors comprised a temporally unfolding auditory material encompassing transitions from randomly arranged to regularly patterned phases and vice versa: such design allowed to compare the differential logics underlying neural processing in experiencing shifts from illogical to meaningful segmentation rather than exposure to the disruption of such expectations. As indicated by the dynamics of modulation of MEG signal, the subjects required an amount of time closely adhering to the lower bound expected from a Bayesian ideal observer to develop sensitivity to emerging regularities, with less than one full repetition of the cycle eliciting an increase in the amplitude of evoked responses, signalling increased predictability. Conversely, mismatch responses to transitions violating already established regularities displayed, in their neural signatures, properties instructing at least two legitimate interpretations. Elicited responses occurred at latencies consistent with those of MMN, licensing an interpretation of the observed drop in amplitude as the down-regulation of the sensory units that signalled incompatibility of the signal with the existing internal model, thus reflecting the attempt to adjust discrepancy-driven surprise until it didn't exceed a tolerability threshold. Further, behaviourally documented asymmetries in temporal dynamics of acquisition/update were mirrored in the latencies of emergence of violation signals: these ones resulted delayed compared with what an optimal Bayesian detector would have predicted, thus meaningfully diverging from the optimality exhibited in patterned regularity acquisition. Overall, the speculation of a cognitive architecture that, in arbitrating the modulation of precision weighting of top-down and bottom-up evidence, shifts the equilibrium toward discovering patterns and against losing acquired structure seem to be empirically tenable.

At this juncture, it may be useful to adopt from the psychophysical literature the concept of "history-dependent hysteresis", a phenomenon robustly assessed under conditions of uncertainty and that epitomizes how attributions of informativeness to analogous objective deviations may differ due to the underlying inferential asset. In probabilistic environments, where prediction errors may equivalently signal meaningful discrepancies with ongoing inferential posture or stochastic variability, this notion seems explanatorily more compelling than simple system's resistance to change or representational inertia. In the cognitive economy, abandoning a predictive regime in response to every mismatch under an invariably applied assumption of occurred structural change would lead to unstable inference and misallocation of processing resources. Instead, learners must determine the most parsimonious trade-off in choosing among inferential error-reduction policies that frame discrepancies as evidence that the regime has become inadequate and must be replaced rather than energy-maintenance strategies that privilege the interpretation of anomalies as random noise still warranting preserved credibility for the current regime.

Instructive about the challenge faced by a learning agent called to define an optimal balance between minimization of computational cost and maximization of the accuracy of prediction is proposed by Lynn et al (2020). The acquisition of higher order structure of statistical relationship is treated as a useful exemplum of how to optimize the accuracy-complexity trade-off by exploiting the free energy principle. The authors suggest that the most successful strategy of acquisition of higher-order dependencies is not to implement a sophisticated hierarchical learning algorithm; rather, their internal representations should emerge from a cognitive system whose architectural entropy is maximized, meaning that their internal models should be maintained at the simplest possible level, namely, minimally constraining their computational status. The authors drew from graph theory and community structure evidence (Schapiro et al., 2013) and complemented such approaches with a modelling attempt to describe the human memory distribution: they demonstrated that a modular arrangement of the network of statistical relations, with densely interconnected nodes and progressively fading edges constituted the topological arrangement allowing acquisition of higher order structure. An essential property in such a framework was the intermediate level used to model the parameter of precision assumed to characterize humans' memory: rather than an infallible, extremely refined recollection device, the model assumed mnestic liability to mental errors, whose temporal trajectory assumed a typical graphic form. Retention fallibility was the paramount feature in allowing the dynamical shift in internal estimates' timescale of transition structure, owing to the progressive vanishing of local dependencies and ensuing availability of temporally broader community-structure patterns.

A different declination of principles allowing the optimization of the resource-accuracy compromise is posited, still within the framework of Bayesian inference, in models of latent structure inference (Gershman & Niv, 2010). They postulated that an equilibrium in the state space of concurrently maintained inductive hypotheses concerning the relations between observable variables and their most explanatorily adequate causal antecedent can be achieved by adhering to the rule of fewest possible latent organizational variables. Learners are assumed to treat a sequential sensory input as arising from hidden states, each implementing its own statistical regularities and implicitly encoded by a probability distribution that models subjective priors about their relative sustainability; the ultimate determination of the latent model of the present situation is reached through an update process of the relative plausibility of probabilistic beliefs via observational data. The authors phylogenetically ground their proposal with the interpretation of post-extinction renewal of distress within reinforcement learning paradigms: merely putting back conditioned rats where they originally experienced shock elicited fear reinstatements that were framed as contextual re-establishment of an inferential regime. Such findings, as well as seminal ethological evidence (Knudsen; 1998) demonstrating rapid adaptive processes in adulthood-occurring reintroductions of perturbative sensorial experiences originally encountered, and extinct, during initial phases of the sensory sensitive period, concur to strengthen two interrelated interpretative hinges. In the logics of a compromise between efficient processing and computational complexity, error-recovery mechanisms asymmetrically privilege the mnestic retention of earlier occurred acquisition: even though subsequently disconfirmed, demonstrations of the quiescent availability of original learning outcomes suggest that shifts in the meaningfully structuring unit of perception don't remove previous commitments. The persistence of predictive regimes in learning mechanisms reflects a rational solution to the inferential problem that emerges from environmental stochastic variability: occasional mismatches can be generated even under a correct model, favouring the development of a certain degree of tolerance to error. Once again, historically driven-phenomena of hysteresis in perception and mnestic retention seem to be supported by empirical evidence, not excluding, at the same time, that learners must remain sensitive to patterns of discrepancy meaningful enough to signal the emergence of a new structure.

The balance in preserving resources and entrusting the correctness of inferential processes is a ubiquitous computational principle that appears to be operating at a distinct unit of analysis and to be particularly relevant for cognitive endeavours that are more demanding and pervasive. As previously established, the dynamics of statistical learning cannot be reduced to a simple accumulation of prediction errors; instead, learners must continuously determine the juncture at which persisting in the current organizational hypothesis transitions from being computationally effortless to becoming unbearably taxing. The ability to establish that improving the quality of the current inferential regime by incremental refinements of its parameters is computationally inadequate ahead of the possibility that the sequence is generated by a different organizational hypothesis altogether is a necessary requisite to achieve optimal statistical learning. However, a plethora of essential substrate prerequisites allow the working logics of this mechanism, first and foremost the aptness to infer the source of prediction error, which in turn depends on whether discrepancies accumulate in a way that undermines the explanatory coherence of the current model. Such evaluation is inextricably tied and critically depends on the learner's beliefs about the temporal stability, or lack thereof, of the environment.

A formalization of the thresholded adjudicating process between regime maintenance versus regime revision is provided by Nassar et al. (2010). They used a novel estimation task to quantify how human subjects updated beliefs when confronted with environmental noise and variability and, specifically, how their choice of learning rate from sensory evidence was influenced by the magnitude of error relative to what was expected, the recency of such an unpredicted change and the relative uncertainty of the learning agent. First, the degree to which estimation errors were considered surprisingly large was contingent upon the variance of the generative distribution supposedly generating the stimuli: in contexts estimated as more stable, the absolute magnitude of an error that would have been adjusted within more noisy environments, elicited instead a tendency to increase subjects' learning rates. Such a policy adhered to an ideal Bayesian observer model, predicting that at change points, reliance on previous evidence should be discarded as no longer useful to interpret the novel pattern, enhancing comparatively the processing of newly arriving information. Emblematically relevant in prediction becomes the concept of hazard rate (Nassar et al.;2010), representing the expected probability that a change point has occurred at any given moment, with higher hazard rates implying relatively frequent regime transitions, biasing the proclivity of learners to attribute discrepancies to structural change. Conversely, a lower hazard rate implies a more stable environment, encouraging learners to interpret discrepancies as noise within the current regime. Consequentially, one should predict that, under situations where estimates about underlying structure were reported as more uncertain, subjects should be inclined to favour the elaboration of new sensory evidence in order to reach a degree of understanding of the environmentally dominating statistics. Crucially, the relative distribution of processing resources favouring accumulated or new evidence depended on the interaction between both hazard-rate beliefs and the learner's uncertainty about the current regime. When the observational sample of the learner, or run length, was short, a paucity of evidence could be exploited and uncertainty about the reliability of that regime remained high: here the assumption of a positive correlation between uncertainty levels and learning rates is substantiated. Conversely, when a regime has governed interpretation over a long run of observations, confidence judgments concerning the variance of the generative distribution allow to likely treat discrepancies as stochastic variation, instantiating an apparently paradoxical negative relationship between uncertainty and learning rates, that induces strikingly unexpected resistance to re-evaluation. Such seemingly contradictory findings can be reconciled if the notion of "estimate uncertainty" is conceptually split into two components having opposite effects on learning rates: the kind emerging from a limited pool of occurring evidence would induce more thorough observational evaluations and novelty processing, the one related to the expected standard deviation of the generative distribution would elicit instead a relative underestimation of the probability of an occurred regime transition when uncertainty is high enough to justify the expectation of constitutively significant noisy variance. The broad interindividual variability documented in the choice of hazard rate underscored the differential profile of strengths and limitation in the updating approach of the prior distribution probability concerning environmental volatility: those with high hazard rates exhibited high sensitivity to change but over-responsiveness to meaningless contextual noise, whereas subjects adopting a low hazard rate were relatively successful in formulating accurate predictions when stability was situationally dominating. Useful validations of how learning rates dynamics and differential degrees of reliance in recent outcome history were contingent upon idiosyncratic priors concerning the statistical reliability of the input is supplied by Behrens et al. (2007). Examining neural signals occurring in different phases of a decision-making task, volatility-related signal change in the region of the anterior cingulate cortex could predict mean behavioural learning rates across individuals, revealing that subjects with higher effects of volatility during the monitoring phase exhibited a higher behavioural learning rate, hinging relatively more in weighting recent sensory information. Understanding statistical learning in these terms illuminates seemingly contradictory statistical learning findings such as learners often appearing remarkably sensitive to subtle regularities in sequential input while sometimes persisting in incorrect structural interpretations despite substantial contradictory evidence can therefore be reconciled. Rather than reflecting inconsistent learning abilities, these patterns can be understood as consequential of a regime-based inference where sensitivity to regularities reflects the learner's capabilities in organizing coherent latent organizational hypotheses, whereas persistence reflects the evaluative stance concerning convenience of inferential stability or flexibility.

However, when the current regime is judged to be inadequate, regime revision is not merely a matter of abandoning one internal model, but to select the alternative organization that should replace it. The system must resolve this competition between a multiplicity of plausible learning approaches, optimally controlling which predictive regime comes to guide interpretation, shape expectation, and ultimately control behavioural output. Crucially, however, inferential revision does not arbitrate within an unconstrained inferential vacuum, rather the hypothesis space is organized through hard and soft learning constraints, shaped by representational biases and inductive priors concerning the relative plausibility of candidate structural organizations.

The importance of a constrained hypothesis space in circumscribing the learnability problem has been emphasized in theoretical work on linguistic acquisition of grammar by Pearl and Goldwater (2011). The representational and computational gradient of structures' admissibility, the learner's present inferential and computational constraints concur indeed in governing the hierarchy of authority of the set of latent organizations and dictate ultimate behavioural expression. As competition is not merely a static contest among stored representations, it can be more accurately understood as a conflict among candidate regimes for behavioural governance, arbitrating a selective readout policy and dynamically shifting in the organizational structure currently privileged as the basis for prediction and action. Notwithstanding the usefulness of Bayesian ideal learners' models in outlining contextually-informed principles of environmental fit estimation, such proposals can't thoroughly exhaust the above outlined issue of behavioural dominance among competing statistical structures.

A contemporary empirical anchor for the notion of a governance bottleneck can be found in recent work by Pedraza et al. (2024), which reports a competitive relationship between executive functions and statistical learning, whose results suggest an antagonistic interaction between the implicit acquisition of probabilistic regularities and goal-directed updating processes supported by prefrontal control systems. Although statistical learning cannot be straightforwardly mapped onto the model-free versus model-based dichotomy developed in reinforcement learning research, the analogy remains instructive. In both cases, behaviour appears to emerge from a selective deployment process operating over multiple concurrently available strategies, each optimized for different trade-offs between flexibility, reliability, and computational cost (Daw, Niv, & Dayan, 2005; Wilson et al., 2014).

The considerations developed thus far converge in outlining a unified characterization of statistical learning as a multi-level process of structural inference operating under bounded computational resources, in which each stage of the predictive hierarchy can instantiate multiple configurations of the variables ultimately shaping perceptual organization. This level of unitized description is however conceptually distinct from the increasingly consolidated view of statistical learning abilities as a componential domain (Siegelman et al.; 2016). As discussed in the first part of this dissertation, modality-specific dissociations, neurobiological correlates and domain-selective performance profiles converge in establishing that statistical learning does not constitute a unitary cognitive faculty and can be more accurately understood as a family of partially independent abilities idiosyncratically declined across sensory and representational systems. The present proposal of a theoretically plausible unification, therefore, does not address the ascertained multi-componential characterization of the statistical learning domain, applying instead at a different explanatory layer. Functionally dissociable manifestations of statistical learning are posited to share a common computational principle: across modalities and tasks, learners appear to structure sequential input through hierarchical processes of predictive inference, progressively integrating evidence across multiple temporal scales while maintaining competing hypotheses about the latent organization of the stream. Further, approaching statistical learning through a systematic effort to delineate the determinants of such processes, and their reciprocal interactions, should not be mistaken for an artificial dissection of a form of learning often described as immediate or implicit. On the contrary, the present argument aims to extend the nested complexity typically attributed to semantic cognition to perceptual inference, regardless of the seemingly contradiction inherent in the relative ease with which perceptual regularities can be acquired or expressed behaviourally. Beyond the attempt to describe a range of plausible computational principles applying ubiquitously to structure humans' cognitive architecture; each stage of the inferential cascade exhibits a strikingly broad range of inter-individual variability. While such dissimilarities seem to undermine, at a first glance, the very commonalities previously posited, potentially invalidating a unified account, however, these divergences point to a more subtle conclusion. Individual idiosyncrasies do not stand in opposition to the general computational logics governing statistical learning; they are the manner in which those logics become instantiated, at each processual stage, within the learner's specific cognitive system. Accordingly, the apparent difficulty of identifying a single, stable parametrization of statistical learning does not imply theoretical incoherence or reduced robustness of the phenomenon under scrutiny but connotes the plausible set of underlying inferential principles through the lens of cognitive architectures' subjectivity: heterogeneous behavioural expressions can therefore be reframed once the subjective weighting of the relevant variables is considered.

A hierarchical distribution of interindividual variability across the multi-processual trajectory of statistical learning

The hitherto proposed theoretical dissertation and the body of empirical evidence discussed thus far hopefully compel to acknowledge the reductionism inherent the attempt to frame the complex, multi-componential architecture of temporal statistical learning as the mere registration of recurrent contingencies. The previous sections aimed instead to progressively substantiate the plausibility of conceiving this phenomenon as a hierarchical inferential process, occurring under conditions of uncertainty and epistemic non-closure, across predictive timescales specifically constraining diagnostic latency and robustness and within a system capable of concurrently sustaining more than one viable organization of the informative stream. Such a set of premises renders the postulation of a uniform learner profile logically untenable, since the range of pathways and intermediate turning points preluding the definitive outcome of stabilized and overtly expressed predictive structure is unconceivably broad. More promising than a simplistic effort to quantify the scalar magnitude of different individuals' learning strength, seems the attempt to hypothesize where, within this heterogeneous cognitive trajectory, principled interindividual divergences should be expected and in which form. Interindividual variability is not to be treated as a nuisance (Bogaerts et al., 2018; Kidd et al., 2018), an accessory complication of an otherwise homogeneous process, simply requiring to be minimized; rather it can be conceived as an exploratory tool to probe the observable variety of manifestations of predictive inference. At each stage of this multi-level inferential hierarchy, two subjects exposed to the same learning context under analogous conditions may diverge in what they treat as a meaningfully structuring pattern, which temporal horizon is privileged as the more trustworthy carrier of prediction, how urgently timing requirements dictate commitment to an emerging organization and the cutoffs of such latent structure's robustness in tolerating discrepancy.

Features of sensitivity to and extraction of statistical structure deserve to be considered as the most elementary, yet explanatorily insufficient locus of interindividual divergence. A first ground where individuals' dissimilarities are expressed concerns their spontaneous proclivity and readiness to resort to patterning logics in order to meaningfully fragment a continuous stream, with some learners that may indeed be more efficient than others in organizing a noisy input through latent relevant subunits. Further, the extension of the predictive horizon instantiated by dependencies spanning different timescales isn't epistemologically neutral but instantiates distinct commitment geometries that ensue from the relative attribution of reliability to features characterizing short- and long-range statistical dependencies. Accordingly, learners may vary in the distribution of functional priority allocated to specific temporal spans: for some, locally dense and rapidly auditable relations are the most valuable features to scaffold anticipation; for others, extended intervals interposed amidst distal relations yield the potential to highlight the stability of the structure, whose robustness emerges through their contrast with neighbourhood's variability. The chosen balance point depends both on contextual determinants and on subjective traits of personality, with interindividual variability and life course revisions contributing a substantial expansion of the range of designated strategies.

An additional fertile field for divergence concerns commitment policies: once the latent organizational principles of the currently operative predictive regime has been established, learning continues to proceed in an irreversibly conditioned fashion, with incoming evidence conveniently bent as they are processed relative to the ongoing expectations. Here, commitment policies designate the criteria by which a learner grants provisional authority to one interpretative regime and regulates its persistence in the face of partial violations. This would render statistical learning a merely-self-confirmatory cognitive process, unless an upper bound for tolerable ambiguity was set, determining a continuously operating evaluative mechanism that capitalizes both on individual attitudes and contextual evaluations. Subjectivity may lead to more conservatively set thresholds of tolerance to ambiguity, where a minimal degree of uncertainty is sufficient to cast doubt on the correctness of the current organizational principle or where people with an impatient attitude accept to only fleetingly engage in evaluative mechanism, thus being extremely responsive to perceived, yet possibly meaningless, changes in the informative stream. On the other hand, a different learning agent may opt to privilege a reassuring solidity of the epistemological framework, embracing the risk of lagged receptivity to environmental requests. Beyond individually determined requirements to commit to or to abandon their representational objects, subjects may differ in assessment of contextual features of stability/volatility, set of factors whose dynamic interaction results in inter-individually uneven policies of settlement between transient irregularity or decisive evidence of structural change.

A further level of variability concerns behavioural governance, namely the process by which one among several available predictive organizations is granted control over overt choice. Within an environment comprising informative streams capable of scaffolding coexistent multiple predictive organizations, the learner may acquire more than one internal model, but the behavioural response cannot express all of them simultaneously, especially when their nature is mutually incompatible. It appears plausible that a plurality of internal models is progressively achieved, with some remaining behaviourally quiescent, whereas others may dominate because they are easier to retrieve, more stable, less computationally demanding, or more directly suited to the response format. From this perspective, differences in overt performance do not necessarily imply differences in what has been learned; reflecting instead differences in which representation is allowed to govern behaviour at the moment of decision. Again, at the interindividual level, such inherent coexistence among predictive inferential frameworks may be variously handled, with subjects attributing temporary authority, thus preferential behavioural deployment, to one interpretative frame according to the idiosyncratic resolution of the trade-off between accuracy and computational resources. From this perspective, variability across learners may appear in the form of overt differences in behavioural performance's indexes even when their underlying representational repertoire doesn't meaningfully diverge.

Once variability is recognized as hierarchically distributed across the architecture of statistical learning, the interpretive status of population-level summaries and their explanatory power inevitably changes. While aggregate means remain descriptively convenient indicators that a regularity can, in principle, be acquired by a group of learners, they fall irremediably short in capturing the inferential trajectory leading to the observed outcome, potentially hindering the full acknowledgment of the constitutive heterogeneity of the pathways through which that outcome is reached. Akin to how acquisition failures inferred from not above-chance accuracy ins statistical learning performance assessment shouldn't be improperly extended to postulate absolute lack of developed sensitivity to the statistic under scrutiny, identical behavioural averages can't be straightforwardly ascribed to comparable underlying organizational solutions to the same learning problem. From this perspective, the structure of variability becomes a unit of analysis bearing theoretically informative potential: differences across individuals can reveal the direction of progressive accumulation of predictive commitments at each level of the inferential hierarchy, elucidating how the idiosyncratic pattern of cumulative evidential stratification and gradual curbing of the hypothesis space makes a perceptual outcome inevitable. Under this perspective, complementing with a systematic analysis of variability more classical evaluations concerning whether performance exceeds chance affords a desirable enlargement of the explanatory landscape that becomes empirically tractable.

Such an approach becomes particularly revealing in contexts where the informational stream supports more than one equally plausible latent organizational principle, whose respective temporal trajectory extend over distinct predictive horizons; under these conditions, not only countervailing evidence is detectable within the same stream, but its undermining potential against the currently operating inferential framework doesn't reside in constituting a random sensory discrepancy. Rather, a structurally plausible alternative predictive candidate is simultaneously instantiated, whose temporal span dictates differential logics of evidential accumulation, revision density and a relative profile of susceptibility to error and resistance to interference. At this juncture, variance in both individual performance and interindividual tendencies become informative. The accumulation of counter-examples places pressure on the stability of predictive commitments, revealing how learners negotiate the balance between willingness to revise prior interpretations and the robustness of established expectations. Further, a structurally analogous, yet epistemically ambiguous environment supplies a promising neutral ground to gauge, from the resulting range of behavioural outcomes, an inter-individually variegated spectrum of predictive stances.

