Review

Enriched learning: behavior, brain, and computation

Brian Mathias [1,2,*] and Katharina von Kriegstein [2,*]

The presence of complementary information across multiple sensory or motor modalities during learning, referred to as multimodal enrichment, can markedly benefit learning outcomes. Why is this? Here, we integrate cognitive, neuroscientific, and computational approaches to understanding the effectiveness of enrichment and discuss recent neuroscience findings indicating that crossmodal responses in sensory and motor brain regions causally contribute to the behavioral benefits of enrichment. The findings provide novel evidence for multimodal theories of enriched learning, challenge assumptions of longstanding cognitive theories, and provide counterevidence to unimodal neurobiologically inspired theories. Enriched educational methods are likely effective not only because they may engage greater levels of attention or deeper levels of processing, but also because multimodal interactions in the brain can enhance learning and memory.

The facilitative role of enrichment in learning

The use of institutions and techniques that are designed to enhance learning is a strictly human phenomenon: schools and educational curriculums have a major role in shaping the development of children and adults. Although classroom-based learning has many advantages, school settings often lack the wealth of information that is available in more naturalistic settings, where learners have at their disposal an abundance of sights, sounds, odors, tastes, and proprioceptive information. At the heart of several approaches to education is the belief that the integration of complementary sensory and motor information into the learning experience, here referred to as **multimodal enrichment** (see Glossary), can enhance learning outcomes by approximating real-world environments. For example, the Montessori Method emphasizes the role of hands-on experience with physical objects and specialized learning materials that integrate the use of multiple senses during learning [1]. **Movement-based learning** techniques, which are widely discussed among educators but not yet widely adopted, encourage the integration of movement into the learning experience [2]. For brevity, we refer to multimodal enrichment as 'enrichment' in this review.

Over the past decade, scientists have gained significant ground in answering the key question of how the presence of complementary information across modalities during learning is capable of benefitting learning outcomes. Recent neuroscientific developments have revealed that the beneficial effects of enriched learning are associated with **crossmodal** brain responses in sensory and motor cortices [3,4]. For example, visual brain regions may respond during the presentation of an auditory stimulus that was paired with a complementary visual stimulus during prior learning. Moreover, it has recently been demonstrated that these crossmodal brain responses are functionally relevant in driving the behavioral benefits of enrichment [5], which is consistent with mathematical theories of brain function [6]. In addition, computational models have begun to account for enrichment benefits [7]. Here, we synthesize these recent advances, identify areas of overlap

Highlights

Multimodal enrichment can enhance learning in a variety of domains, such as letter and vocabulary acquisition, reading, mathematics, music, and spatial navigation.

Cognitive, neural, and computational theories of enrichment attribute the benefits of enriched learning to either multimodal or unimodal mechanisms.

Neuroscience research shows that sensory and motor areas in the brain respond crossmodally to enriched items following learning. These brain responses occur during early stages of perceptual processing.

Recent neurostimulation studies reveal that crossmodal brain responses contribute causally to the behavioral benefits of enrichment. These findings support multimodal theories of enrichment and update and constrain cognitive and computational theories.

The translation of neuroscience-based principles to classroom experiments has yielded promising results.

[1]School of Psychology, University of Aberdeen, Aberdeen, UK
[2]Chair of Cognitive and Clinical Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

*Correspondence:
brian.mathias@abdn.ac.uk (B. Mathias) and
katharina.von_kriegstein@tu-dresden.de (K. von Kriegstein).

between cognitive, neuroscientific, and computational theories of enriched learning, and unify these approaches, with the goal of better understanding why multimodal enrichment serves as an effective educational tool.

Principles of enrichment

The notion that the presentation of information in two or more modalities during learning can be beneficial for learning is not necessarily intuitive. Consider, as an example, the learning of the four lexical tones found in Mandarin Chinese (Figure 1A). If a tone-naïve learner's goal is to auditorily distinguish the pitch contours associated with each tone, then they might intuit that the presentation of information in an additional modality (e.g., visual) to the primary (auditory) learning modality may not be very helpful for learning. The additional information could be distracting, use up scarce attentional resources, and strain working memory systems, thereby hindering overall learning [8,9]. Surprisingly, however, behavioral research conducted over the past two decades indicates that such crossmodal information can, in fact, benefit learning [10,11]. These behavioral benefits of enriched learning are ubiquitous across various types of learning and range from improvements in perception [12] to conceptual memory [13] and mathematical cognition [14] (Box 1).