METHODS

The present paradigm was motivated by the already discussed limitation common to several strands of the statistical learning literature: each feature of this complex and heterogeneous empirical domain is often investigated secluded from the others, as if dissecting the complexity of the phenomenon would by itself provide explanatory clarity. As illustrated throughout the introductory part of this dissertation, a substantial body of empirical work has established that learning agents can successfully acquire both adjacent and non-adjacent dependencies from an unfolding input stream, particularly when the temporally extended statistical structure is conveyed through auditory or speech-like materials, affording the possibility to capitalize on modality-specific resolution acuity and efficiency of processing (Gómez, 2002; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996). Parallelly, the tradition of visual statistical learning has reliably substantiated that observers are capable of both extracting higher-order temporal regularities and tracking multiple statistics from continuous streams of abstract stimuli without requiring perceptually explicit segmentation cues (Fiser & Aslin, 2001; Hunt & Aslin, 2001).

Notably, despite the presence of empirical evidence relative to the acquisition of single classes of statistical dependencies or (although conspicuously less often) the parallel tracking of multiple regularities, statistical contingencies are not typically placed in a position of direct behavioural competition. The conceptual landscape, therefore, leaves underspecified a seemingly fruitful issue: does the cognitive system maintain as concurrently available distinct predictive structures? Are statistical contingencies that instantiate incompatible predictions relatively selected and weighted by the learner? Amidst a shortage of empirical evidence concerning this topic, the "statistical garden path" paradigm (Gebhart, Aslin, & Newport, 2009) proves to be uniquely instructive about a potential dissociation between learning and behavioural deployment, whereby learners' commitment to a specific statistical organization may hinder processes of prediction revision as alternative statistical structures become more relevant in a subsequent experimental phase. Nevertheless, the scarcity of existing paradigms does not provide a controlled setting in which multiple statistical contingencies of distinct predictive range both converge on a shared outcome space, scaffolding the instantiation of direct conflict among predictions. Accordingly, the rationale of the designed paradigm stems from an interest to jointly address several research questions: are temporal learning mechanisms able to preserve sensitivity to multiple contingencies, whose nature is instantiated probabilistically and defined over distinct temporal spans? If sensitivity to concurrently available, yet discordant, statistical structures has been established, how is behavioural expression administered?

The present work proposes a novel visual statistical learning paradigm designed to instantiate, within the same continuous stream, two probabilistic dependencies differing in temporal scope but, crucially, matched in formal predictive strength. More specifically, the task addressed whether participants could develop sensitivity to both an adjacent and a non-adjacent predictive relation, and whether one of these relations would gain greater behavioural authority under conditions of conflicting predictions in which the two contingencies supported simultaneously incompatible predictions. The design attempts an operationalization of the broader assumption that statistical learning should not be treated as a homogeneous capacity, whose quantification may be only captured through above-baseline accuracy; rather, performance may depend on partially dissociable components, including the concurrent availability of learned structure, the relative strength of competing regularities, the deployment of those regularities under conflict, and response-level signatures such as facilitation or variability.

The task used abstract black visual shapes presented on a white background, employing a stimulus vocabulary that was derived from the visual statistical learning tradition, in which unfamiliar, non-semantic shapes are used to minimize the influence of pre-existing lexical, conceptual, or object-based associations. The full stimulus set, including both the familiarization and testing phases, consisted of 32 abstract figures, thus entailing additional items relative to the original, narrower, set (Fiser & Aslin, 2001). The additional items were obtained through manipulations of orientation and minimal transformations of a subset of the original shape, editing and normalizing in both size and resolution the final visual vocabulary in order to obtain a perceptually homogeneous set. The assignment of stimuli to functional roles was constrained: items serving as predictors and legal outcomes were preferentially selected from the non-transformed set, whereas orientation-modified variants were primarily used to populate filler categories.

The core set of shapes belonging to the familiarization stream was organized around four predictive items, two legal endings, and a set of filler items, with two of the predictors instantiating a longer range (i.e. global) dependency and the remaining ones instructing a local contingency. The two legal endings constituted a common outcome space common to instances belonging to each class of dependency: a precaution allowing the global and local regularities to converge on the same two possible final outcomes, therefore allowing the two predictive structures to become directly comparable and mutually competitive.

Each familiarization unit was a triplet composed of three sequentially presented shapes and could be classified into a *global* kind, characterized by the instantiation of predictive relation between the item in Position 1 and its relative legal ending in Position 3, and *local* triplets, in which the predictive relation connected the second and third position's items. The remaining positional slots in both classes of triplets were instead occupied by a not consistently varying filler item, which were not merely incidental placeholders. Their function was to preserve the planned 3-item format while preventing from the risk entailed by experiencing a repeatedly occurring pairing of the filler with a specific ending, possibly leading participants to rely on unintended filler–ending associations. The familiarization lists were therefore generated under the precaution of keeping fillers statistically neutral: for each list, filler usage was tracked across the eight blocks composing the exposure phase, separately for co-occurrences with specific category predictors or with particular legal ending, allowing to arrange a sampling assignment of candidate fillers sufficiently balanced across positions and outcome contexts, thereby preventing a filler to acquire systematic predictive value.

Crucially, global and local dependencies were equated in objective predictive validity as both dependency classes were governed by the same 80/20 probabilistic structure, according to which the within-class predictive relationship between predictors and legal endings were complementary (i.e. in both global and local classes, one predictive element was followed by one legal ending on 80% of its occurrences and by the alternative ending on the remaining 20%, whereas the spare predictive item belonging to the same category instantiated the opposite mapping).

The familiarization phase (Fig.1) followed a block wise generative structure: it contained 8 blocks, each one respectively entailing 20 triplets, yielding, across the whole exposure phase, a total of 160 familiarization triplets. Overall, a block consisted of 10 global triplets and 10 local ones, and each class was arranged in such a way to implement 5 occurrences of each of the two global predictive elements and 5 occurrences of each of the two local predictors. Notably, the desired probabilistic ratio was achieved by imposing, for each 5-trial predictor set, that 4 occurrences followed the dominant mapping, and one trial instantiated the low-probability alternative, thereby implementing the 80/20 contingency consistently within each block. The triplet boundaries were not perceptually marked; rendering the input sequence a perceptually continuous stream of individual shapes rather than a sequence of explicitly segmented units, as it was paramount to preserve the inferential structure of the task, allowing to test whether the regularities could be extracted from temporal co-occurrence patterns rather than from overt grouping cues. The order of trials was randomized while preserving constraints on the distribution of dependency type and rule-consistent versus rule-violating trials, while additionally warranting that each block encompassed the full set of predictors–outcome contingencies. Each shape appeared for 800 ms and was followed by a 200 ms blank interval and during the whole familiarization phase participants were not required to respond and were not informed about the presence of the underlying regularities. The above illustrated empirical architecture was adopted since it made the stream probabilistic, structured but not explicitly segmented, and informative only through the accumulation of exposure across the sequence.

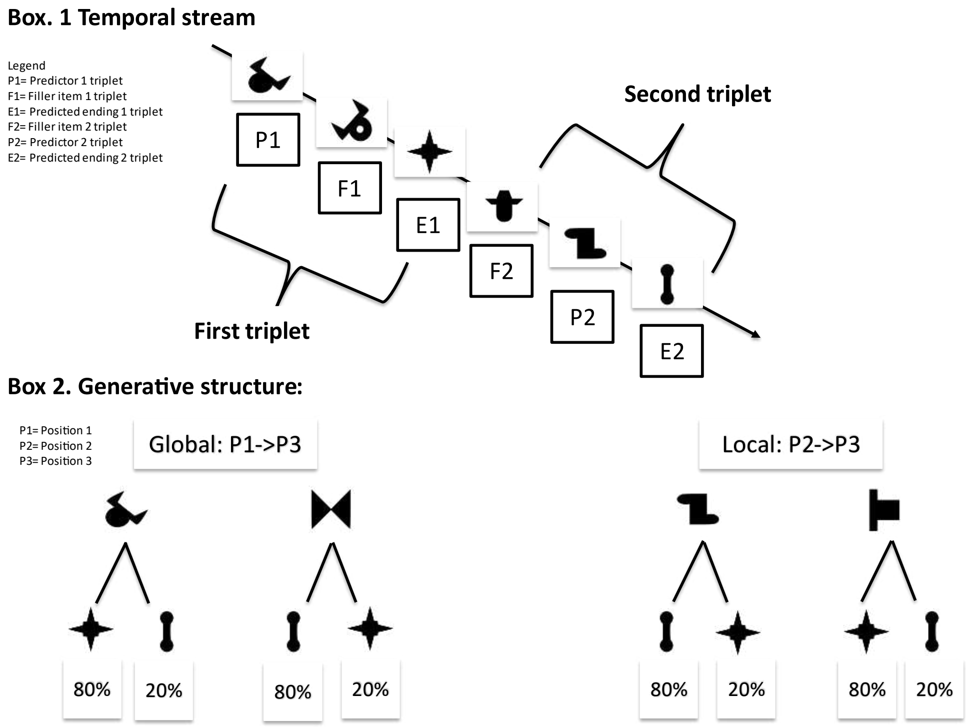

Figure 1. Temporal embeddings and concurrent predictive structure. Predictive contingencies are embedded within a continuous temporal stream. P1 is the structurally predictive element in the 1st position of "global triplets", diagnostic with 80 % probability of one of the 2 possible ending shapes in position 3 (E1, E2). P2 occupies the 2nd position in "local triplets", maintaining the probabilistic dependency of 80/20. Predictors of both local and global triplets belong each to 2 distinct shape categories, instantiating complementary predictions relative to 2 possible outcomes. To complete the triplet a filler appears in the remaining position: those items don't instantiate predictivity since each one appears equally in the 1st and middle position (i.e. in local and global contexts)

The testing phase was designed to translate the described generative structure into separable and meaningful behavioural probes. Participants completed a two-alternative forced-choice (2-AFC) task in which each trial displayed a two-item prompt and two possible candidate shapes providing the alternatives among which the item adequate for the triplet completion must be selected. The instructions provided to the experimental subjects did not frame it as an explicit rule test, as subjects were simply asked to choose the option that completed the sequence according to their intuition, further imposing a time limit of 4 seconds for deciding that was intended to discourage prolonged explicit deliberation.

The full test phase (Fig.2) encompassed 56 trials: 12 global trials, 12 local trials, 24 structurally matched baseline trials, and 8 contrast trials: all those structural types were intermixed and randomized in their presentation, keeping response alternatives counterbalanced across screen side, and not providing either any information concerning the underlying rational of statistical dependencies, nor subjective performance feedback.

A first subset of trials assessed sensitivity to global dependencies: here the prompt preserved the Position-1 predictor, allowing to assess whether the subject would choose the ending that was most consistent with the non-adjacent relation experienced during familiarization. A symmetric set of 12 local test trials assessed instead sensitivity to the adjacent dependency, presenting the Position-2 diagnostic element, and defining the correct response as the high-probability ending associated with that predictor during exposure. For each learning branch, half of the trials contrasted the 2 elements belonging to the pool of legal endings, whereas the other half contrasted it with a filler item that had not been experienced as a legal final outcome in the previous phase.

A central concern in the construction of the testing phase was that the experimental trials themselves could introduce unintended statistical or familiarity-based cues, rendering raw accuracy measures in the experimental trials an inadequate metric to estimate statistical learning in an uncontaminated way. Indeed, in the global and local trials, the nominally correct alternative was not simply one response option among two equivalent stimuli, as it belonged to the restricted class of legal endings that had occupied the terminal position during familiarization. Especially in conditions where the predicted ending was contrasted with a filler, participants could obtain accuracy by relying on terminal-position familiarity, which directly affects the degree of plausibility of legal endings, rather than on the designed specific predictor–ending. However, even in trials where the distractor option was the alternative legal ending, a potential cue could unintendedly ensue within the testing phase itself, owing to the repeated prompt–target pairings during test. To address the possibility that participants might have chosen the nominally correct item not as a function of actual learning of the predictor–ending contingency, but because the target appeared more familiar, more plausible, or more consistent with the test format, a parallel baseline set of 24 trials was included (Fig.2). These baseline trials structurally reproduced the same two-alternative forced-choice structure, the same distinction between legal-ending and filler alternatives, and the same arbitrary assignment of a correct option, but with items that had not been experienced during familiarization.

The final component of the test phase consisted of 8 contrast trials, whose purpose was not to measure correctness, but to index behavioural deployment under conflict. In these trials, the prompt entailed both a global and a local predictor, which, crucially, reliably cued opposite endings with equated probabilistic strength. This empirical class of trials aimed to cognitively activate the two types of dependencies differing for their predictive span, instantiating a conflictual decisional arbitration owing to their inferential co-availability. Accordingly, the response alternatives consisted exclusively of one legal ending consistent with the global predictor and the other consistent with the local one and, since each alternative was supported by one of the trained structures, contrast trials could not be coded in terms of objective correctness. Methodologically this choice implied an explicit dissociation between learning and use, so that branch-specific assessment of baseline-adjusted accuracy could be complemented with pieces of information concerning the subjectively administered balance and the relative gain of behavioural authority of each class of dependencies (Fig.2).

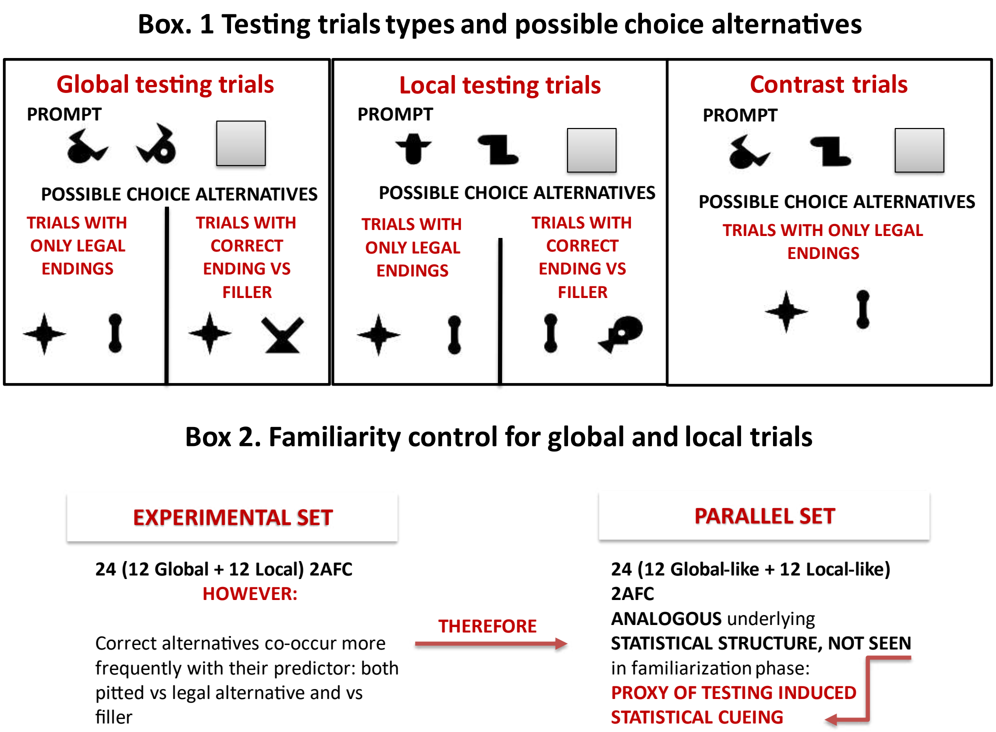

Figure 2. Testing phase dissociates predictive availability from familiarity-driven responding.

Box 1. Testing trials' structure. In global and local trials, participants perform a 2-alternative forced choice (2AFC) task among alternatives that may include the rule-consistent ending paired either with a filler or with the alternative legal ending. Contrast trials directly oppose global- and local-consistent endings, assessing arbitration among concurrently available predictive models.

Box 2. To control for familiarity induced during testing phase by the mere frequency of co-occurrence of the correct pair predictor-predicted, global and local trials are mirrored in a parallel set instantiating an analogous statistical structure but not present during the familiarization phase, to isolate genuine learning effects.

EXPERIMENT 1: In-person implementation

METHODS

Experiment 1 constituted the controlled in-person implementation of the global–local visual statistical learning paradigm above described. Its primary aim was to evaluate whether the task architecture, once protected as far as possible from item-specific and list-specific confounds, generated systematic group-level evidence of sensitivity to the embedded probabilistic dependencies. Because the proposed paradigm was novel and the whole index architecture, although theoretically motivated, was partly newly defined, this in-person version can be treated as an exploratory pilot, an empirical test where, rather than a confirmatory demonstration of a pre-established effect, the underlying rationale was to assess whether the task produced coherent behavioural signatures across accuracy, behavioural deployment, and response dynamics. The in-person implementation was specifically suited to group-level inference because it relied on constrained familiarization lists, a choice that directly stemmed from the structure of the paradigm and the nature of the visual materials. Abstract shapes are customarily adopted in visual learning paradigms (Fiser & Aslin, 2001), with the aim to neutralize potential semantic and lexical confounds; however, they do not necessarily eliminate perceptual idiosyncrasy. Indeed, individual figures may differ across individuals in perceived salience, confusability, elicited mnestic associations, or spontaneous response appeal. Accordingly, the administration of a single configuration of figures may unintendedly determine that these accidental properties and the entailed idiosyncratic perceptual biases become inseparable from the functional role assigned to a stimulus (i.e. global or local predictor, legal ending, or filler). A list-based generative logic was therefore adopted to prevent the empirical task from being reducible to one arbitrary arrangement of visual items and produced 10 familiarization and corresponding test-lists. Across lists, the probabilistic grammar of the task remained invariant, but the concrete distribution of fillers and the pairing history of individual shapes were varied under explicit constraints allowing to attribute the metrics deemed relevant to the generation of a systematic behavioural profile having controlled for list-specific variance.

Sixty-four Italian participants were included in the final behavioural dataset, ranging in age from 22 to 64 years (*M* = 32.67, *SD* = 10.13; median = 29), and the final sample included 36 women and 28 men. Because the computation of some dependent measures required reaching a minimal threshold of valid observations in specific trial subsets, the effective number of participants contributing to each index could vary slightly, particularly for reaction-time and variability-based measures.

An inferential question of primary interest concerned whether participants showed an above-chance selection of the outcome licensed by the trained predictor–ending contingency as the continuation of the prompt-instantiated predictive cue. Therefore, with a logic consistent with the long-standing use of forced-choice performance as an overt behavioural index of exposure-dependent learning in statistical learning research (Saffran et al., 1996; Fiser & Aslin, 2001), accuracy has been adopted in the current paradigm as the most direct overt measure of whether exposure successfully biased participants toward the trained structure. The present work acknowledges (and conceptually shares) the most recent concerns that accuracy adequacy may be task-dependent and insufficient as a stand-alone measure of a latent learning ability (Siegelman, Bogaerts, Christiansen, & Frost, 2017); nonetheless accuracy was retained as the core dependent variable for several reasons. First, the task was explicitly exploratory because it deliberately probed the phenomenon of statistical learning under particularly demanding boundary conditions. Instead of isolating a single regularity, the paradigm instantiated a dense and heterogeneous landscape of probabilistic dependencies spanning different temporal granularities, and did so in the visual modality, where temporal structure is arguably less naturally exploited than in auditory or speech-like domains. This choice was a calculated methodological risk as it allowed us to ask whether uninstructed sensitivity could emerge in the presence of concurrent and partially competing regularities, and whether different modes of evidence accumulation, uncertainty tolerance, and behavioural deployment would become observable within the same learning environment. As a corollary, accuracy was deemed as sufficiently theoretically appropriate because global and local experimental trials had a defined high-probability (80%), converging in granting accuracy-based measures the status of core analytic spine. At the same time, accuracy was not treated as an exhaustive measure of learning, as the broader theoretical framework of the dissertation assumes that statistical learning may be expressed through multiple, partially dissociable behavioural signatures, including concurrent availability, overt deployment, response time facilitation, stabilization, and sensitivity to competition between legal structures. Notably, such measures may enrich the characterization of the phenomenon of interest. potentially unveiling specific features of statistical structure yielding a somewhat privileged processing status. For instance, faster or less variable responding may reflect acquisition-related modulations of decisional fluency, uncertainty reduction, or task-specific heuristics, making them valuable measures to be retained, not as primary evidence of acquisition, but as exploratory process-sensitive descriptors.

The primary learning indices were proportion-based and baseline-adjusted. Concerning branch-specific learning metrics both a *Global (GLI)* and *Local Learning Index (LLI)* were obtained through an analogous underlying logic, which compared accuracy in both classes of experimental trials with accuracy in the respective matched parallel baseline. Both indexes quantified whether and the extent to which performance in the trained contingencies exceeded execution under structurally equivalent but untrained conditions, as formalized in the following equations whose resulting positive values would indicate a measurable gain relative to the baseline, whilst values around zero would denote no measurable advantage relative to the not instructed set:

$$GLI_i = P\ (\ Correct \mid Global_{Exp,i}) - P\ (\ Correct \mid Global_{Par,i})$$

$$LLI_i = P\ (\ Correct \mid Local_{Exp,i}) - P\ (Correct \mid Local_{Par,i})$$

To further characterize the individual learning balance between non-adjacent and adjacent contingencies, a *Preferential Learning Index (PLI)* was defined as the difference between the above-described metrics, capturing the relative direction of accuracy-based, baseline-adjusted performance, with positive values indicating a more global-oriented profile and negative values indicating a more local-oriented profile. Crucially, as this index is inherently comparative, it was possible to interpret it only in relation to the two branch-specific learning indices, according to which a positive balance could for instance reflect stronger global learning, weaker local learning, below-baseline local performance, or a combination of these possibilities.

$$PLI_i = GLI_i - LLI_i$$

Experimental conditions of direct conflict, whereby the first and the second predictive element of the prompt instantiate concurrently incompatible completions, provided the optimal ground to quantify the tendency to preferentially deploy one or the other rule-consistent behavioural responses. Such a measure, therefore, shifts from reflecting retrieval strength to directly index arbitration policies among concurrently available predictive models (Daw et al., 2005 Gebhart et al., 2009). The underlying rationale assumed that rather than stochastic fluctuations in behaviour, subjects will exhibit systematic, internally consistent biases, formally assessed through the *Predictive Deployment Index (PDI)*, which specifically established which dependency gained behavioural authority when the two, equally reliable predictive structures were placed in direct opposition. Positive values here indicated preferential selection of the globally consistent ending, whereas negative values signalled preferential deployment of the locally consistent structure.

$$PDI_i = P\ (\,Global\ Choice \mid Contrast_i\,) - P\ (\,Local\ Choice \mid Contrast_i)$$

Finally, a *Corrected global–local orientation index* combined the subjective accuracy-balance component of learning and the conflict-deployment component after having rescaled them onto comparable ranges. This composite score was not used as a pure acquisition score, but as a descriptive summary of the relative accordance or mismatch between the individual performance profiles across acquisition and deployment, without conflating the two components conceptually.

Having assumed the perspective of the multi-componential nature of statistical learning, the obtained pool of data was additionally examined through a set of exploratory measures, organized into theoretically motivated families devised to determine whether the task generated additional signatures of structured processing. One of such families addressed the availability status of classes of statistical dependencies and was composed of 2 metrics: the first that, collapsing global and local trials, assessed whether experimental performance exceeded baseline, while the other queried whether both dependency branches showed simultaneous evidence of above-baseline availability by taking the weaker of the two branch-specific learning scores. This lower-bound logic was adopted because an apparently balanced global–local profile does not necessarily imply that both dependencies were learned above chance levels; as balance can also arise when both branches are weak or when one branch is selectively disadvantaged.

Concerning the family of reaction-time-based indices, the label *process-based learning indices* was used in a deliberately restricted sense. The present measures should not be treated as canonical "process-based" or "implicit" measures, terms that in statistical learning research are referred to more online or reaction-time-based paradigms (such as serial reaction time variants), often introduced to reduce explicit reflection and to capture processing consequences of predictability. This classificatory caution is particularly relevant for the present task, because reaction times were not recorded during familiarization as trial-by-trial indices of prediction formation, but as complementary measures collected during the post-exposure forced-choice test; thus, not allowing their interpretation as direct measures of acquisition. They could be more parsimoniously conceptualized as *response-dynamic indices*: secondary measures assessing whether, for the restricted subset of correctly selected responses, performance was accompanied by greater decisional fluency or by changes in response consistency. This cautionary terminology also aligns with psychometric concerns raised in the statistical learning literature, where group-level effects and individual-difference measures may diverge, and where low trial counts (as in the present case), weak chance-level performance, and measurement noise can compromise the reliability and validity of behavioural indices (Siegelman et al., 2017). Having established the required epistemological constraints, the present analysis distinguished two families of response-dynamic measures: the first indexing correct-*response facilitation*, which were operationalized as faster median reaction times for correct experimental trials relative to their matched baselines and the second addressing *correct-response stability*, defined as reduced or increased dispersion of correct reaction times, estimated through the interquartile range. This dissociation was postulated because facilitation and stability of the response need not necessarily coincide: with partially available structure possibly failing to accelerate responses, yet still compressing their variability, or, conversely, yielding faster but less stable responding. Notably, the multicomponential logic was extended in this paradigm not only across different dependent variables (accuracy, conflict choice, and reaction time) but also within the reaction-time domain itself, still maintaining the subordinate and exploratory status of these measures relative to the core accuracy-based learning indices.

A further exploratory measure was included to describe the degree of directional coherence between accuracy-based learning and behavioural deployment under conflict. The rationale of such a metric, referred to as the *learning–deployment alignment index* and computed as the product of the global–local accuracy-balance index and the conflict-deployment preference index, was to assess whether the dependency that was relatively favoured in baseline-adjusted accuracy was the same preferentially deployed under conditions of behavioural conflict. Positive values therefore would indicate a directional alignment between learning and deployment, while negative ones would signal a directional dissociation, namely cases in which the dependency relatively favoured in non-conflict accuracy was not the one governing choice in contrast trials. The inclusion of this measure was conceptually motivated by the claim that statistical sensitivity and behavioural authority are not necessarily equivalent and that a partial representational availability of a structure in non-conflict trials could emerge in subjects without becoming the dominant basis for choice. Conversely, conflict deployment may reveal a response policy grounded on idiosyncratic heuristics or decisional preference that are not merely reducible to the branch showing the strongest apparent acquisition.

The final exploratory indices addressed *structured competition costs* and *functional robustness*. Those measures were conceptualized as assessments of whether performance suffered from experimental trials where the target ending was pitted against another legal ending rather than against a filler alternative, as a proxy of the relative availability status of a statistically licensed continuation, which may be easy to select when the alternative is outside the learned outcome space, yet fragile when the alternative is itself structurally plausible.

All indices were computed at the participant level and then aggregated across participants, allowing to conduct inferential analyses consisting of one-sample tests on participant-level indices. The metrics expressing improvement or gain relative to baseline were tested in the theoretically predicted direction, while indices expressing relative orientation were tested two-sidedly because deviations in either direction were theoretically meaningful.

This analytic strategy defined for Experiment 1 within the thesis was purposefully designed to determine whether a constrained, list-based version of the conceived paradigm produced coherent group-level signatures across the behavioural channels deemed to be relevant and involved in the current task. However, the novelty of the paradigm circumscribed the scope of analyses to remain within an exploratory level and constrained the derived claims to the evaluation of whether the task yielded an interpretable group-level profile, not allowing to suggest anything meaningful in terms of stable individual differences. That question was reserved for Experiment 2, where the online implementation was used to evaluate whether the indices derived from the task were reliable enough to support cautious individual-difference interpretation.

status	family	conceptual_role
core	core_proportion	Baseline-adjusted global learning accuracy.
core	core_proportion	Baseline-adjusted local learning accuracy.
core	core_proportion	Relative balance between global and local learning.
core	core_proportion	Preference for global versus local choices under conflict.
core	core_proportion	Corrected 0-1 global-local preference combining learning preference and conflict deployment.
exploratory	availability_accuracy	Combined learning index collapsing global and local experimental trials relative to filler baseline.
exploratory	availability_accuracy	Dual Availability Index; conservative lower-bound estimate of simultaneous global and local availability.
exploratory	process_stabilization_iq	Global learning-related RT variability/stabilization index.
exploratory	process_stabilization_iq	Local learning-related RT variability/stabilization index.
exploratory	process_stabilization_iq	Relative global-local stabilization preference.
exploratory	process_facilitation_rt	Global RT facilitation index.
exploratory	process_facilitation_rt	Local RT facilitation index.
exploratory	process_facilitation_rt	Relative global-local RT facilitation preference.
exploratory	conflict_deployment	Conflict deployment preference.
exploratory	conflict_deployment	Learning-Deployment Alignment Index.
exploratory	conflict_deployment	Absolute residual from the learning-deployment regression.
exploratory	conflict_rt_cost	Overall choice-contingent conflict RT cost.
exploratory	conflict_rt_cost	Global-choice conflict RT cost.
exploratory	conflict_rt_cost	Local-choice conflict RT cost.
exploratory	conflict_rt_cost	Relative RT difference between global and local conflict choices.
exploratory	conflict_rt_cost	RT cost for choices incongruent with the participant's learning tendency.
exploratory	conflict_variability_iqr	Global-choice conflict RT variability change.
exploratory	conflict_variability_iqr	Local-choice conflict RT variability change.
exploratory	conflict_variability_iqr	Relative variability difference between global and local conflict choices.
exploratory	conflict_variability_iqr	IQR difference between incongruent and congruent conflict choices.
exploratory	structured_competition	Overall structured competition cost.
exploratory	structured_competition	Global structured competition cost.
exploratory	structured_competition	Local structured competition cost.
exploratory	structured_competition	Normalized Structured Competition Selectivity index.
exploratory	robustness	Global robustness index combining global learning gain and global competition cost.
exploratory	robustness	Local robustness index combining local learning gain and local competition cost.

Figure 3. Schematic overview of analytical metrics

RESULTS

As previously stated, the analyses of Experiment 1 aimed to establish whether a controlled in-person version of the devised paradigm could be successful in producing a coherent and interpretable group-level behavioural profile. As imposed by the speculative nature of the task, the obtained outcomes were not interpreted as independent pieces of evidence, corroborating either valid acquisition or learning failures in terms of a unitary learning score. The to be discussed metrics are conceived as complementary pieces of information about how the different components hypothesized to be engaged in the paradigm behaved once translated into quantifiable indices.

The first and most theoretically constraining result is that the core set of accuracy-based indices did not support a classical statistical-learning profile. Indeed, participants did not show above-baseline accuracy for either structural dependency (Fig.4), revealing an analogous pattern of nonsignificant difference from zero for both the Global learning index ($M = -0.031$, $SD = 0.209$, $t(61) = -1.16$, $p = .875$) and the Local learning one ($M = -0.068$, $SD = 0.251$, $t(62) =$

−2.14, $p = .982$). Therefore, both the local and global learning branches fell short in exhibiting evidence for positive learning, as indicated by the absence of any statistically significant advantage in the completion of trials whose patterned regularity was experienced during the familiarization phase over their matched baselines, lacking previous exposure. Overall, in the most direct accuracy-based sense, the pilot showed that participants did not reliably select the statistically licensed continuation in either the global or the local branch.

The relative learning balance between the two branches was also weak as shown by the Preferential Learning Index (Fig. 4) that, although suggesting a descriptive asymmetry in the global direction, was not statistically reliable in its effects, $M = 0.054$, $SD = 0.265$, $t(61) = 1.60$, $p = .114$. Thus, although the relative comparison between the two baseline-adjusted branches did not reveal a preferential accuracy profile in the current data, a directional interpretation of this index requires caution in being evaluated as an asymmetrically robust global or local learning tendency. Indeed, the balance index is computed as a comparative measure, specifically as the difference between other two difference-based indices (i.e. branch-specific learning scores), preventing any interpretation in terms of acquisition if not supported by the branch specific direct learning.

A slightly more encouraging pattern emerged when the two dependencies were placed in direct competition. Here, participants appeared to select the globally consistent ending more often than the locally consistent shape in conditions where both structures supported concurrently incompatible responses, as shown by a significantly positive Predictive Deployment Index (Fig.4), $M = 0.102$, $SD = 0.342$, $t(61) = 2.35$, $p = .022$. Moreover, the Corrected global–local orientation Index, which combines the participant-specific relative accuracy component in non-conflictual trials and the contrast-deployment constituent after rescaling them onto a common metric, was also significantly above the neutral value of .50, $M = 0.532$, $SD = 0.097$, $t(61) = 2.62$, $p = .011$. Notably, such a picture can't be taken as evidence of global learning in the standard accuracy-based sense; nevertheless, a behavioural decisional bias in favour of the global alternative was displayed at a group-level, when the task required participants to arbitrate between two structurally, equally plausible, yet simultaneously incompatible predictions. An apparent dissociation seems therefore to emerge between what has been termed in the present dissertation accuracy-based acquisition and behavioural deployment under conflict.

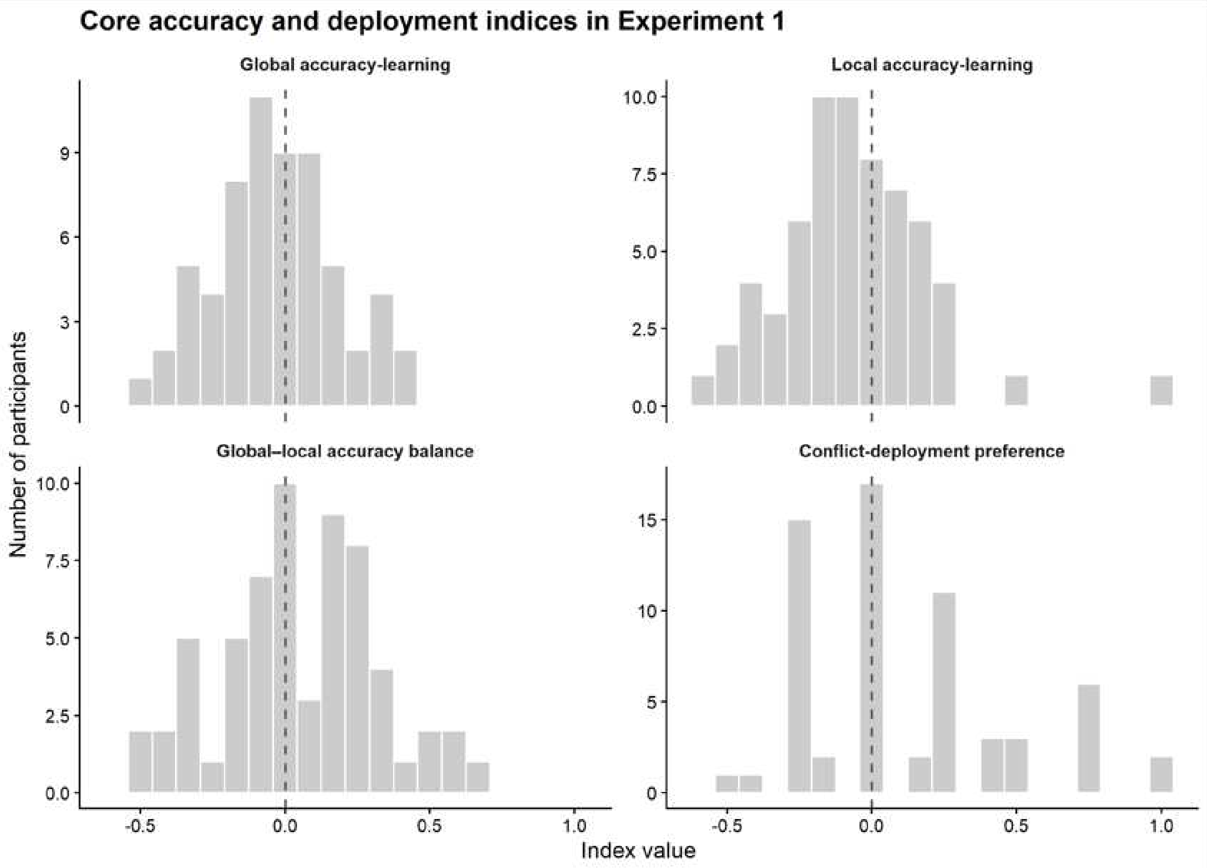

Figure 4. Distribution of core accuracy and deployment indices in Experiment 1.
The figure shows participant-level distributions for the 4 principal proportion-based indices: Global learning index, Local learning index, Preferential Learning Index, and Predictive Deployment Index. Global and local learning indices quantify baseline-adjusted performance for each dependency type, with positive values indicating higher accuracy in experimental trials than in the corresponding parallel baseline. The Preferential Learning Index expresses the relative direction of accuracy-based performance, with positive values indicating a more global-oriented profile and negative values indicating a more local-oriented profile. The Predictive Deployment Index quantifies choice behaviour in contrast trials, with positive values indicating preferential selection of the globally consistent ending and negative values indicating preferential selection of the locally consistent ending.

The availability indices reinforced the absence of a classical dual-learning profile. The *Combined learning Index* which, collapsing branch-specific accuracy scores, provided a combined measure of acquisition, was negative and converged with the absence of the predicted positive acquisition effect shown by the single structural dependencies, $M = -0.057$, $SD = 0.163$, $t\,(61) = -2.77$, $p = .996$. Consistently, the *Dual-availability index,* defined as the weaker of the two branch-specific learning scores, was negative, $M = -0.145$, $SD = 0.240$, $t\,(62) = -4.82$, $p > .999$. As this index provides a lower-bound, conservative test of whether both dependencies were simultaneously expressed above baseline, the absence of an effect was reasonably expected at group-level. Indeed, even before the absence of statistically significant acquisition was assessed in the present sample, this index was predicted to be more instructive at the level of subjective profiling and implausible to emerge as a consistent effect across participants, owing to the nature of the task, conceived to be particularly demanding. Taken together, these pieces of evidence complement the core set of accuracy-based metrics in supporting the empirical claim that absence of robust learning is not limited to one isolated branch, but also discourage the originally hypothesized possibility of the co-availability of both patterned regularities' classes.

The *Learning–deployment alignment Index* enrich the already delineating picture: this index emerged as significantly positive, $M = 0.021$, $SD = 0.095$, $t\ (61) = 1.76$, $p = .042$, suggesting a modest directional correspondence between participants' relative accuracy profile and their decisional inclination in conflict choices. Although not interpretable as an independent measure of acquisition (since the index is derived from the product of 2 already-computed measures) the intrinsic suggestion that is worthwhile to be considered is that deployment under conflict could be not wholly arbitrary when considered relatively to non-conflict accuracy. Participants who were relatively more global or local in their accuracy profile in nonconflictual trials tended, to some extent, to deploy the corresponding statistical structure when forced to choose under conflict. However, the regression analysis (Fig.5) provided an important qualification to this interpretation. Although the slope relating *Preferential Learning Index* to *Predictive Conflict Deployment* was positive, it was not statistically reliable, slope = 0.227, $R^2 = .031$, $p = .172$, therefore critically preventing from stating a strong participant-level predictive coupling between what was relatively expressed in non-conflict accuracy and what was deployed under conflict. The most defensible conclusion seems not that learning balance robustly determined deployment, but that the two dimensions showed a weak form of directional coherence without collapsing into a single behavioural axis. Specifically, in the in-person sample, conflict choices appeared to be partially organized by prior performance structure, at least in terms of directionality, yet they also retained a considerable degree of autonomy, consistent with a cautious claim in terms of partial directional coherence of the learning and usage components, which may be treated as coupled but non-equivalent elements of performance.

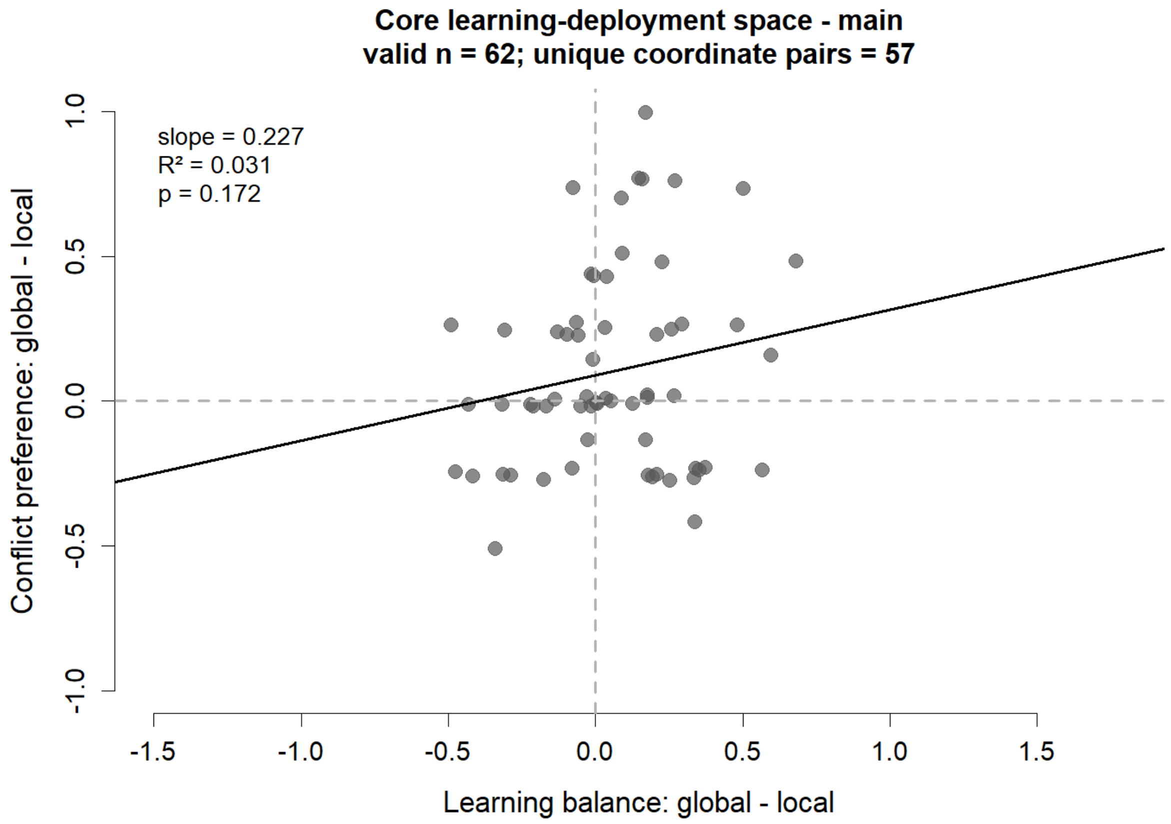

Figure 5. Core learning–deployment space in Experiment 1. The scatterplot represents the relation between the Preferential Learning Index and the Predictive Deployment Index at the participant level. Positive values on the x-axis indicate relatively more global-oriented baseline-adjusted accuracy, whereas negative values indicate relatively more local-oriented accuracy. Positive values on the y-axis indicate preferential selection of the globally consistent ending in contrast trials, whereas negative values indicate preferential selection of the locally consistent ending. Dashed lines mark the neutral value of zero on both axes, dividing the space into four quadrants of learning–deployment correspondence or dissociation. The regression line shows a positive but non-significant association between accuracy balance and conflict deployment, suggesting that learning and deployment were only weakly coupled in the in-person sample.

The conceptual architecture of the exploratory response-dynamic indices encompassed 2 macro-categories, with the purpose to complement accuracy-based learning measures by addressing whether less overt features of the behavioural output, such as the degree of facilitation or stabilization in correctly responding to experimental non-conflict trials as a proxy of the extent of developed sensitivity to their embedded statistical structure. Correct-response facilitation was not observed for either branch: the *Global correct-RT facilitation index* was negative and non-significant, $M = -148.55$ ms, $SD = 424.01$, $t\ (59) = -2.71$, $p = .996$, a pattern mirrored by the *Local correct-RT facilitation index*, $M = -31.27$ ms, $SD = 452.16$, $t\ (60) = -0.54$, $p = .704$. Likewise, the relative facilitation index did not reliably differ from zero, $M = -94.32$ ms, $SD = 604.31$, $t\ (58) = -1.20$, $p = .235$. Correct-response stability indices were similarly uninformative, and the very limited number of participants meeting the validity criteria for the IQR-based learning measures further limits their interpretability.

Conflict-related trials substantiated the conceptual rationale underlying the operated dissociation between velocity and variability-based response dynamics-indices. Here, the original prediction hypothesized a generalized conflict slowing: reaction times characterizing correct responses in conflict trials were supposed to be slower than the matched-class responses to non-conflictual trials. Notwithstanding the negative and non-significant overall *RT conflict cost index*, $M = -80.37$ ms, $SD = 361.67$, $t\ (60) = -1.74$, $p = .956$, and the lack of reliable RT costs for branch-specific conflict choices relative to their corresponding non-conflict baselines, the *Congruency-based conflict* cost was significantly positive, $M = 109.88$ ms, $SD = 416.01$, $t\ (48) = 1.85$, $p = .035$. Namely, choices that were incongruent with a participant's own learning-balance profile were slower than choices congruent with the Preferential Learning Index. This effect is, again, exploratory, but it converges conceptually with the *Learning–deployment alignment* measure: even in the absence of robust above-baseline learning, conflict behaviour may be hypothesized to still carry some traces of participants' relative performance structure, either in terms of directional alignment with the structural class preferentially deployed or as an RT cost reflected in disregarding the favoured statistical regularity.

The complementary set of variability-based conflict indices added a further, though cautious, qualification. Although the *Global-choice conflict variability index* was not significant, the *Local-choice conflict variability index* was significantly negative, $M = -129.70$ ms, $SD = 404.46$, $t\ (52) = -2.33$, $p = .023$, therefore indicating a branch-specific effect of reduced RT dispersion for local-consistent choices under conflict relative to the corresponding non-conflict baseline. Crucially, this result should not be over-interpreted, however, the curious directionality of the assessed effect, opposite to the predicted one, could suggests that conflict might have narrowed the response regime, as the outcome of a heuristic-driven responding strategy implemented under conditions which were, arguably, more demanding (i.e. concurrent incompatible predictions).

Finally, the competition family did not add an additional reliable axis to the group-level profile and cannot be used, therefore, as a strong inferential asset to characterize the present sample; remaining overall compatible with the broader conclusion that the task did not yield robust functional acquisition.

To summarize, Experiment 1 did not reveal a clear and direct group-level learning pattern in the standard accuracy-based sense: accuracy-based evidence was weak for both dependencies, and neither combined nor dual availability was robustly supported. The clearest behavioural signal emerged instead under conflict: when the task required participants to resolve conflict between the two trained structures, conditions under which participants showed a global-oriented deployment preference that was, further, modestly aligned with their relative accuracy profile. Conflict also appeared to selectively affect response dynamics, as revealed by slower reaction for choices incongruent with the subjective learning balance and reduced RT variability for local-consistent conflict choices. In a particularly demanding multi-regularity environment, behavioural organization couldn't therefore be described as straightforward acquisition, but it did reveal that global and local structures were not behaviourally equivalent once placed in conditions of arbitration under ambiguity.

DISCUSSION

The in-person pilot of the proposed novel visual statistical learning paradigm aimed to examine whether the cognitive demands imposed by a multi-regularity environment, constructed to embed within the same continuous stream two probabilistic dependencies, matched in formal predictive strength but differing in temporal scope, could reveal identifiable components of statistical behaviour. To achieve this purpose the construct of statistically-derived sensitivity was operationally fragmented: beyond the attempt to reproduce the classical demonstration that exposure to structured input yields above-baseline performance in a subsequent forced-choice task, the constructs of branch-specific learning, relative global–local balance, deployment under conflict, and response-dynamic modulation were additionally addressed. The empirical pattern that emerged from the present sample imposed an important constraint on the original ambition: the experiment did not provide evidence for robust above-chance acquisition of either the global or the local dependency, with neither branch-specific accuracy-learning index differing from baseline in the predicted positive direction.

The present discussion will treat this pattern of negative findings as the selected theoretical backbone of the interpretation. If passive exposure would have succeeded in producing a stable, representational co-availability of statistical contingencies spanning different temporal granularities, the predicted statistical-learning outcomes' profile would have been required to show higher accuracy for global and local experimental trials than for their structurally matched baselines, possibly with the *Dual-availability Index* confirming that both dependencies were concurrently expressed above chance levels. This was definitely not the observed pattern.

This outcome is particularly instructive because the learning indices were not raw, proportion-based accuracy scores, but baseline-adjusted metrics, structurally designed to control for response biases that could be elicited by the test structure. Indeed, besides the need to control for generic idiosyncratic preference or plausibility attribution to legal endings rather than filler alternatives, it was paramount to consider that the operational architecture itself may have unintendedly produced familiarity for terminal position-shape mapping as a function of the different frequency of concurrent appearance of predictor and ending figure. Therefore, the failure to obtain positive baseline-adjusted learning at group level cannot be reduced to the trivial claim that participants did not perform above chance. The most appropriate interpretation is instead that experimental trials, which embedded the statistical dependencies to which subjects were exposed during familiarization, did not elicit significantly more accurate choices than structurally analogous baseline trials that, although reproducing in their underlying structure the format of the test triplets, lacked the exposure supposed to instantiate predictor–ending contingencies. The significance of the parallel baseline is the methodological function it performs: preventing an unparsimonious interpretation of forced-choice accuracy as evidence of learning when performance could instead be plausibly explained as a function of familiarity or test-induced regularities.

At the same time, the absence of accuracy-based learning should not be mistaken for conceptual refusal of the paradigm as theoretically uninformative, given that the task was deliberately constructed as a demanding stress test of the boundaries of visual statistical learning. Unlike canonical paradigms that pursue a clean strategical effort to isolate a single regularity and subsequently assess discrimination between familiarized and non-familiarized items through more explicit or implicit performance qualifiers, the present task exposed participants to two concurrent probabilistic dependencies that converged on the same outcome space and would be later placed in direct conflict. The current design is conceived to substantially increase interpretative and computational demands, as the participant was not simply required to develop sensitivity to a repeated structure, but to extract partially independent informative streams, identifying in an undifferentiated input which, if any, predictive relation occurred consistently among specific elements, while tolerating residual uncertainty induced by the probabilistic 80/20 mappings, in order to later determine the criterion of behavioural dominance among the supposedly acquired classes of contingencies. The present methodological and conceptual choices directly stem from an increasing tendency in recent literature to shift away from conceiving the construct of statistical learning as a single homogeneous ability, expressed uniformly, at the intra- and inter-individual level, across tasks. Recent work has indeed emphasized that statistical learning performance depends strongly on the type of statistic, the sensory modality used to instantiate structured regularities, the domain-specific format of presentation, and the behavioural measure used to assess learning. Growns, et al (2020), for instance, argue that even within visual statistical learning, different tasks may index partially distinct abilities rather than a single unified mechanism. Convergingly, Siegelman et al (2017) question the framing of statistical learning as a monolithic faculty, rather suggesting that the phenomenon encompasses a set of partially dissociable individual capacities, and Arnon (2020) highlight how strongly conclusions can depend on the reliability and structure of the specific task used. These theoretical contributions prove useful in warning against immediate conclusions, that would inappropriately lead to interpret failure in the behavioural expression of learning as exhausting the possible structure of the task or the whole learning space. The absence of a robust group-level accuracy effect in this pilot may reflect the amplification of the inferential burden produced at multiple levels: from the extraction of probabilistic rather than deterministic regularities, to the use of temporally distinct rather than homogeneous structural granularities, to the taxing demands posed to the visual system by the temporal resolution required by the task and, finally, to the necessity to arbitrate behavioural dominance in conflictual conditions. A sustainable reading of the absence of straightforward accuracy gains could, accordingly, be diagnostically instructive about the fact that the paradigm may have exceeded the conditions under which uninstructed learners can reliably express coexisting visual dependencies through overt post-exposure forced-choice performance.

At this point, the difficulty of the task is not to be considered as an auxiliary limitation but part of its theoretical meaning. The intrinsic value of the paradigm is to place learners in a high-ambiguity environment, rather than exposing them to an informative stream structured in an unequivocal, precisely predetermined manner. Under the present conditions, the learner was not simply asked to segment a stream or recognize a familiarized structure, devoid of any informational noise, but, in a manner supposedly more adherent to the ecological reality, to extract predictive relations while tolerating residual uncertainty and potential competition between explanatory models, thus instantiating a conflict whose resolution cannot be definitive. This situation precisely clarifies why behaviour may be questioned as a direct readout of acquired structure, with failures in accuracy gains are not straightforwardly attributable to the fact that nothing was learned, but potentially could emerge from the task's requirements to select among partially informative cues whose relevance depends on the current response context. It should be noted that the present data do not allow us to infer hidden acquisition from failed accuracy; however, they do justify the more cautious conclusion that overt accuracy is one among multiple potential behavioural channels allowing to probe different layers of sensitivity to statistical structure. Overall, Experiment 1 highlight the limit of the current paradigm as a conventional accuracy-based learning task, producing behavioural outcomes that, however, are valuable in defining what the experiment did not show. This negative constraint should frame and curb interpretative precision of the subsequent conflict findings: in this sense a consequential and empirically faithful question is why, in the absence of robust accuracy-based acquisition, participants nevertheless showed a systematic global-oriented tendency when global and local predictions were directly opposed. A potentially interesting hypothesis concerns the sensitivity of the present task, which may less reliably capture stable acquisition as measured by isolated forced-choice accuracy, while being more sensitive to the individuals' arbitration policies through which the dominance balance between partially structured alternatives forces one predictive model to gain behavioural authority over the other.

Within this framework, a behavioural dissociation between accuracy-based expression and conflict-based deployment is documented. The significant preference for the globally consistent ending in explicitly conflictual conditions cannot license overclaims, once considered under the epistemological constraints imposed by the *Global learning Index*, which didn't show a reliable above-baseline advantage. Accordingly, the relevant empirical distinction should not be posed as "global" versus "local learning," but between accuracy-based expression in non-conflict trials and behavioural arbitration under conflict. Likewise, one meaningful insight from the pilot's should be that the same exposure history may not produce robust accuracy gains yet succeed in structuring behaviour when the response context forced arbitration between incompatible predictions.

The global-oriented deployment effect is especially relevant because the contrast trials were not merely a more difficult version of the structure-specific experimental ones: indeed, they instantiated a qualitatively different behavioural problem. In global and local test trials, the participant could in principle rely on a single predictive cue, given it was recognised as such after familiarization, and select the ending most consistent with the trained contingency; in contrast trials, instead, both cues were present and both response alternatives were equally sustainable at face value. Correctness could not be relied upon to drive individual decisional policies, since there was no objectively correct response; the trial instead required the participant to determine the extent to which the two candidate predictive models could be deemed reliable. Accordingly, the contrast trials operationalized a form of behavioural governance aiming to address not only whether a regularity had left a trace, but to assess whether, across subjects, a representational structure may become robustly dominant when the task converted co-availability into mutual incompatibility. The finding therefore aligns with broader evidence that statistical learning can be framed as something more than passive accumulation of transitional probabilities and that the logics of commitment, updating, and deployment policies may be imported from the predictive-processing domain to be adequately extended to the statistical learning phenomenon. This interpretative stance resonates with pieces of evidence highlighted by the statistical garden-path literature, although the implementation of the present paradigm cannot be directly traced back to this research tradition. Gebhart et al (2009) showed that learners' early commitments to one kind of statistical organization can constrain subsequent sensitivity to a changed structure, suggesting that the construct of statistical learning is inherently tied to mechanisms of commitment, persistence, and revision of an interpretative regime. Although the experimental manipulation of the present experiment did not employ midstream changes, but direct competition between coexisting predictive models; the conceptual implication is similar: behavioural output depends not only on the regularities present in the input, but on which structure is selected as behaviourally authoritative in a given context. Under this perspective, the global-oriented conflict preference can be described as something akin to a deployment bias or arbitration tendency.

Indeed, the pilot suggests that these two dimensions of acquisition and deployment are related enough to be meaningfully compared, but not sufficiently overlapping to be treated as equivalent. Empirically, this claim is supported, at least in a directional sense, by the positive alignment index, that suggested a modest coherence between the participants' relative orientation in accuracy and their preference in conflict choices: participants who were relatively more global or local in non-conflict trials tended, to a limited extent, to deploy the corresponding structural class under conflict. However, and crucially, the regression relating *Preferential Learning Index* to *Predictive Conflict Deployment* was descriptively positive but not statistically reliable and explained only a small amount of variance. The combination of these two pieces of non-converging evidence is theoretically instructive: on one side it prevents the interpretation of conflict deployment as merely arbitrary, because some directional coherence was present; on the other, it prevents the stronger claim that conflict choices were directly determined and could be directly attributable to the individual balance of accuracy-based learning. This partial coupling preserves the central conceptual premise of the paradigm while preventing from overstatements: if learning and deployment had been completely unrelated, the contrast trials would have been difficult to interpret as anything different than idiosyncratic response preference or task-induced bias, on the other hand, if the two dimensions had been largely overlapping, the conflict index would not have brought any informative addition to the learning indices, being merely redundant. The intermediate observed pattern suggests that indeed an additional layer of task performance was captured by contrast trials: not acquisition per se, but the transformation of weakly encoded structure into a more categorical choice when the task forced one dependency to dominate the other.

A cautious neuroscientific framing can help to elucidate the meaning of this dissociation without inappropriately overextending neural evidence to behavioural data. Contemporary accounts of statistical learning increasingly suggest a view of statistical learning as a temporally and representationally layered phenomenon, thus offering a plausible argument in favour of conceiving behavioural expression as the selective weighting or prioritization of statistical structures, whose cognitive availability cannot be directly inferred from responses. Specifically, neural and behavioural evidence converge in suggesting that learners are able to track regularities at multiple levels of organization. For instance, Ding et al. (2016) work shows that cortical activity can concurrently track hierarchical linguistic structures unfolding at different temporal granularities, and that the emergence of a higher-order structure does not necessarily imply the disappearance of lower-level sensitivity. This account does not necessarily imply that the present visual task, employing, incidentally, abstract non-linguistic material, would be expected to elicit the same neural mechanisms, nor that the global conflict preference could be taken as indexing hierarchical neural tracking. It does, however, provide a principled framework to both expect that coexisting temporal regularities may become behaviourally dissociable depending on the demands of the test context and justify not automatically equating weak accuracy with absence of all structure.

Notably, the global orientation observed under conflict is particularly interesting because the global dependency was the one spanning a more extended temporal scale and that required maintaining a predictive relation across an intervening item. In principle, empirical research in this field would make non-adjacent dependencies predictably more fragile or harder to acquire than adjacent ones, as non-adjacent learning is known to depend on representational format, intervening variability, and the learner's ability to abstract across predictively non-diagnostic material (Gómez, 2002; Newport & Aslin, 2004). The present results apparently contradict this literature-corroborated expectation, an evidence that can be exploited to suggest that when the task made both predictions explicit and mutually incompatible, the discrepant set of activated expectations may provide an informative, self-explanatory cue about the task' latent purpose to probe sensitivity to statistical structure. At first counterintuitively, a tentative epistemological perspective may be that conflict trials, originally conceived as more difficult, may retain some cueing potential about what they aim to assess, therefore benefiting relatively more the statistical structure that is supposedly less cognitively available. In this sense, the conflict effect may raise the possibility that the global predictive cue acquired a different functional status, manifested in a privileged decisional authority when the task explicitly required arbitration between two incompatible predictive models, therefore operating, in ambiguous conditions, as a broader organizing hypothesis over the triplet, consistently more often than the local cue. The fact that local structure, conversely, appears as behaviourally "suppressed" is particularly noteworthy as adjacent dependencies are often proved to be behaviourally efficient in simpler statistical learning contexts; possibly alluding to the fact that their advantage may depend on whether local information remains *uniquely* diagnostic.

Overall, the most sustainable and parsimonious, yet theoretically productive position yielded by the in-person pilot could be that, in complex probabilistic environments, the behavioural signature of statistical learning may sometimes emerge less through uniform improvement in accuracy than through the way learners resolve competition between partially informative alternatives.

The exploratory response-dynamic indices provide an important boundary around the interpretation of the deployment effect. It is crucial to acknowledge that exploratory RT measures cannot be used as compensatory evidence for acquisition, due to their interpretability restraints. Indeed, consistent with recent methodological concerns in statistical learning research, methodological purity of reaction-time or process-sensitive measures as indices of learning cannot be automatically assumed and the liability of their reliability to the influence of trial number and preprocessing choices requires some consideration. Moreover, reaction-time measures collected during a post-exposure forced-choice test are not what can be defined as a direct online measure of implicit acquisition; thus, they cannot be treated as privileged access to the learning process itself. In the present design, RT indices merely reflect selected attributes of how participants executed decisions after familiarization, conditional on the structure of the test and on having selected the statistically licensed response. They can therefore be better understood as response-dynamic descriptors rather than independent assessments of learning and, especially in a framework where core accuracy indices were weak, the role of such exploratory measures is restricted to a refinement of the behavioural profile, not licensing to transform a non-acquisitive accuracy pattern into a successful learning result.

Bearing in mind all the above constraints, a speculative stance might prompt the suggestion that, to a certain extent, an individual's developed sensitivity to a statistical structure might be queried through both a static metric of choice accuracy and more dynamic performance's attributes, thus making the conflict-related response dynamics informative. The original hypothesis of a generalized conflict RT cost (i.e. slowing responses elicited by a more taxing experimental condition) was disconfirmed by the data: participants were not simply slower whenever both predictive structures were made simultaneously relevant. Instead, choices incongruent with the participant's relative accuracy profile were revealed to be slower, and local-consistent choices under conflict showed reduced variability compared to the response range characterizing their matched baseline. This effect might conceptually complement the learning–deployment alignment in cautiously suggesting that conflict choices were not random fluctuations but were embedded in a decisional regime revealing some degree of congruency with the direction of structural acquisition emerged under non-conflictual conditions. Under this perspective, incongruent choices may have required additional deliberation because they contradicted the participant's own weakly expressed accuracy orientation, whereas reduced variability for local conflict choices may indicate a stereotyped, heuristic-like response modality when participants selected the local option under explicit competition. An additionally interesting inferential direction may even propose that such a compressed response range specific for the local branch could be conceptually convergent with the relative group-level tendency to prefer globally-aligned alternatives in conflict trials, possibly hinting at a selective response re-organization contingent on the task's ambiguity.

The sparse and faint pattern of group-level effects that delineates from the in-person implementation of the task should not be treated simplistically as conceptual or methodological weakness but may instead strengthen the interpretation and provide a corollary to the empirical exploration of the boundaries inherent in the phenomenon of statistical learning. The present empirical conditions were ambitious in aiming to maintain deterministically inconclusive but inferentially informative predictive streams, additionally demanding in the extent of cognitive discrepancies between the computational operations, kind of sensitivity and format of uncertainty tolerance brought about by each temporal span implemented by the two classes of regularities. Under these premises, a clean dual-learning effect would have been theoretically elegant but arguably very unlikely to emerge consistently across individuals. Here, the observed absence of above-baseline accuracy is particularly informative as it suggests that the paradigm may have lacked sufficient input clarity, sensorial salience or, more generally, may have exceeded the cognitive proficiency of uninstructed learners, preventing them from robustly express both dependencies through overt forced-choice performance.

At the same time, a conflict-deployment effect is shown to be elicited by the task which can suggest that the current paradigm may prove valuable in probing arbitration under uncertainty rather than pure acquisition test. As contrast trials did not require participants to identify the objectively correct option, their relative performance is the ideal empirical ground to reflect multiple channels individually deemed relevant at a decisional level: not only what has been learned, but what is selected, trusted, and operationalized when multiple predictive commitments compete. Under this perspective, the probabilistic nature of learning proves crucial in configuring the acquisition process as one unfolding under residual uncertainty, allowing to connote the later deployed behavioural output as something more layered than the transparent expression of a single internal representation. An additional theoretically relevant, but still exploratory pilot's contribution lies in demonstrating that, under controlled conditions, the current statistical learning paradigm might produce a group-level global deployment tendency despite weak accuracy-based learning, allowing to establish a behavioural dissociation between the two components of performance.

At this point it is crucial to acknowledge some limitations of this first implemented version of the paradigm in order to constrain the discussed interpretative directions. First, the accuracy indices were dependent on a relatively small number of trials per condition, particularly for the conflict class (i.e. 8 experimental trials), due to the timing logistics imposed by the online version of the paradigm, that was proposed within an extended battery aimed at probing interindividual variability across statistical learning's tasks. To keep the design of the two implementations informatively specular, also the in-person version entailed a reduced number of testing trials, clearly limiting their sensitivity and increasing measurement noise. Second, the response-dynamic indices, especially those based on IQR and conflict subsets, were additionally constrained by the imposed validity thresholds, further limiting the pool of responses that could be interpreted in the analysis and accordingly circumscribing their epistemological status as secondary descriptors rather than stable response-dynamics measures.

Third, the use of abstract visual stimuli reduced semantic confounds but could not eliminate idiosyncrasies of perceptual or mnestic nature, even though the list generation system was designed to distribute such variance across participants. Finally, and directly following from the list-generating logic, the in-person implementation was optimized for group-level control rather than stable individual-difference measurement, therefore preventing from establishing whether the observed deployment tendency reflects what can be defined as a stable participant-level profile.

These limitations define the framework motivating the second experiment. Experiment 1 established that the paradigm can generate an interpretable group-level deployment signature, but it also showed that the task does not yield robust accuracy-based acquisition for either dependency; at this juncture, Experiment 2 serves a different inferential purpose. The main question it poses concerns which, if any, of the indices derived from the current paradigm are reliable enough to support individual-level interpretation, therefore aiming to evaluate the measurement of the behavioural architecture defined in the in-person implementation of the task at group-level. Accordingly, the two experiments should not be simply framed as a replication sequence, but as complementary empirical steps in assessing the novel paradigm, through a reliability-oriented estimation of whether the behavioural organization stressed at group-level might be used to describe interindividual variability and to be informative in terms of individual-level characterization.

Summarizing, Experiment 1 supports a view of statistical learning as a multi-componential process in which acquisition, cognitive availability, and behavioural authority may diverge, making it worthwhile to address the phenomenon through multiple, complementary measurement channels depending on independent, yet related performance attributes, in order to reveal forms of behavioural organization that classical accuracy-based measures alone would miss.

EXPERIMENT 2: Online implementation

METHODS

Building on the task architecture described in Experiment 1, Experiment 2 retained the same distinction among accuracy-based sensitivity, potential availability, and conflict deployment, while shifting the inferential target from group-level interpretability to participant-level characterization. The central methodological issue was to determine whether the indices previously illustrated were sufficiently stable, interpretable, and non-redundant to support a cautious characterization of interindividual variability in terms of acquisition, deployment, availability, response organization, and competition sensitivity.

This shift in aim required a different treatment of stimulus-list variability. As in Experiment 1 employing multiple constrained lists was an essential measure to distribute item-specific perceptual and associative idiosyncrasies across participants, consenting to protect group-level inference from undesired dependence on a unique arbitrary stimulus arrangement, in the online version the principal interest concerned the relative position of participants on the same task-derived metrics. Under this perspective, administering participants with different lists would not have licensed to straightforwardly attribute between-subject variability to differences in acquisition, deployment, availability, or competition sensitivity, as they could have possibly reflected differences in the measurement instrument itself. For this reason, the online measurement context was standardized, and participants received a single list selected from the set of constraint-validated lists that was generated for the in-person implementation, which, therefore, preserved the same probabilistic grammar, filler-balancing logic, and overall test architecture.

The final online behavioural dataset included 61 participants, each completing the 56-trial test phase. Demographic descriptors were available from the Prolific recruitment export for a subset of 59 participants, where participants appeared to range in age from 20 to 50 years (M = 34.44, SD = 8.17), and included 34 women and 25 men. 31 participants were resident in the United States and 28 in the United Kingdom and all reported English as their primary language and normal or corrected vision. The online procedure preserved the temporal and decisional structure described in the general Methods section and was implemented through Pavlovia to participants recruited through Prolific, which completed the familiarization stream without overt responses and without being informed about the embedded regularities. They then completed the post-exposure two-alternative forced-choice test, including global trials, local trials, structurally matched baseline trials, and contrast trials. Since the architecture of the task, the role of the parallel baseline, and the logic of contrast trials have already been described, they would not be reintroduced here. The primary difference was that the present task was not proposed as an isolated experiment, but as the final task of a broader battery designed to assess interindividual variability across several statistical learning paradigms and that included tasks differing in sensory modality, structural format, and type of dependent measure, including more implicit and more explicit probes of learning. It is important to acknowledge this broader framework, as the repeated exposure to statistical-learning tasks may have increased participants' general sensitivity to the possibility, commonly extended across tasks, that the stimuli proposed in each one of them contained hidden structure. To address such a criticality, the order of the tasks was organized so as to interleave modalities and response demands, thereby reducing the likelihood that the immediately preceding task would provide direct cues for the immediately following paradigm and, additionally, temporal constraints were imposed between each session. Thus, although a general task-set or structure-searching orientation cannot be completely ruled out in the online sample, the previous tasks, all instantiating more traditional and validated statistical learning assessments, did not provide specific information about the global–local contingencies, or the response structure of the current task.

At this point it becomes essential to discuss the methodological status of the participant-level indices and on the diagnostic procedures used to determine which of them could be meaningfully interpreted. The core indices preserved the main inferential structure of the paradigm: baseline-adjusted global and local accuracy learning, global–local learning balance, conflict-deployment preference, and corrected global–local orientation. Those metrics, even though summarized at group level to preserve continuity with the in-person pilot, yielded a different evidential emphasis and epistemological status, as greater methodological weight was given in the present analysis to participant-level structure than to the sole presence or absence of mean-level effects. This distinction recognizes a crucial methodological dissociation whereby a task may fail to produce a robust group-level acquisition effect while still containing structured variability across individuals, or, conversely, may show a group-level tendency but remain unsuitable for profiling if participants are shown not to be consistently ordered on it.

For this reason, a methodological crucial choice was to treat the exploratory indices not as an undifferentiated extension of the core battery, but to first submit them to a diagnostic evaluation designed to determine which measures could be sustainably interpreted as participant-level descriptors. This diagnostic step was necessary because, although the inclusion of each metric was theoretically motivated, their interpretation needs to be constrained by the respective degree of statistical support, which substantially differed across exploratory families. Indeed, availability and accuracy-based indices were computed from relatively broad trial classes, whereas reaction-time, variability, and conflict-contingent indices often depended on smaller subsets defined by correctness, choice type, and response times, yielding therefore a reduced subset of trials deemed sufficiently valid. Under this perspective, the diagnostic analysis functioned as a methodological filter located between theoretical relevance and extent of interpretative use, whose role lies in identifying indices that are too unstable, sparse, or redundant to contribute substantively to the individual-variability narrative.

The diagnostic evaluation pipeline integrated three complementary criteria. The first one, *distributional adequacy, was* defined as the graphically derived evaluation of whether an index showed participant-level distributions that were excessively skewed, too sparse, or dominated by a small number of extreme observations. The second metric was *split-half reliability*, which assessed whether participants retained comparable relative positions when the index was recomputed on independent subsets of trials (i.e. halves of the task). The third criterion leveraged the concept of *within-family redundancy*, therefore assessing whether an index contributed distinguishable information within its theoretical family or was largely overlapping to the point of duplicating another measure. Together, these criteria operationalized the conceptual rationale motivating a hierarchical treatment of the exploratory battery rather than a uniform one: some indices could support substantive interpretation, others could be retained as secondary qualifiers, and others could remain descriptive only.

As above stated, Spearman–Brown-corrected split-half reliability was used to quantify the extent to which an index preserved participants' relative ordering across independent and arbitrarily defined portions of the task. The underlying measurement logic can be formalized in a relatively straightforward question: if the same behavioural construct is estimated from two independent subsets of trials, do participants who score relatively high in one half also tend to score relatively high in the other half? Therefore, the Spearman–Brown correction estimates the reliability to be expected for each metric directly from the correlation observed between the two half-length estimates.

At a practical level, the split-half reliability was estimated by repeatedly and randomly dividing the relevant trials into two halves and recomputing the participant-level index separately in each half, then correlating the two half-level estimates across participants, and applying the Spearman–Brown correction; with the whole procedure iterated across 500 random splits. Median Spearman–Brown reliability was retained as the primary reliability estimate, as less sensitive than mean values to unstable or extreme split configurations.

The diagnostic procedure above described was deliberately not implemented as a rigid psychometric pass-or-fail rule, a criterion that would have been admittedly inappropriate for a novel exploratory paradigm which yielded only a limited number of trials per condition after separating trials by branch, correctness, conflict choice, or response-time dispersion. Reliability was therefore used as a "methodological safeguard" against interpretative over-reaching, instructing a methodologically motivated ranking of the exploratory indices on the basis of the interpretability of their distributions, the extent to which split-half reliability was deemed as comparatively acceptable, and redundancy values. In this sense, the diagnostic analysis decided the evidential weight that the index could legitimately carry, outlining a hierarchical arrangement of the most defensible exploratory descriptors within each conceptually defined family and their status in terms of individual profiling.

More broadly, the interpretation of these coefficients was guided by recent methodological work on individual differences in statistical learning and experimental cognitive tasks more generally. It is relatively common in statistical learning paradigms to show low-to-moderate reliability, including even those tasks that are considered well-established and are consistently employed across studies, owing to the fact that many such paradigms were originally optimized to detect group-level sensitivity to statistical structure rather than to provide precise participant-level measurement. Arnon (2020), for example, showed that commonly used auditory and visual statistical-learning tasks can have limited reliability as measures of stable individual differences, especially when used in developmental or cross-task contexts, with Siegelman et al. (2017) similarly arguing that individual differences in statistical learning are psychometrically demanding, as issues relative to task specificity, measurement noise, and limited convergence across paradigms complicate the interpretation of a single behavioural score as a stable learning ability. It must be considered that this problem is not unique to statistical learning: Hedge et al (2018) work is foundational in describing a broader "reliability paradox", whereby robust experimental effects may yield poor individual-difference reliability precisely because successful experimental paradigms often minimize between-subject variability. All those considerations are especially relevant for the present paradigm, where many of the indices were difference scores or higher-order, derived measures of contrasts. Such indices can reasonably be expected to be more fragile than simple accuracy scores because each additional subtraction or conditional split, although necessary in conceptual terms, sensibly reduces the amount of stable participant-level information available for estimation. Work on learning-task reliability has further emphasized how reliability estimates can be highly sensitive to the number of trials, the splitting procedure, and the kind of dependent variable: whether it is accuracy-based, RT-based, or derived from more complex contrast structures.

On the other hand, within-family redundancy was assessed by computing correlations among all the indices belonging to the same theoretical family and summarizing, for each index, its mean absolute correlation with the other family members. Importantly, high correlations were not treated as intrinsically problematic, owing to the fact that, when several indices are deliberately designed to capture aspects of the same theoretical construct, a certain degree of convergence is expected and may even support the coherence of the family itself. Accordingly, two indices, belonging to the same theoretical family, that are strongly associated, may indicate that their operationalization's logic is correctly sampling a shared underlying dimension rather than unrelated behavioural noise. However, a diagnostic concern might emerge under conditions of excessive redundancy in relation to interpretative use, with very high within-family correlations suggesting that what were originally presented as multiple indices should not be used as independent sources of evidence. Conversely, low correlations within a family may indicate useful differentiation, but, on the other hand, signal that the family is empirically heterogeneous or that some indices are too noisy or imprecisely defined to capture the intended construct. Within-family correlation needs therefore to be interpreted as an equilibrium criterion along two possibly problematic polarities: accordingly, sufficient convergence can be taken as evidence that the conceptual family had empirical coherence, whereas excessive overlap limits by definition the evidential independence of its members. Under this perspective, the redundancy diagnostic factor was employed to regulate evidential weight rather than to automatically exclude correlated measures. Moreover, bearing in mind the outcomes of the diagnostic analysis, highly correlated indices could be retained when they represented theoretically distinct operationalizations of the same family, given that their interpretation is circumscribed as complementary descriptions of a shared dimension instead of separate and converging findings.

Overall, the selection of exploratory qualifiers depended on the joint and critical evaluation of reliability, distributional adequacy, and within-family correlation, with the final choice of the metrics to be preserved covering what is deemed as theoretically relevant to measure a multi-component behavioural system in which acquisition, deployment, conflict processing, and competition sensitivity are postulated to be dissociable, while avoiding an uncontrolled proliferation of outcomes.

The individual-profile analysis was constructed after this diagnostic step as a descriptive mapping of the behavioural state space generated by the task, not aiming to define stable cognitive styles or trait-like learning styles, but to describe specific configurations of core and selected exploratory dimensions within individuals. The primary profile dimensions were derived from the proportion-based core indices and their construction adhered strictly to the present pool of data as profile labels were generated through sample-relative magnitude thresholds. Practically, for each relevant index, a threshold was computed as the median of the absolute finite values of the index itself, yielding specific labels as a function of the direction of the considered value relative to the sample-derived cut-off. Accordingly, values were labelled as positive or global-oriented when above the positive threshold, as negative or local-oriented when below the negative threshold, while those between the two thresholds were classified as weak or neutral, allowing to discriminate between stronger directional expressions and near-neutral qualifiers.

The first categorical family was the *Core acquisition profile*, whose main function was to describe how acquisition-like evidence was distributed across the two branches obtained by crossing global and local baseline-adjusted learning, therefore defining configurations that could be assigned to the label of dual positive learning, selective global or local positive learning, branch-specific disadvantage, dual disadvantage or weak evidence.

The second family pattern was the *Learning–deployment profile*, which was obtained by crossing the direction of global–local learning balance with the direction of conflict-deployment preference. This profile allowed therefore to distinguish aligned configurations, mismatched configurations, but also peculiar patterns such as deployment without clear learning or learning without clear deployment, and, finally, weak learning with weak deployment. This profile family was particularly informative in terms of operationalizing at the individual level the distinction between what appeared relatively favoured in non-conflict accuracy and what became behaviourally authoritative under conflict.

A third descriptor summarized the *Corrected global–local orientation profile*, intuitively derived from the corrected global–local orientation index and providing a compact summary of overall global or local orientation, although not interpretable independently of the two-dimensional learning–deployment profile, as possibly compressing different underlying configurations. In parallel, participants were also located in a continuous learning–deployment space defined by the orthogonal dimensions of learning balance and conflict preference, defining meaningful quadrants of this space corresponding to broad orientation labels. In this graphical arrangement, a conceptually essential attribute is the participant's distance from the origin: participants close to neutrality were not treated as equivalent to participants showing strong directional expression within the same quadrant.

While the core profile described the main acquisition–deployment architecture of the participant's behaviour; those that have been termed as exploratory qualifiers were used to hone the established picture by addressing whether additional behavioural channels supported, complicated, or qualified the emerging architecture. These qualifiers were structured into families whose primary object of inquiry was, respectively, auxiliary availability, conflict-related RT costs, conflict-related variability, correct-response facilitation or stabilization, competition asymmetry, and local functional robustness.

Finally, several cross-tabulations, both count- and percentage-based, were produced to examine how profile dimensions related to each other; describing, for instance, how core acquisition profiles mapped onto learning–deployment patterns, whether acquisition profiles were accompanied by auxiliary availability, how the dimensions of deployment profiles was distributed across conflict-related RT or variability qualifiers, and whether acquisition profiles were linked to structured competition asymmetry. Crucially, these cross-tabulations were strictly descriptive and responded to a systematic attempt to organize the individual-variability space and identify recurring configurations, not to establish categorical learner types.

Experiment 2 reveals it valuable potential in affording to change the level of analysis while preserving inferential discipline, allowing between-subject variability to be examined without list-induced variance, therefore rigorously attempting to address whether the behavioural architecture originally hypothesized and subsequently identified in the in-person pilot could support cautious individual-level characterization, while remaining constrained by distributional adequacy, split-half reliability, and redundancy among indices.

RESULTS

Since Experiment 2 evaluated for the behavioural architecture derived from the global–local paradigm the possibility to support participant-level characterization, the pattern of group-level effects was inspected only as a preliminary, secondary piece of evidence, merely evaluating which aspects of in-person profile were reproduced and which were not.

At the level of sample means, the online implementation converged with the in-person version insofar as it did not provide evidence for robust accuracy-based acquisition, with neither branch-specific learning index showing evidence of above-baseline performance. Indeed the *Global Learning Index* was slightly negative, $M = -0.029$, $SD = 0.238$, $t(59) = -0.95$, $p = .828$, whereas the Local Learning Index was slightly positive but, again, non-significant, $M = 0.017$, $SD = 0.254$, $t\ (59) = 0.52$, $p = .303$ and no reliable global–local learning imbalance was documented here, as the *Preferential Learning Index* was also non-significant, $M = -0.046$, $SD = 0.286$, $t\ (59) = -1.25$, $p = .215$. The global-oriented deployment profile that was observed in the in-person pilot was not reproduced in the present sample, as shown by the *Predictive Deployment Index*, which did not attest any directional conflict preference, $M = -0.061$, $SD = 0.386$, $t\ (59) = -1.23$, $p = .223$. Thus, the online sample did not yield either a classical group-level acquisition profile nor a shared deployment policy.

Notably, when compared with the partially incongruent pattern observed in Experiment 1 sample, the present pool of data revealed that the significantly positive *Learning–deployment Alignment Index* significantly positive, $M = 0.048$, $SD = 0.151$, $t\ (59) = 2.46$, $p = .008$, was in this case complemented by a significant effect for the regression that predicted conflict deployment from accuracy balance, slope = 0.563, $R^2 = .174$, $p = .0009$. The online outcome pattern, in this sense, is only partly complementary to the in-person pilot. While Experiment 1 revealed a shared global-oriented deployment tendency that was, however, not fully corroborated by the weak participant-level coupling between learning balance and deployment, Experiment 2 lacked that shared global tendency, but exhibited a stronger relation between each participant's relative accuracy profile and the preferential structural deployment under conflict. Accordingly, it can be suggested that, once all participants were exposed to the same validated list, the task was less appropriate to express structure as a common group-level bias and more so to capture an individually organized correspondence between accuracy balance and conflict choice.

The online data do not rescue a dual-learning account; indeed the pattern encompassing both the null *Combined Learning Index*, $M = -0.006$, $SD = 0.201$, $t\ (59) = -0.22$, $p = .588$, and the negative *Dual-availability Index*, $M = -0.119$, $SD = 0.220$, $t\ (59) = -4.18$, $p > .999$ argue against robust simultaneous availability of both dependencies at the level of overt accuracy. The online sample, therefore, redirect interpretation toward a more circumscribed, yet particularly informative claim: participants did not express reliable acquisition or preferential deployment of both structures as a group, but their relative accuracy profiles were meaningfully related to how they resolved conflict.

The diagnostic analysis was a crucial intermediate evaluative component of Experiment 2, allowing to evaluate each measure through the joint contribution of three criteria: whether its distribution was interpretable, the extent of relative stability in the positioning of each participant across repeated split halves, and the meaningfulness of each metric within its conceptual family. This threefold logic was necessary to address a layered logic, purposefully protecting against unstable participant ordering, leveraging on distributional inspection to prevent sparse or outlier-driven variables from being treated as meaningful dimensions and aiming to regulate the evidential independence of indices belonging to the same conceptual family while also testing its empirical coherence.

Adhering to the outlined rationale, the availability family represented the clearest diagnostically defensible component of the whole exploratory architecture. Indeed, the *Combined Learning Index* and the *Dual-availability Index* showed what can be seen as a sufficiently strong split-half reliability, with median Spearman–Brown coefficients of .498 and .418, respectively. Their participant-level distributions were interpretable and showed meaningful variability around an overall weak or negative availability profile. At the same time, the two indices were highly correlated within family, with mean absolute within-family correlation = .911. This convergence is not a weakness: it supports the idea that both measures sample a shared availability dimension, however, it also limits their evidential independence. CLI and DAI should therefore be interpreted as complementary operationalizations of a coherent availability construct, rather than as separate confirming findings.

The learning–deployment family showed a different diagnostic profile, with the *Learning–Deployment Alignment Index* showing sufficiently favourable split-half reliability, median Spearman–Brown = .441, and low redundancy with the basic conflict-deployment measure. This outcome outlines how the metric is especially valuable as it captures an evidential insight not merely duplicating conflict preference; but enriching it with the within-participant convergence, or lack thereof, between the direction of relative learning and the direction of conflict deployment. Its conceptual value is further strengthened by the fact that the online sample, beyond directional alignment, also showed a significant continuous relation between learning balance and deployment, that instead lacked in the in-person pool. Accordingly, *LDAI* emerges as a particularly valuable metric to be retained, owing to the fortunate combination of its theoretical centrality, its sample-dependence in how it is related with regression analysis' outcomes, its acceptable participant-level stability, and its within-family position, which suggests incremental information beyond simple deployment.

The core indices require a more differentiated interpretation relative to exploratory measures, owing to their status of originally defined analytic probes due to their conceptual centrality. Branch-specific learning showed a modest participant-level stability, with the *Local Learning Index* yielding a median Spearman–Brown of .369, and the *Global Learning Index*, exhibiting a median Spearman–Brown of .272. Notwithstanding the non-ideal reliability value of these indices, they however remain central as they define the primary acquisition axes of the task. By contrast, the *Preferential Learning Index* was unreliable, median Spearman–Brown = −.006, an outcome that may be expected once one acknowledges the conceptually relevant, but methodologically fragile nature of the metric itself, specifically as a derived difference of differences' scores. Therefore, the learning-balance index remains descriptively useful for defining the learning–deployment space, especially when considered relationally through its association with deployment and within the broader profile structure, but participant-level inferential conclusions should not rely on it. The *Predictive Deployment Index* was likewise theoretically indispensable due to its direct operationalization of arbitration under conflict, but its modest reliability, median Spearman–Brown = .119, make it less adequate as a strong psychometric variable. The diagnostic pattern emerging from these two last comparative core measures requires to be additionally specified in light of the structural constraints affecting them. First, as above noted, the *Preferential Learning Index* is a difference between two already baseline-adjusted difference scores; a condition that determines its inheritance of the noise associated to both branch-specific estimates while, consequentially, reducing the stable variance available for participant ordering. Difference scores and higher-order contrasts are known to be particularly problematic as individual-difference measures, particularly when derived from elementary component measures that are themselves noisy or, as in the present case, moderately reliable. Second, the *Predictive Deployment Index* is estimated at most from only eight contrast trials (before any exclusion or validity filtering), making its modest reliability interpretable as a consequence of limited sampling. The present analysis therefore retains learning balance and conflict deployment as architecturally necessary indices, while considering their constraints in terms of participant-level interpretation.

The response-dynamic families were significantly less diagnostically convincing, with *Global RT facilitation* reaching comparatively acceptable reliability, median Spearman–Brown = .352, *Local RT facilitation* being essentially unreliable, and the relative RT facilitation index showing, actually, a negative reliability. IQR-based stabilization indices were weaker still, with negative median reliability for global, local, and relative stabilization measures. Thus, although response facilitation and stabilization remain interesting, secondary behavioural probes, the online diagnostics prevents from treating them as meaningful individual-level descriptors. Conflict-related RT and variability measures further illuminate how a group-level effect may describe the sample, but it cannot be automatically implied to provide a stable participant-level descriptor, justifying a dissociation in treating, in this case, conflict variability as an exploratory response-organization phenomenon, but not as a reliable profiling dimension. Exemplar in this sense is the *Congruency-based conflict variability index* which, although showing a significant group-level effect, $M = -140.82$ ms, $SD = 336.95$, $t\,(46) = -2.87$, $p = .006$, nonetheless exhibited strongly negative split-half reliability.

The structured competition family provided a more promising but still uneven diagnostic picture. For instance, the *Global Competition Cost* showing the strongest and most coherent profile within this family, with acceptable reliability (median Spearman–Brown = .408) and an interpretable distribution, while the normalized *Structured Competition Selectivity Index*, although showing high median reliability, .399, exhibited a bimodal distribution requiring some caution. By contrast, the *Local Competition Cost* was significantly positive at group-level analysis, $M = 0.087$, $SD = 0.277$, $t\,(59) = 2.44$, $p = .009$, suggesting an effect of local competition vulnerability, but presented a negative median reliability. Therefore, while in the current sample local performance may be suggested to be vulnerable when both the choice alternatives were structurally "legal"; however, this branch-specific effect can be addressed only as an exploratory group-level pattern and a secondary descriptive qualifier but not as a stable individual-difference variable.

Overall, the hierarchy emerging from the diagnostic analysis is not a strict "good-to-bad" ranking, but a useful constraint of evidential use, ensuring methodological discipline in defining learning profiles, that can be achieved by weighting all the original indices according to the extent to which each measure can support participant-level interpretation (Fig.6).

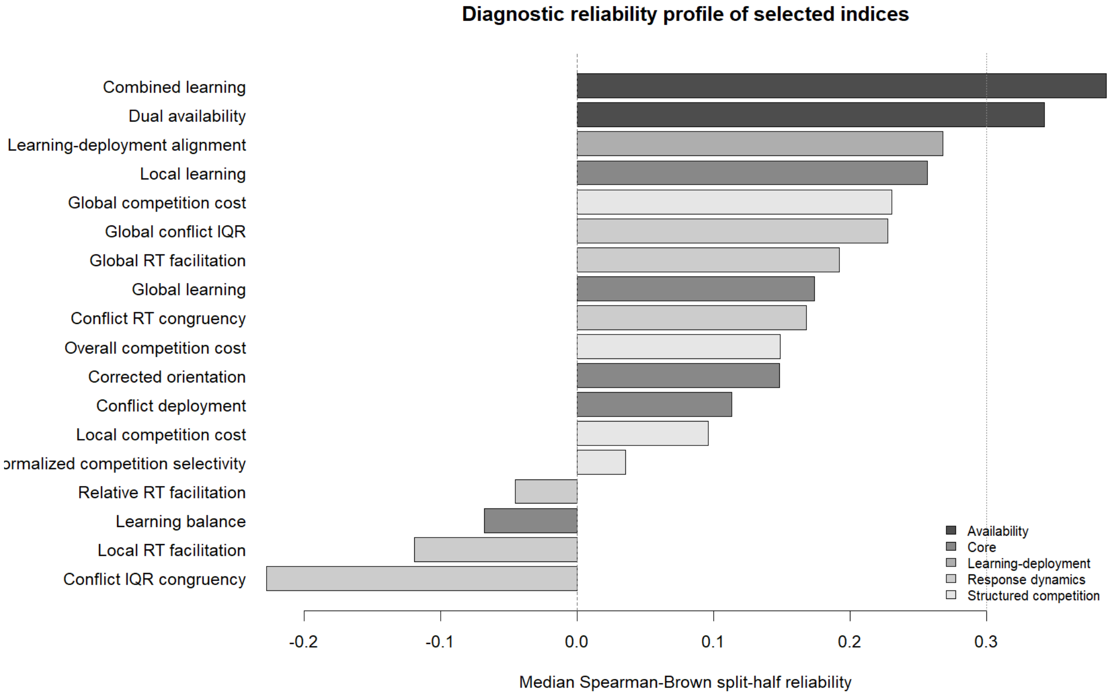

Figure 6. Diagnostic reliability profile of core and selected exploratory indices in Experiment 2. Bars represent median Spearman–Brown-corrected split-half reliability. The figure illustrates the heterogeneous measurement status of the task architecture: availability and learning–deployment alignment indices showed the strongest reliability, whereas several difference-based, RT-based, and conflict-contingent indices were less stable. The dashed vertical line at zero marks the boundary between negative and non-negative reliability; the dotted line at .30 indicates the arbitrary threshold used to identify comparatively defensible participant-level descriptors.

The individual-profile analysis revealed that the online sample did not form a clean typology of learners nor a clearly segregated acquisition space, showing instead a heterogeneous behavioural state space (Fig.7). The largest core acquisition category was weak or mixed learning, including 29.51% of the whole sample. Selective global disadvantage (i.e. worst performance in experimental than in baseline matched trials) accounted for 10 participants, 16.39%, and the mirroring selective local disadvantage was documented for 8 participants, 13.11%. Local-only positive learning was assessed for 9 participants, 14.75%, compared to the 3.28% of the learning profile presenting only global advantage. 5 participants, covering 8.20% of the sample appeared as particularly proficient learners showing a profile of dual positive acquisition, whilst a dual disadvantage is present in another 5 participants. The acquisition space therefore does not show a dominant dual-learning or branch-specific pattern; but is dispersed across weak, asymmetric, and peculiar disadvantage-based configurations.

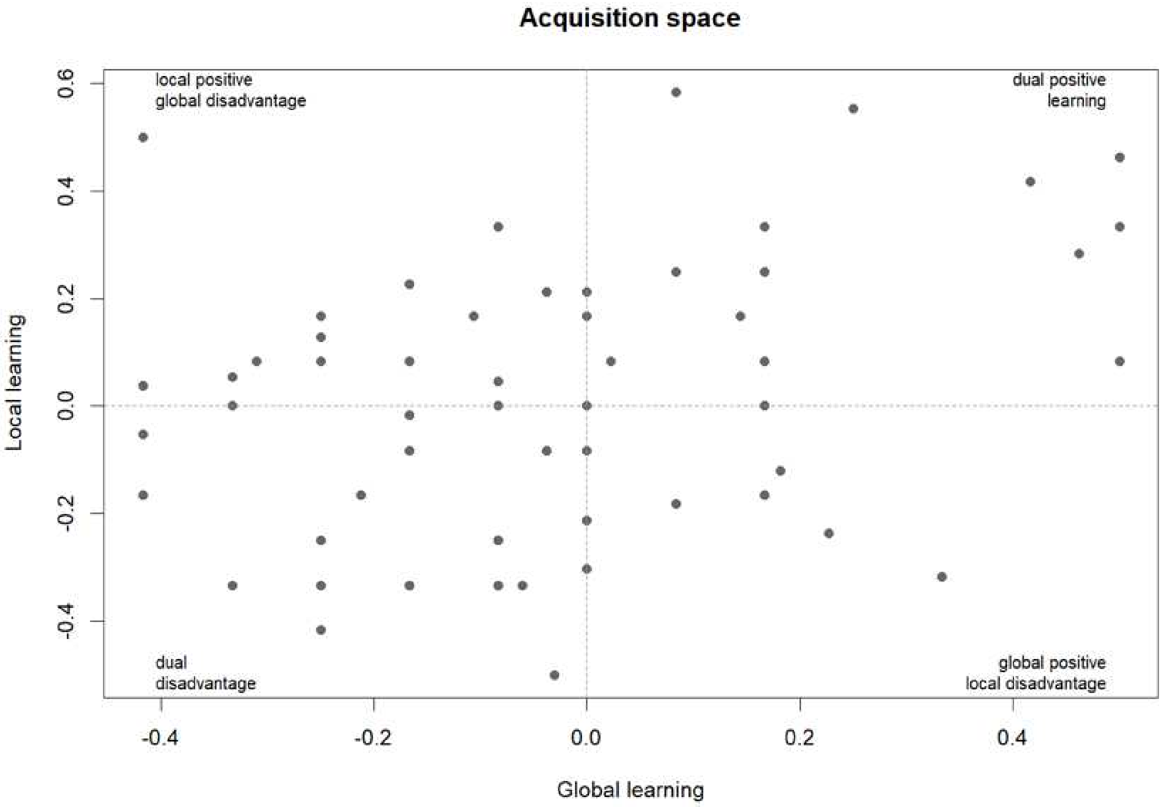

Figure 7. Acquisition space in Experiment 2. Participant-level distribution of baseline-adjusted global and local learning indices. Dashed lines indicate the neutral value of zero for each branch. The upper-right quadrant corresponds to dual positive learning, whereas the lower-left quadrant corresponds to dual disadvantage. The distribution illustrates the absence of a dominant dual-learning pattern and the heterogeneity of acquisition-like profiles.

The learning–deployment profiles provided an informative addition to hone the emerging description of this heterogeneous organization. Here, the most frequent category was a convergence of weak learning with weak deployment, observed in 20 participants, 32.79%. Local learning without clear deployment accounted for 13 participants, 21.31%, while the mirroring profile of global learning without clear deployment was shown by 7 participants, 11.48%. Fully aligned profiles were present but not dominant and evenly distributed across structural dependencies, with 5 participants, 8.20%, showing global learning complemented by global deployment, and 5 participants, 8.20%, showing local acquisition with local behavioural preference. An interesting profile of deployment without clear learning was also observed, with 4 participants, 6.56%, showing global deployment without clear learning and 6 participants, 9.84%, showing local deployment without clear learning. Overall, these outcomes refine the continuous alignment results, validating that learning and deployment were statistically coupled at the group-level, but showing that their categorical individual expression remained partial, weak, and not attributable to a single dominant orientation, being instead equally distributed across contingency types.

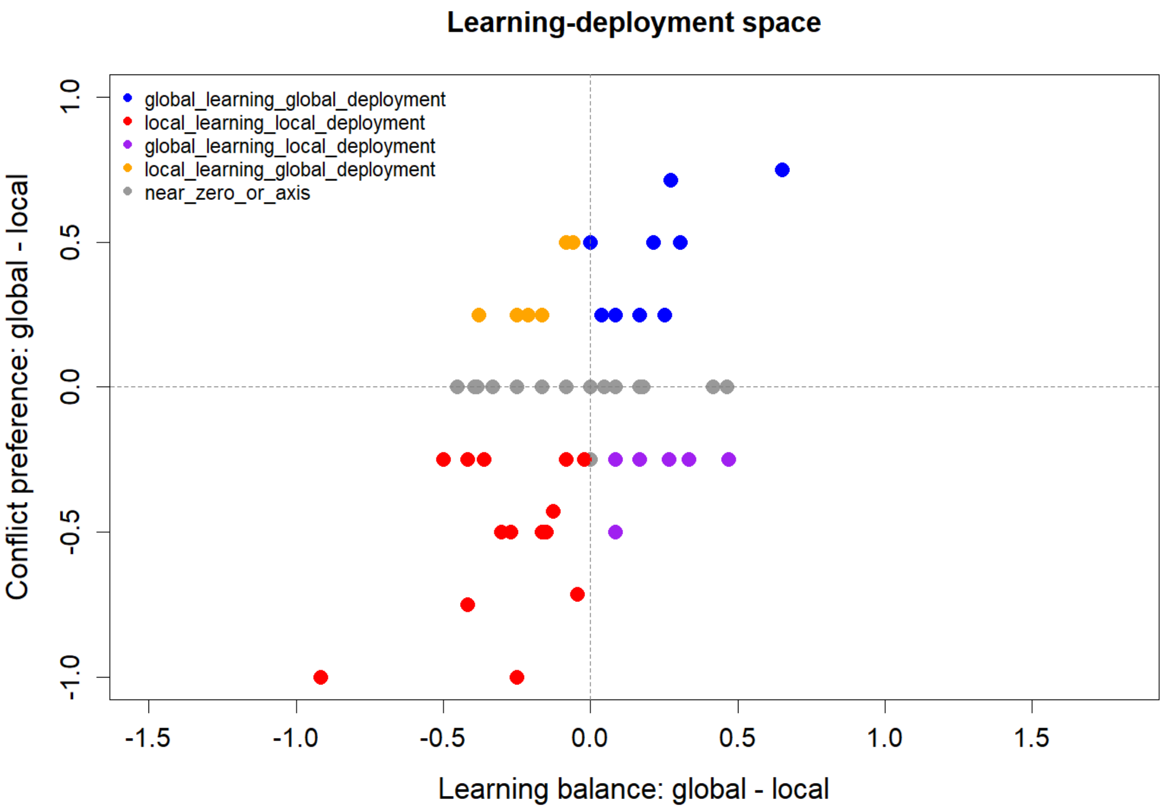

Figure 8. Learning–deployment space in Experiment 2. Participant-level relation between global–local learning balance and conflict-deployment preference. Positive values on the x-axis indicate relatively more global-oriented baseline-adjusted accuracy, whereas positive values on the y-axis indicate greater selection of the globally consistent ending under conflict. Dashed lines indicate neutrality on both dimensions. The quadrant structure distinguishes aligned learning–deployment configurations from mismatched ones, while points near the axes indicate weak or incomplete directional expression.

The broad learning–deployment space was defined by the quadrants' arrangement that emerged from crossing the orthogonal dimensions of learning and deployment in order to concurrently characterize each individual location within the resulting space. Such a representation allows to visually confirm a partial learning–deployment coupling for the online sample, but not a categorical division into stable learner types. Twenty participants, 32.79%, fell near zero or on one of the axes, indicating weak or incomplete directional expression. Instead, among participants occupying directional quadrants, aligned configurations were more frequent than mismatches: 15 participants, 24.59%, showed local learning with local deployment, and 11 participants, 18.03%, showed global learning with global deployment. Notably, mismatched configurations were less frequent but still present, with 9 participants, 14.75%, showing global learning with local deployment and 6 participants, 9.84%, showing local learning with global deployment.

Auxiliary profile dimensions further constrained this picture. Availability profiles combined both the *Combined Learning Index* and the *Dual Availability Index,* defining a pattern whereby 35 participants, 57.38%, showed no auxiliary availability support, whereas a non-negligible portion of 14 participants, 22.95%, showed dual or combined availability support and 11 participants, 18.03%, showed combined learning without dual availability. This is overall consistent with the negative DAI result and reinforces the conclusion that dual availability was definitely not the dominant online profile, although it appears to be frequently co-occurring with sample-relative, above-threshold combined learning in a sizeable portion of subjects. Competition asymmetry profiles suggested a cautious local-vulnerability pattern, convergent with the group-level significance: 21 participants, 34.43%, were classified as showing greater local competition sensitivity, compared with 5 participants, 8.20%, exhibiting greater global competition sensitivity, while 34 participants, 55.74%, remained weak or neutral. The local-competition pattern is therefore directionally meaningful but not pervasive and identifies a tendency for the local-branch to be specifically affected in trials where the correct option competed with a plausible, globally-adherent alternative rather than an implausible filler.

The cross-tabulations showed that profile dimensions were not arbitrary co-occurrences but can be organized in a structured way, with subjects classified as dual positive learners consistently associated with auxiliary availability support. At the same time, acquisition profiles did not automatically map onto deployment profiles, with a non-negligible portion of participants showing branch-specific acquisition patterns but lacking clear deployment, whereas others showed, less intuitively, preferential deployment without clear learning. This dissociation is therefore convergently assessed at both group and individual level, showing the meaningful presence, also at the level of the subjective profile space, of configurations in which acquisition, availability, and deployment are related but non-equivalent dimensions of behaviour.

Taken together, Experiment 2 group-level meaningful insights concern how, once list-induced variance was removed, the online implementation revealed a meaningful coupling of learning balance and conflict deployment at the participant level. Overall, the diagnostic analysis selectively constrained the space of metrics that can sustainably be deemed usable for participant-level interpretation: availability and learning–deployment alignment were the most defensible dimensions, selected competition measures added a cautious qualifier, whereas several RT- and IQR-based measures remained secondary or unstable. Finally, the profiling analysis contributes a level of subjective descriptive characterization, a constrained tool for mapping how different components of statistical behaviour may combine within individuals and revealed a structured but heterogeneous behavioural space characterized by weak or mixed acquisition, limited dual availability, partial learning–deployment coupling, and cautious evidence of local competition sensitivity.

DISCUSSION

Experiment 2 was conceived as a measurement-oriented incremental development of the in-person pilot rather than a mere replication of its group-level effects. Nonetheless, group-level pieces of information may be exploited, although marginally, to further probe the extent to which the weak tendencies assessed in the in-person version of the novel paradigm embody sufficiently meaningful epistemological axes, so much so to be reliably replicated across independent samples. The in-person shared global-oriented deployment bias under conflict, for instance, was not reproduced in the online implementation. Instead, once all participants were exposed to the same validated list, the task revealed a comparatively stronger relation between participants' relative accuracy profile and the structure they deployed under conflict, both in direction and in predictive strength. Consistent with the former version's outcomes, neither global nor local learning showed a reliable above-baseline advantage, combined learning was essentially null, and dual availability was not supported; results instructively converging in concluding the absence of either dependency classes' acquisition in a conventional forced-choice sense. Not to be confused with a strict dismissal of the paradigm, these commonalities in the cross-experiment pattern of results could be better formulated as the absence of a sufficiently pervasive formation of stable, overtly accessible representations of both structures across-subjects. At this point, a seemingly fruitful question concerns the peculiar configuration relating the failure, common across samples, of group-level acquisition, with the extent to which individual non-conflict accuracy coheres or diverge with conflict-based deployment. Indeed, Experiment 2 appears theoretically informative in showing a positive learning–deployment alignment, presenting a specularly inverted pattern compared to the in-person pilot. While in the first experiment a group-level global deployment tendency emerged, lacking however a strong participant-level coupling, the online sample didn't exhibit a common deployment direction but revealed a stronger correspondence between learning balance and conflict choice, expressed, akin to the former implementation, in the directionality of the 2 components, but in the present case further corroborated in the significative predictive relation among them. This pattern may be taken to suggests that, once list-induced variance was removed, the task expressed structure less as a shared orientation across the whole sample and more as an individually organized relation between non-conflict accuracy and conflict-based arbitration. Despite being modest, the finding is conceptually important in supporting a partial distinction between acquisition-like expression and behavioural deployment, at the same time maintaining their complementarity in characterizing the construct of statistical learning. Therefore, while not collapsing into a single behavioural dimension, deployment under conflict was shown not to be arbitrary with respect to participants' own relative performance profile. Here, the online version is particularly useful in making visible those partially coupled components of performance at the level of participant-specific configuration rather than at the level of mean accuracy or mean deployment. This interpretation appears overall aligned with the broader contemporary, theoretical landscape, where multiple empirical accounts agree in suggesting treating statistical learning as multi-componential and task-dependent rather than as a homogeneous latent ability (Siegelman et al., 2017; Growns et al., 2020).

The diagnostic analysis is in the present case not a technical appendix to the results; it rather represents the necessary constraint that makes the above and following participant-level interpretations legitimately meaningful. The outcome was accordingly deliberately hierarchical, with availability and learning–deployment alignment constituting the most defensible dimensions in light of the convergence of acceptable reliability, distributional adequacy, and conceptual coherence. Moreover, the structured competition family provided a theoretically useful, yet uneven, exploratory qualifier, while process-dynamics based measures, although conceptually interesting, were mostly relegated to secondary descriptive status as their diagnostic profile was too weak to support strong profiling claims.

The modest reliability of some core indices requires a further discussion before being interpreted. Above all, it must be considered that the *Preferential Learning Index* is a difference between two already baseline-adjusted difference scores, that, as such, inherits noise from both the original global and local estimates. Furthermore, the *Predictive Deployment Index* is theoretically indispensable, because it is the only core measure that directly operationalizes arbitration between global and local predictions, however, it must be acknowledged that the metric is estimated from only eight contrast trials before any exclusion or validity filtering. Both the constraints of higher-order score construction and limited trial sampling make psychometric instability quite expected once one considers broader work dissecting how reliance on few trials and computation from difference scores (Hedge, 2018) are amongst the factors that dramatically affect the reliability of cognitive and learning tasks.

For a novel paradigm in which the contrast trials instantiate the most distinctive theoretical manipulation, the operational strategy that was deemed more appropriate was neither to discard the fragile core indices nor to over-promote them. Learning balance and deployment were therefore retained as the primary metrics defining the conceptual architecture of the paradigm, not as a stand-alone measure of stable individual traits: their value emerges in the current sample in a relational sense, when interpreted within the broader diagnostic hierarchy and in combination with more stable descriptors such as availability and learning–deployment alignment. To further illustrate this principle, it is useful to consider the availability family where both the indices of combined learning and dual availability exhibited comparatively favourable reliability and interpretable distributions but were highly correlated with one another. This convergence can be interpreted as suggesting that the family provides two component metrics coherently querying a shared availability dimension. Although, in light of such evidence, they cannot be treated as independent evidence generating a new claim, they could be considered as providing a stable auxiliary dimension against which core acquisition profiles can be interpreted. An additional methodological consideration can be made concerning the response-dynamic measures, that are exemplar in how conceptual attractiveness cannot overcome empirical evidence of weak measurement support. RT facilitation and IQR stabilization were originally included because statistical learning may, in principle, modulate behavioural outcomes at a finer granularity, possibly revealing peculiar fluency, uncertainty, or response consistency patterns above and beyond accuracy. Although they are not part of those that are defined as implicit measures and that aim to capture individual differences in statistical learning without relying on familiarity-based or recognition-like decisions, their original purpose was to address more process-sensitive attributes of performance than accuracy. In the present analysis of the online sample, however, such measures don't directly contribute to the profiling characterization owing to their methodological weakness.

Overall, the profiling attempt based on sample-relative thresholds did substantiate only a cautious interpretation: indeed, and converging with group-level outcomes, the sample was not clearly divided into clear categories of global, local, or dual learners and the largest acquisition category was weak or mixed learning, with a dual positive pattern of acquisition that was present only in a small minority of participants. This heterogeneity directly mirrors the partially dissociable components of statistical behaviour, with individual profiles showing that participants exhibited various configurations of a multifarious landscape encompassing acquisition-like performance, availability, deployment, and competition sensitivity. Accordingly, a behavioural state space emerges from Experiment 2 rather than an exact taxonomy of learners.

Within that state space, the non-negligible presence of what is labelled as a "disadvantage profile" deserves a more nuanced interpretation than simple learning failure. Because the core indices were baseline-adjusted, negative values can be taken as indices of either a comparatively poorer experimental performance, a relatively stronger baseline performance, or both. While this phenomenon appears as a, at least peculiar, instantiation of "anti-learning", the underlying logic motivating such a paradoxical effect may be more articulate. In a task deliberately designed to be particularly demanding, the post-exposure 2AFC additionally requires translating a weak, if present, implicitly acquired structure into an explicit choice between alternatives, a requirement that has been increasingly criticized in literature for partly relying on familiarity or meta-cognitive decisional processes rather than directly reflecting the emergence of an implicit acquisition. Given the restrict outcome space, the issue that dictated the introduction of a parallel baseline may become anew problematic: indeed, across the test phase, some participants may resort to a heuristic grounded on the arguably more immediately available cue of familiarity and completely rely on it. As the mere frequency of occurrence of the predictor-correct coupling, an alternative may become diagnostic if a subject captures its preferential status of co-occurrence relative to filler or mirroring legal ending and such a cue could be unambiguously leveraged to provide a response. Moreover, it may provide an indication that is more easily processed compared to a sensitivity which, assuming it was actually developed, ensued from an uncertain and sometimes contradictory environment, requiring cognitive components that go beyond mere acquisition and entailing different decisional policies and ambiguity tolerance. Under a cognitive economy perspective, therefore, a more immediate, unequivocal cue, may be attributed a greater extent of behavioural authority in an uncertain environment compared to a combination of two weak problematic sensitivities.

Another not marginal profile, that is therefore worthy to be discussed, is the deployment-without-clear-learning one, that provides a second critical hinge. These profiles are theoretically informative because they instantiate, in an admittedly unexpected direction, the dissociation between non-conflict accuracy and conflict-based use. Beyond the possibly more conservative interpretation that such profiles reflect response heuristics, decisional preferences, or strategy-like behaviour emerging when learning evidence is weak, a further hypothesis may be put forward. Indeed, an originally neglected but not implausible possibility could be that contrast trials changed the decision problem itself. While in non-conflict trials participants had to identify the statistically licensed continuation against a baseline architecture that could itself contain familiarity cues, contrast trials may have, in an apparently counterintuitive way, made the structural rationale of the task more visible. These experimental conditions indeed entailed the simultaneous presence of global and local predictors: further, both response alternatives were legal, each one licensed by a different predictive horizon. Conflict may therefore have functioned not only as a more demanding condition, but as a cue that foregrounded the relational nature of the task, providing potentially a clearer manifestation of the experimental rationale. Despite not proving hidden acquisition, such considerations may suggest that conflict can make behavioural arbitration more manifest than isolated accuracy.

The structured competition further qualifies the discussed result, with the online data suggesting that local performance was more vulnerable when the competing alternative was itself structurally legal. This pattern should not be overstated, especially because the local competition index was not diagnostically strong as an individual-level measure; nonetheless its theoretical meaningfulness at the group level may provide an empirically motivated ground to discuss it. It is overwhelmingly recognized that adjacent dependencies are easier or more robust in their acquisition in simpler statistical learning contexts, as they allow rapid prediction and may leverage on locally constrained information, thus not requiring elapsed maintenance of active, yet unresolved, contingencies. However, their advantage may depend on whether they remain *uniquely* diagnostic. In the present task, local predictions competed within a shared legal outcome space and when the alternative was itself structurally plausible, local performance appeared more fragile. Conversely, the global dependency, inherently assumed to be more difficult (and indeed not expressed as stronger accuracy learning), may have benefited relatively more in contexts where the task explicitly required arbitration. Such an exploratory interpretation may complement the above hypothesis concerning the constitutive difficulty of contrast trials: if the cueing power of such experimental condition is indeed presumed, the deriving advantage should be selectively more helpful for the structure that is comparatively impoverished at baseline. Overall, such a consideration, although tentative, is especially interesting in complicating a simple proximity-based account and supporting a more dynamic interpretation than a simple equation between empirically documented acquisition effects and robustness profile. It has been previously, and extensively, argued that local and global structures are not merely easier and harder versions of the same computation, but different predictive regimes whose usefulness depends on the context in which they must guide behaviour (Gómez, 2002; Newport & Aslin, 2004; Deocampo, King, & Conway, 2019). Accordingly, local information may be especially relied upon when it is uniquely diagnostic, but may not necessarily be as decisive when another legal and structurally licensed alternative is present, while, conversely, globally-derived predictions may be harder to express in ordinary accuracy yet become more functionally relevant when the task foregrounds relational arbitration. The online battery context also qualifies the interpretation, since participants completed the task as the final component of a broader individual-differences battery involving several statistical learning paradigms. This makes Experiment 2 different from the in-person pilot not only in administration format and list standardization, but also in cognitive context, preventing definitively excluding that participants may have become generally sensitized to the possibility that experimental materials contained hidden structure. Aligned to the contrast trials hypothesis, it could be claimed that participants may have become generally sensitized to the possibility that experimental streams contain hidden structure, thus, again, providing a cueing adjunct that was comparatively more useful for global structure. However, it must be noted that the previous tasks differed in modality, materials, structure, and dependent measures, and did not provide any specific indication about the contrast between global and local contingencies, that was relevant only for the present task. I would therefore parsimoniously opt for not dismissing online findings as mere task-set effects, while neither interpreting them as contextually identical to a naïve first exposure either.

A crucial significance of Experiment 2 lies in its treatment of variability that, rather than being treated as noise around a true learning effect, becomes itself one of the objects of analysis, allowing to conceive statistical learning under probabilistic conflict as not exhausted by the presence or absence of above-baseline accuracy. Rather, establishing a conceptual continuum with the theoretical architecture of the present work, a better conceptualization could instruct a definition akin to "*structured behavioural negotiation among partially available predictive commitments, their decisional authority, and the contexts in which they are required to govern choice*".

GENERAL DISCUSSION

The present dissertation was originally motivated by a central limitation that can be identified in the statistical learning literature: namely, the tendency to treat behavioural success as an exhaustive, revealing proxy about the nature of the acquired representation. The purpose of the presented paradigm was to deliberately challenge the fragility of this and other canonical statistical learning assumptions, as well as the overwhelmingly shared rationale of its empirical implementation. To this aim a learning environment was constructed, which maintained an ecologically aligned continuity with occurrences of ordinary learning, in which statistical structure is rarely singular, deterministic, or behaviourally univocal. This was practically achieved by defining the embedding of two probabilistic dependencies, matched in formal predictive strength but differing in temporal scope, within the same visual stream, which converged on the same legal outcome space, and would be later placed in direct competition. An architecture of this sort can present some empirical criticality, in that it was not designed to produce the cleanest possible acquisition effect. However, I consider the paradigm a valuable asset, insofar as it offers the opportunity to test what happens when the learner is confronted with a situation closer to the theoretical problem that defines ordinary learning environments; where multiple partially informative structures are available, none eliminates uncertainty, and behavioural response requires a process of selection favouring one predictive commitment over another. This logic is continuous with earlier discussed arguments that propose an interpretation of statistical learning as a composite, multifaceted system, encompassing cognitive computations that span from evidence accumulation to availability maintenance, from prioritization to deployment and that can definitively be not conceived as a homogeneous mechanism that a single score is sufficient to capture.

The strongest empirical conclusion is restrictive: neither the in-person nor the online implementation showed reliable above-baseline accuracy in both global and local dependencies, and the availability indices exclude any claim of simultaneous clear behavioural expression of both dependencies. Therefore, the paradigm, in its present form, cannot be presented as a successful demonstration that uninstructed participants acquired two coexisting probabilistic structures. Crucially, such a limitation is not theoretically sterile, as it indicates, and replicates, that when regularities are probabilistic, temporally arranged, visually instantiated, mutually concurrent, and mapped onto a shared response space, a question such as "*did learning occur*?" may become too underspecified. Under this perspective a dissecting, more elementary logic can be adopted and potentially become informative: once failure can be attributed to different sequential steps, an outcome that lacks stabilization becomes attributable to an interindividually shared or a subjectively idiosyncratic breakdown at each level of the whole process.

Under this perspective, in order to unveil the attributes of the phenomenon of statistical learning, a principled attempt of decomposing a negative result may be beneficial. While a simple null effect in a standard forced-choice cannot disambiguate between equally sustainable hypotheses of learning failures, lack of behavioural expression, or reliance on a decision criterion misaligned with the intended structure, here, the index architecture was deliberately built to make that ambiguity analytically distinguishable. For instance, baseline-adjusted learning indices tested predictor-specific expression while controlling for unintended, but potentially effective, additional learning cues; availability indices addressed whether the two structures were representationally co-active and, finally, conflict trials tackled the issue of behavioural dominance under ambiguity. The paradigm therefore yields the potential to refine an observed weak acquisition pattern by inquiring from multiple directions where the behavioural chain becomes, thus placing originally isolated evidence within an informative map.

The critical empirical dissociation across experiments concerns deployment: while in the first version the reliable group-level tendency to deploy the global ending under conflict was not mirrored by an above-chance global accuracy in non-conflict trials, Experiment 2, revealed a stronger participant-level relation between relative accuracy balance and conflict deployment (although in the absence of the shared global tendency of the previous sample). This interesting discrepancy suggests how the same paradigm can be used to expose different layers of behavioural organization, strictly depending on the inferential use of the design itself. In the present case, as it would be expected, the counterbalanced in-person version protected group-level inference from item-specific idiosyncrasies deriving from a material that, although abstract and consistently used in SL paradigms, cannot be excluded to elicit perceptual, mnestic, imaginative preferential biases, thus emerging as a more adequate way to reveal collective arbitration tendencies. On the other hand, the removal in the single-list online version of list-induced variance made participant-level correspondence more interpretable, providing a methodologically informative divergence that licenses treating counterbalancing and standardization not as merely alternative procedural choices, but as different epistemic instruments.

Furthermore, an important conceptual advance concerns the epistemological status of conflict trials and the need to understand them as something more than difficult test probes. This set of trials is indeed shown to significantly alter the epistemic nature of the task: while during non-conflict, the participant is asked to select the continuation associated with one dependency against a matched alternative, here both alternatives are legal, and both are supported by different predictive histories. The request shifts from a simple recognition to a more articulate decisional policy, hinging on learning features, extent of ambiguity tolerance, strategic approaches to learning environments and arbitration between candidate organizations of the stream. Under this perspective, conflict trials can be conceived to operate as probes to classify both subjective and common regime-selection, a framing that is closer to latent-structure accounts of learning, whereby organisms infer the hidden cause or regime that best explains current evidence, than to a simple associative account, in which exposure linearly and predictably strengthens one response tendency (Gershman & Niv, 2010).

Another construct that is honed by the present data is what has been defined as the global–local contrast. Rather than simply corroborating the shared tendency to invariably subordinate acquisition of non-adjacent statistical contingencies to locally instantiated dependencies, the current pool of data seems to suggest that the functional value of a regularity depends on the decision ecology within which it is queried. Local dependencies undeniably provide dense and rapidly testable evidence, yielding the potential to either confirm or refute the ongoing predictive organization, but this consistency advantage may be argued to be particularly useful in stable environments, where the process of evidential accumulations proceeds in a quasi-linear deterministic fashion. In a more volatile environment, the status of extended under-determinacy of non-adjacent evidence may favour flexibility and the deriving uncertainty may better align to a dynamic informational stream. Indeed, when a local ending competes with another legal and statistically plausible alternative, proximity appears to be no longer sufficient to guide choice, while, conversely, a global dependency may become comparatively more useful when the task explicitly foregrounds relational arbitration in an ambiguous environment. The structured competition findings point in the same direction, with the local dependency appearing more vulnerable when the competing alternative was itself structurally legal. Although excluding strong claims about a stable local-deficit mechanism, it can be assumed that adjacent dependencies, while retaining a more immediately accessible epistemological status, may also more strictly constrain interpretation. In a condition where ambiguity cannot be ultimately and definitively resolved by relying on either dependency, the representationally stronger alternative may suffer comparatively more than one whose epistemic status is constitutively open, in which a degree of residual uncertainty is not sufficient to exclude its relative plausibility. Such explanations remain tentative; however, they aim to provide a more nuanced reading than a simple difficulty hierarchy: local and global structures are expected to instantiate different predictive regimes, each one with distinct conditions of robustness and fragility and preferentially relied upon under different task requirements.

The overall componential logic applies both to the more conceptual characterization of the phenomenon, but also to the methodological contribution of the diagnostic analysis, whose hierarchical classification outcome adds a fine-grained lesson: reliability is not a property of "the task" alone, but of each component derived from it. This consideration could motivate how the present paradigm should be treated and how it could be developed.

A first and more conservative improvement concerns the diagnostic power of the task itself, as the present implementation was constrained by its position within a broader online battery (and the methodological choice to make the in-presence version exactly mirroring the design that would be used for the online sample). Such a logistic condition constrained the design by limiting the number of informative trials available for several indices, especially contrast-based and conditional measures. The limitation is not trivial, both in terms of the somewhat limited power to identify group-level tendencies without being strongly affected by trial-specific performance fluctuations, and in terms of the detrimental impact on participant-level reliability when estimating indices from few observations. A future implementation should therefore definitively increase the number of trials particularly contrast trials, before adding any further complexity to the paradigm, thus strengthening the diagnostic resolution of the task and allowing negative, weak, or unstable indices to be more unambiguously interpreted as properties of the behavioural architecture or as consequences of insufficient measurement opportunity.

Beyond this necessary increase in diagnostic resolution, a particularly fruitful direction could be to extract from it what can be considered as an experimental grammar, whose components can be recombined to isolate distinct inferential operations. The more consequential issue is that the current task succeeded in creating the intended theoretically relevant ambiguity, but did not yet fully decompose it, thus placing the learner in a situation where contingency acquisition, terminal position familiarity, legal-ending preferential status and conflict-based deployment could all, indistinguishably, contribute to the same overt choice. This indeed was the reason why the paradigm was deemed informative in the first place; however, it is also the critical factor in explaining why its behavioural outcomes could not be unequivocally assigned to a single underlying operation. A potentially fruitful future development could therefore implement a principled decomposition of the current design, aimed at specifically distinguishing which component of statistical behaviour is responsible for the effect that appears as well as which component collapses when the effect is not there.

A first point of decomposition concerns the relation between predictor–ending contingency and legal-ending status. As repeatedly acknowledged, in the present task the endings were not completely neutral response objects: some shapes had acquired the status of possible terminal outcomes during familiarization and could therefore be treated as plausible completions independently of the specific predictor–ending mapping. Such a feature could allow participants to rely on a broader, less relationally specific source of evidence than the originally intended contingency. A future implementation could leverage on this criticality and, accordingly, directly manipulate the relation between familiarity and contingency validity, therefore allowing the critical contrast to not be simply restricted to trained and untrained continuations but extending it to compare endings that are familiar as legal endings without being valid for the current predictor, valid for the current predictor without being independently familiar as legal endings, both familiar and valid, or neither. Such a manipulation would contribute a precious insight on disentangling the extent to which subjects are guided by the relational structure linking predictor and ending rather than their global familiarity, therefore providing a solid ground to empirically address one of the central ambiguities exposed by the "disadvantage profiles": whether negative baseline-adjusted learning reflects absence of sensitivity, or the adoption of an aspecific but more immediately available familiarity criterion.

The second issue concerns the unresolved definition of conflict as a test of deployment, or, instead, as a potential, foregrounding cue to structure. A future implementation should manipulate the temporal and informational status of contrast trials. Conflict trials' placement could be in this sense particularly informative: if all the conflictual set would be administered only after a complete non-conflict test phase, they could be used to assess whether a previously acquired structure becomes behaviourally dominant when alternatives compete, if they would instead be introduced earlier, they may be evaluated in the extent to which they train participants to attend to the global–local opposition itself. Comparing these conditions would possibly allow to distinguish actual deployment as the expression of an already available predictive commitment, from fictitious deployment, that could be better defined as a strategy elicited by the structural transparency of the test.

An additional juncture of decomposition concerns the temporal *locus* of measurement, as the current response-dynamic measures were informative only in a limited sense, being collected during the post-exposure 2AFC phase and therefore indexing how participants executed a decision after the phase of familiarization, rather than how predictive structure emerged during exposure. The constructs of coexistence and prioritization entail a dynamic evolution and progression, a learning trajectory that can be sampled in a future version of the current paradigm. To this aim, a range of possible metrics can be exploited, spanning from anticipatory probes, intermittently required prediction judgments, oculomotor information, or more classically employed implicit exposure-phase markers. Altogether, this would allow an online exploration of whether global and local dependencies are tracked in parallel, if one becomes dominant over time, or they both remain weak and untraceable until the test format imposes an arbitration problem. The present proposal would be not merely adding "implicit" measures, but to distinguish the alignment of each process-dynamic metric to the stage of the inferential process it is supposed to index.

A further refinement involves the formal treatment of behaviour. A criticality that has been identified in the current indices was their contrastive nature, making them particularly fragile measurement-wise. A potential alternative could be to methodologically translate the notion that each trial-level response may be the product of several simultaneously available evidential channels, thus adopting a characterization in terms of the latent evidence-weighting profile. Accordingly, it could be modelled whether a subject relies mainly on a predictor-specific contingency, on ending familiarity or on a conflict-specific decision policy; therefore, avoiding the assumption that each behavioural dimension can be "cleanly" recovered from a single, derived score.

The rationale underlying all the proposed methodological refinements reflects the attempt to make the paradigm a genuinely meaningful tool for statistical learning research, aiming to target a gap that canonical tasks rarely address, namely, the transition from having been exposed to an structurally ambiguous stream to allowing one structure to dictate behaviour when another plausible structure competes with it, isolating the juncture at which representation, uncertainty, and action meet.

Even though, currently, the present data cannot support claims about neural mechanisms, the same logic applied to the behavioural section can be mirrored in defining a broader neuroscientific implementation of the project. A relevant research direction here would concern whether the different moments of the learning-to-use trajectory can be temporally and representationally separated, using a finer-neural grain to evaluate whether global and local dependencies are tracked during exposure and the different temporal dynamics through which they stabilize, as well as to discriminate whether one such contingency gains a preferential availability before the test or requires behavioural arbitration under conflict to be selectively amplified.

If anything, the current behavioural version shows that post-exposure choice might be too late in the processing chain to determine whether weak accuracy reflects absent learning, poor availability, criterion competition, or fragile deployment, therefore, neural measures could prove especially useful to address this limitation. During familiarization, frequency-tagging and analyses of neural entrainment could test the potential co-occurring monitoring of global and local regularities by the system, connecting the paradigm to evidence that cortical activity can simultaneously track hierarchical structures at different temporal scales, (Ding et al., 2016), or to work showing that temporal receptive windows are hierarchically organized across cortical areas (Lerner et al., 2011). Nonetheless, it should be clarified that the present paradigm's critical question would not concern the aspecific existence of multiscale tracking, but also whether two experimentally matched predictive horizons can coexist when they are instantiated in the same visual stream and later compete for behavioural authority.

Concerning the first issue of representational availability, it was only indirectly inferred, at a behavioural level, from baseline-adjusted indices and was definitely not classifiable as dual overt access. In this sense, a neural version could test whether this absence of dual availability reflects a genuine failure of encoding or a defect stemming from behavioural access. Using representational similarity analyses (RSA) it could for instance be evaluated whether items belonging to the same predictive regime and defined by the same temporal granularity become more similar in neural representational space after exposure. This would be especially useful in showing that statistical structure can reorganize neural representational spaces, possibly through medial-temporal and cortical systems' contribution to relational binding and event-structure learning.

Finally, conflict trials should be considered as simultaneously the most theoretically distinctive and the most interpretatively ambiguous (at least from behavioural evidence) component. A neural implementation could prove essential in disentangling the current ambiguity, by relating the interpretation of final responses in conflict trials to the dynamics of evidence weighting during exposure. If a global or local neural signature is already present during familiarization and becomes selectively amplified once conflict trials appear, deployment could be attributed merely to behavioural arbitration and selection of a pre-existing representational candidate; if instead the relevant signature emerges only after repeated contrast exposure, the interpretation should favour a conceptualization of contrast trials as effectively restructuring learning.

Such a programme would also refine the interpretation of individual variability, honing the interesting evaluation of global or local learners' profiles through evidence concerning possible differences in the stage at which one predictive horizon becomes dominant; effectively shifting to a process-level account of variability, in which individual differences are localized along a holistic trajectory from encoding to availability to deployment.

I would frame the final value of the paradigm as the extent to which it preserves what I deem as an interesting theoretical ambition, while recognizing what needs to be improved to upscale its inferential resolution. The current version is not yet a robust measure of dual statistical learning, nor a mature instrument for individual-difference assessment, yet it identifies a criticality that can be yield promising research avenues: when several probabilistic structures are simultaneously plausible, learning cannot be equated with recognition, and recognition cannot be equated with behavioural authority. Under this perspective, the most informative insight offered by the present work is the separation, currently imperfect but theoretically motivated, between what can be captured through ordinary accuracy, what becomes representationally available as a candidate structure, and what is in the end authorized to guide behaviour when competing predictions are placed in conflict. Statistical learning, in this sense, is more than an intrinsic, phylogenetically and ontogenetically continuous proclivity to silently register order within multiplicity; it is the fragile act of determining what is worthwhile to put trust upon, in order to prospect one possible future among many.

BIBLIOGRAPHY

A Single Paradigm for Implicit and Statistical Learning - Monaghan - 2019 - Topics in Cognitive Science - Wiley Online Library. n.d. Retrieved 31 January 2026. https://onlinelibrary.wiley.com/doi/full/10.1111/tops.12439.

Arciuli, Joanne, Janne von Koss Torkildsen, David J. Stevens, and Ian C. Simpson. 2014. 'Statistical Learning under Incidental versus Intentional Conditions'. *Frontiers in Psychology* 5. doi:10.3389/fpsyg.2014.00747.

Baldassano, Christopher, Janice Chen, Asieh Zadbood, Jonathan W. Pillow, Uri Hasson, and Kenneth A. Norman. 2017. 'Discovering Event Structure in Continuous Narrative Perception and Memory'. *Neuron* 95(3):709-721.e5. doi:10.1016/j.neuron.2017.06.041.

Behrens, Timothy E. J., Mark W. Woolrich, Mark E. Walton, and Matthew F. S. Rushworth. 2007. 'Learning the Value of Information in an Uncertain World'. *Nature Neuroscience* 10(9):1214–21. doi:10.1038/nn1954.

Benjamin, Lucas, Ana Fló, Marie Palu, Shruti Naik, Lucia Melloni, and Ghislaine Dehaene-Lambertz. 2023. 'Tracking Transitional Probabilities and Segmenting Auditory Sequences Are Dissociable Processes in Adults and Neonates'. *Developmental Science* 26(2):e13300. doi:10.1111/desc.13300.

Bogaerts, Louisa, Noam Siegelman, and Ram Frost. 2016. 'Splitting the Variance of Statistical Learning Performance: A Parametric Investigation of Exposure Duration and Transitional Probabilities'. *Psychonomic Bulletin & Review* 23(4):1250–56. doi:10.3758/s13423-015-0996-z.

Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns | PNAS. n.d. Retrieved 31 January 2026. https://www.pnas.org/doi/10.1073/pnas.1508523113.

Brysbaert, Marc. 2024. 'Designing and Evaluating Tasks to Measure Individual Differences in Experimental Psychology: A Tutorial'. *Cognitive Research: Principles and Implications* 9(1):11. doi:10.1186/s41235-024-00540-2.

Clark, Andy. 2013. 'Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science'. *The Behavioral and Brain Sciences* 36(3):181–204. doi:10.1017/S0140525X12000477.

Conway, Christopher M. 2020. 'How Does the Brain Learn Environmental Structure? Ten Core Principles for Understanding the Neurocognitive Mechanisms of Statistical Learning'. *Neuroscience and Biobehavioral Reviews* 112:279–99. doi:10.1016/j.neubiorev.2020.01.032.

Conway, Christopher M., Holly E. Jenkins, Alice E. Milne, Sonia Singh, and Benjamin Wilson. 2025. 'Addressing the Theory Crisis in Statistical Learning Research'. *Npj Science of Learning* 10(1):68. doi:10.1038/s41539-025-00359-6.

Creel, Sarah C., Elissa L. Newport, and Richard N. Aslin. 2004. 'Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences'. *Journal of Experimental*

Psychology. Learning, Memory, and Cognition 30(5):1119–30. doi:10.1037/0278-7393.30.5.1119.

Daw, Nathaniel D., Yael Niv, and Peter Dayan. 2005. 'Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control'. *Nature Neuroscience* 8(12):1704–11. doi:10.1038/nn1560.

Deocampo, Joanne A., Tricia Z. King, and Christopher M. Conway. 2019. 'Concurrent Learning of Adjacent and Nonadjacent Dependencies in Visuo-Spatial and Visuo-Verbal Sequences'. *Frontiers in Psychology* 10. doi:10.3389/fpsyg.2019.01107.

de Diego-Balaguer, Ruth, Anna Martinez-Alvarez, and Ferran Pons. 2016. 'Temporal Attention as a Scaffold for Language Development'. *Frontiers in Psychology* 7. doi:10.3389/fpsyg.2016.00044.

Ding, Nai, Lucia Melloni, Hang Zhang, Xing Tian, and David Poeppel. 2016. 'Cortical Tracking of Hierarchical Linguistic Structures in Connected Speech'. *Nature Neuroscience* 19(1):158–64. doi:10.1038/nn.4186.

Emberson, Lauren L., Christopher M. Conway, and Morten H. Christiansen. 2011. 'Timing Is Everything: Changes in Presentation Rate Have Opposite Effects on Auditory and Visual Implicit Statistical Learning'. *Quarterly Journal of Experimental Psychology* 64(5):1021–40. doi:10.1080/17470218.2010.538972.

Emerson, Samantha N., and Christopher M. Conway. 2023. 'Chunking Versus Transitional Probabilities: Differentiating Between Theories of Statistical Learning'. *Cognitive Science* 47(5):e13284. doi:10.1111/cogs.13284.

Eronen, Markus I., and Laura F. Bringmann. 2021. 'The Theory Crisis in Psychology: How to Move Forward'. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16(4):779–88. doi:10.1177/1745691620970586.

Expectations Do Not Alter Early Sensory Processing during Perceptual Decision-Making | Journal of Neuroscience. n.d. Retrieved 31 January 2026. https://www.jneurosci.org/content/38/24/5632.

Feldman, Harriet, and Karl Friston. 2010. 'Attention, Uncertainty, and Free-Energy'. *Frontiers in Human Neuroscience* 4. doi:10.3389/fnhum.2010.00215.

Friston, Karl. 2010. ‘The Free-Energy Principle: A Unified Brain Theory?’ *Nature Reviews Neuroscience* 11(2):127–38. doi:10.1038/nrn2787.

Frost, Ram, Blair C. Armstrong, and Morten H. Christiansen. 2019. ‘Statistical Learning Research: A Critical Review and Possible New Directions’. *Psychological Bulletin* 145(12):1128–53. doi:10.1037/bul0000210.

Frost, Rebecca L. A., Andrew Jessop, Samantha Durrant, Michelle S. Peter, Amy Bidgood, Julian M. Pine, Caroline F. Rowland, and Padraic Monaghan. 2020. ‘Non-Adjacent Dependency Learning in Infancy, and Its Link to Language Development’. *Cognitive Psychology* 120:101291. doi:10.1016/j.cogpsych.2020.101291.

Gebhart, Andrea L., Richard N. Aslin, and Elissa L. Newport. 2009. ‘Changing Structures in Midstream: Learning Along the Statistical Garden Path’. *Cognitive Science* 33(6):1087–1116. doi:10.1111/j.1551-6709.2009.01041.x.

Gebhart, Andrea L., Elissa L. Newport, and Richard N. Aslin. 2009. ‘Statistical Learning of Adjacent and Nonadjacent Dependencies among Nonlinguistic Sounds’. *Psychonomic Bulletin & Review* 16(3):486–90. doi:10.3758/PBR.16.3.486.

Gershman, Samuel J., and Yael Niv. 2010. ‘Learning Latent Structure: Carving Nature at Its Joints’. *Current Opinion in Neurobiology* 20(2):251–56. doi:10.1016/j.conb.2010.02.008.

Gómez, R. L., and L. Gerken. 2000. ‘Infant Artificial Language Learning and Language Acquisition’. *Trends in Cognitive Sciences* 4(5):178–86. doi:10.1016/s1364-6613(00)01467-4.

Gómez, Rebecca L. 2002. ‘Variability and Detection of Invariant Structure’. *Psychological Science* 13(5):431–36. doi:10.1111/1467-9280.00476.

Growns, Bethany, Noam Siegelman, and Kristy A. Martire. 2020. ‘The Multi-Faceted Nature of Visual Statistical Learning: Individual Differences in Learning Conditional and Distributional Regularities across Time and Space’. *Psychonomic Bulletin & Review* 27(6):1291–99. doi:10.3758/s13423-020-01781-0.

Hasson, Uri, Janice Chen, and Christopher J. Honey. 2015. ‘Hierarchical Process Memory: Memory as an Integral Component of Information Processing’. *Trends in Cognitive Sciences* 19(6):304–13. doi:10.1016/j.tics.2015.04.006.

Hedge, Craig, Georgina Powell, and Petroc Sumner. 2018. 'The Reliability Paradox: Why Robust Cognitive Tasks Do Not Produce Reliable Individual Differences'. *Behavior Research Methods* 50(3):1166–86. doi:10.3758/s13428-017-0935-1.

Heilbron, Micha, and Maria Chait. 2018. 'Great Expectations: Is There Evidence for Predictive Coding in Auditory Cortex?' *Neuroscience* 389:54–73. doi:10.1016/j.neuroscience.2017.07.061.

Isbilen, Erin S., Rebecca L. A. Frost, Padraic Monaghan, and Morten H. Christiansen. 2022. 'Statistically Based Chunking of Nonadjacent Dependencies'. *Journal of Experimental Psychology. General* 151(11):2623–40. doi:10.1037/xge0001207.

Ivanov, Yavor, Jan Theeuwes, and Louisa Bogaerts. 2024. 'Reliability of Individual Differences in Distractor Suppression Driven by Statistical Learning'. *Behavior Research Methods* 56(3):2437–51. doi:10.3758/s13428-023-02157-7.

Kabdebon, C., M. Pena, M. Buiatti, and G. Dehaene-Lambertz. 2015a. 'Electrophysiological Evidence of Statistical Learning of Long-Distance Dependencies in 8-Month-Old Preterm and Full-Term Infants'. *Brain and Language* 148:25–36. doi:10.1016/j.bandl.2015.03.005.

Kabdebon, C., M. Pena, M. Buiatti, and G. Dehaene-Lambertz. 2015b. 'Electrophysiological Evidence of Statistical Learning of Long-Distance Dependencies in 8-Month-Old Preterm and Full-Term Infants'. *Brain and Language* 148:25–36. doi:10.1016/j.bandl.2015.03.005.

van der Kant, Anne, Claudia Männel, Mariella Paul, Angela D. Friederici, Barbara Höhle, and Isabell Wartenburger. 2020. 'Linguistic and Non-Linguistic Non-Adjacent Dependency Learning in Early Development'. *Developmental Cognitive Neuroscience* 45:100819. doi:10.1016/j.dcn.2020.100819.

Kurby, Christopher A., and Jeffrey M. Zacks. 2008. 'Segmentation in the Perception and Memory of Events'. *Trends in Cognitive Sciences* 12(2):72–79. doi:10.1016/j.tics.2007.11.004.

Lengyel, Gábor, Goda Žalalytė, Alexandros Pantelides, James N. Ingram, József Fiser, Máté Lengyel, and Daniel M. Wolpert. 2019. 'Unimodal Statistical Learning Produces Multimodal Object-like Representations' edited by J. Diedrichsen, M. J. Frank, M. Landy, and S. J. Gershman. *eLife* 8:e43942. doi:10.7554/eLife.43942.

Lerner, Yulia, Christopher J. Honey, Lauren J. Silbert, and Uri Hasson. 2011. 'Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story'. *Journal of Neuroscience* 31(8):2906–15. doi:10.1523/JNEUROSCI.3684-10.2011.

Liu, Helen, Tess Allegra Forest, Katherine Duncan, and Amy S. Finn. 2023. 'What Sticks after Statistical Learning: The Persistence of Implicit versus Explicit Memory Traces'. *Cognition* 236:105439. doi:10.1016/j.cognition.2023.105439.

Lynn, Christopher W., Ari E. Kahn, Nathaniel Nyema, and Danielle S. Bassett. 2020. 'Abstract Representations of Events Arise from Mental Errors in Learning and Memory'. *Nature Communications* 11(1):2313. doi:10.1038/s41467-020-15146-7.

Maheu, Maxime, Stanislas Dehaene, and Florent Meyniel. 2019. 'Brain Signatures of a Multiscale Process of Sequence Learning in Humans' edited by F. de Lange and T. E. Behrens. *eLife* 8:e41541. doi:10.7554/eLife.41541.

Marcus, G. F., S. Vijayan, S. Bandi Rao, and P. M. Vishton. 1999. 'Rule Learning by Seven-Month-Old Infants'. *Science* 283(5398):77–80. doi:10.1126/science.283.5398.77.

Nassar, Matthew R., Robert C. Wilson, Benjamin Heasly, and Joshua I. Gold. 2010. 'An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment'. *Journal of Neuroscience* 30(37):12366–78. doi:10.1523/JNEUROSCI.0822-10.2010.

Nemeth, Dezso, Karolina Janacsek, Gabor Csifcsak, Gabor Szvoboda, James H. Howard Jr, and Darlene V. Howard. 2011. 'Interference between Sentence Processing and Probabilistic Implicit Sequence Learning'. *PLOS ONE* 6(3):e17577. doi:10.1371/journal.pone.0017577.

Newport, Elissa L., and Richard N. Aslin. 2004. 'Learning at a Distance I. Statistical Learning of Non-Adjacent Dependencies'. *Cognitive Psychology* 48(2):127–62. doi:10.1016/s0010-0285(03)00128-2.

Ordin, Mikhail, Leona Polyanskaya, and Arthur G. Samuel. 2021. 'An Evolutionary Account of Intermodality Differences in Statistical Learning'. *Annals of the New York Academy of Sciences* 1486(1):76–89. doi:10.1111/nyas.14502.

(PDF) Time for Consensus: Open Challenges and Recommendations for Research on Individual Differences in Statistical Learning. 2025.

https://www.researchgate.net/publication/395708831_Time_for_Consensus_Open_Challenges_and_Recommendations_for_Research_on_Individual_Differences_in_Statistical_Learning.

Pearl, Lisa, and Sharon Goldwater. 2016. 'Statistical Learning, Inductive Bias, and Bayesian Inference in Language Acquisition'. in Vol. 1, edited by J. L. Lidz, W. Snyder, and J. Pater. Oxford University Press.

Pedraza, Felipe, Bence C. Farkas, Teodóra Vékony, Frederic Haesebaert, Romane Phelipon, Imola Mihalecz, Karolina Janacsek, Royce Anders, Barbara Tillmann, Gaën Plancher, and Dezső Németh. 2024. 'Evidence for a Competitive Relationship between Executive Functions and Statistical Learning'. *Npj Science of Learning* 9(1):30. doi:10.1038/s41539-024-00243-9.

Perruchet, Pierre, and Sebastien Pacton. 2006. 'Implicit Learning and Statistical Learning: One Phenomenon, Two Approaches'. *Trends in Cognitive Sciences* 10(5):233–38. doi:10.1016/j.tics.2006.03.006.

Pudhiyidath, Athula, Neal Morton, Rodrigo Duran, Anna Schapiro, Ida Momennejad, Demitrius Hinojosa-Rowland, Robert Molitor, and Alison Preston. 2022. 'Representations of Temporal Community Structure in Hippocampus and Precuneus Predict Inductive Reasoning Decisions'. *Journal of Cognitive Neuroscience* 34:1736–60. doi:10.1162/jocn_a_01864.

Reber, Arthur S. 1967. 'Implicit Learning of Artificial Grammars'. *Journal of Verbal Learning & Verbal Behavior* 6(6):855–63. doi:10.1016/S0022-5371(67)80149-X.

Romberg, Alexa R., and Jenny R. Saffran. 2010. 'Statistical Learning and Language Acquisition'. *Wiley Interdisciplinary Reviews. Cognitive Science* 1(6):906–14. doi:10.1002/wcs.78.

Rouder, Jeffrey N., and Julia M. Haaf. 2019. 'A Psychometrics of Individual Differences in Experimental Tasks'. *Psychonomic Bulletin & Review* 26(2):452–67. doi:10.3758/s13423-018-1558-y.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996a. 'Statistical Learning by 8-Month-Old Infants'. *Science* 274(5294):1926–28. doi:10.1126/science.274.5294.1926.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996b. 'Statistical Learning by 8-Month-Old Infants'. *Science* 274(5294):1926–28. doi:10.1126/science.274.5294.1926.

Schapiro, Anna C., Emma Gregory, Barbara Landau, Michael McCloskey, and Nicholas B. Turk-Browne. 2014. 'The Necessity of the Medial Temporal Lobe for Statistical Learning'. *Journal of Cognitive Neuroscience* 26(8):1736–47. doi:10.1162/jocn_a_00578.

Schapiro, Anna C., Timothy T. Rogers, Natalia I. Cordova, Nicholas B. Turk-Browne, and Matthew M. Botvinick. 2013. 'Neural Representations of Events Arise from Temporal Community Structure'. *Nature Neuroscience* 16(4):486–92. doi:10.1038/nn.3331.

Schapiro, Anna C., Nicholas B. Turk-Browne, Kenneth A. Norman, and Matthew M. Botvinick. 2016. 'Statistical Learning of Temporal Community Structure in the Hippocampus'. *Hippocampus* 26(1):3–8. doi:10.1002/hipo.22523.

Siegelman, Noam, Louisa Bogaerts, Morten H. Christiansen, and Ram Frost. 2017. 'Towards a Theory of Individual Differences in Statistical Learning'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372(1711):20160059. doi:10.1098/rstb.2016.0059.

Siegelman, Noam, Louisa Bogaerts, Ofer Kronenfeld, and Ram Frost. 2018. 'Redefining "Learning" in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities?' *Cognitive Science* 42:692–727. doi:10.1111/cogs.12556.

Siegelman, Noam, and Ram Frost. 2015. 'Statistical Learning as an Individual Ability: Theoretical Perspectives and Empirical Evidence'. *Journal of Memory and Language* 81:105–20. doi:10.1016/j.jml.2015.02.001.

Smith, Cybelle M., Sharon L. Thompson-Schill, and Anna C. Schapiro. 2024. 'Rapid Learning of Temporal Dependencies at Multiple Timescales'. *bioRxiv* 2024.01.15.575748. doi:10.1101/2024.01.15.575748.

The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex | PLOS One. n.d. Retrieved 1 February 2026. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036399.

Turk-Browne, Nicholas B., Justin Jungé, and Brian J. Scholl. 2005. 'The Automaticity of Visual Statistical Learning'. *Journal of Experimental Psychology. General* 134(4):552–64. doi:10.1037/0096-3445.134.4.552.

Turk-Browne, Nicholas B., Brian J. Scholl, Marvin M. Chun, and Marcia K. Johnson. 2009. 'Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities without

Awareness'. *Journal of Cognitive Neuroscience* 21(10):1934–45. doi:10.1162/jocn.2009.21131.

Wilson, Robert C., Andra Geana, John M. White, Elliot A. Ludvig, and Jonathan D. Cohen. 2014. 'Humans Use Directed and Random Exploration to Solve the Explore-Exploit Dilemma'. *Journal of Experimental Psychology. General* 143(6):2074–81. doi:10.1037/a0038199.

Xu, Luzi, Chris L. E. Paffen, Stefan Van der Stigchel, and Surya Gayet. 2024. 'Statistical Learning Facilitates Access to Awareness'. *Psychological Science* 35(9):1035–47. doi:10.1177/09567976241263344.

Yon, Daniel, and Chris D. Frith. 2021. 'Precision and the Bayesian Brain'. *Current Biology: CB* 31(17):R1026–32. doi:10.1016/j.cub.2021.07.044.

Zacks, Jeffrey M., Nicole K. Speer, Khena M. Swallow, Todd S. Braver, and Jeremy R. Reynolds. 2007. 'Event Perception: A Mind/Brain Perspective'. *Psychological Bulletin* 133(2):273–93. doi:10.1037/0033-2909.133.2.273.

Zhou, Haoyu, Sabine van der Ham, Bart de Boer, Louisa Bogaerts, and Limor Raviv. 2024. 'Modality and Stimulus Effects on Distributional Statistical Learning: Sound vs. Sight, Time vs. Space'. *Journal of Memory and Language* 138:104531. doi:10.1016/j.jml.2024.104531.