What sorts of combinatory principle govern enrichment benefits? A key principle is between-modality **congruency**, which can be established across multiple levels of stimulus complexity ranging from congruency of perceptual to semantic features. For example, visual tone marks (perceptual congruence; Figure 1B) are commonly used to learn to differentiate between, and produce, Mandarin tones and, when paired with corresponding auditorily presented tones, benefit post-learning tone recognition [15]. In addition, enriching the learning of specific Mandarin words with drawings that illustrate the meaning of each word (semantic congruence; Figure 1C) can be used to learn the meanings of individual novel words [16–18].

The combination of congruent information across two or more sensory modalities (e.g., auditory and visual) can benefit learning and is called **multisensory** enrichment (Figure 1B,C). Enrichment is also beneficial when it involves a motor component, referred to as **sensorimotor** enrichment. Physical movements can be integrated into learning in a variety of ways. For example, one study found that, during learning, the performance of directional gestures (perceptual congruence; Figure 1D) that imitate the trajectories of Mandarin tones can benefit subsequent tone recognition [19]. The performance of incongruent directional gestures had no benefit (see also [20]). Gestures that enact the meanings of novel words [21][i] are also beneficial (semantic congruence, not shown), unlike the performance of mismatching or meaningless gestures [22].

Perceptual congruencies often arise between stimuli with shared physical sources, such as voices and faces [23]. There are also many examples in everyday life where a consistent arbitrary relation leads to semantic congruency, such as the co-occurrence of the spoken word 'spoon' with the sight of a spoon. Novel statistically regular multimodal associations can be quickly and passively learned, thereby influencing beliefs about congruency and, in turn, perceptual judgments [24,25]. Given that enrichment benefits arise not only from perceptual, but also semantic congruencies, it is likely that they can be induced by relatively arbitrary stimulus combinations, as long as knowledge about the regularity of their co-occurrence has been acquired through prior experience.

Stimuli presented in different modalities can never serve as exact replicas of one another, because they take different physical forms (e.g., light vs. sound) and are registered by different types of receptor. Nevertheless, congruent stimuli can overlap in terms of the information that

they convey. Experiments using highly controlled perceptual stimuli, such as blinking visual dots paired with pure tone beeps, have shown that even the pairing of highly redundant stimuli during learning can benefit post-learning performance on unisensory tasks, such as the perceptual discrimination of visual rhythms [12,26]. These findings suggest that a large amount of information overlap between modalities does not preclude a stimulus from qualifying as enrichment, or may even be a prerequisite for enrichment benefits to occur [23].

Interindividual differences

Does enrichment benefit everyone? Most studies exploring enrichment benefits present group-averaged results. From studies of the **face benefit**, we know that ~76% of people show a benefit of 2 min of face–voice exposure on subsequent auditory-only voice-identity recognition, a proportion replicated in three different experimental samples [27–29]. However, the exact proportion of individuals who benefit is likely to depend on multiple factors, such as the type of learning that one is engaged in, its duration, and the post-learning task from which enrichment benefits are assessed [30].

In addition, although strong evidence for enrichment benefits has been shown in groups of typically developed learners in laboratory studies (e.g., [31]) and educational contexts (e.g., [16]), more rigorous research is needed to understand how enrichment might best be used to support learning in neurodevelopmental disorders and rehabilitation (Box 2).

Why does enrichment benefit learning?

Several explanations for why enriched learning (Figure 2A, Key figure) leads to enrichment benefits (Figure 2B) are offered within the domains of cognitive science, neuroscience, and computational modeling. We categorize these theories into two main camps: multimodal theories and unimodal theories. Multimodal theories propose that enrichment benefits are supported by crossmodal interactions that take place following enriched learning [30,32]. By contrast, unimodal theories propose that enrichment benefits are attributable to general cognitive mechanisms, such as enhanced attention during enriched learning, which boost the sensitivity of unimodal brain responses following learning [33].

Multimodal theories

Cognitive

Cognitive theories (Figure 2C) assume that multimodal stimulus encoding establishes multiple routes to memory retrieval and that interactions between routes can make retrieval faster and more accurate. One example of a multimodal theory in cognitive science is the dual coding theory, which proposes that verbal information, such as spoken and signed language, is represented in a verbal modality, while nonverbal information, such as shapes, environmental sounds, and actions, is represented in a nonverbal modality [32,34]. During recall, each modality can crossmodally activate the other ('crossmodal interaction'; Figure 2C), resulting in enhanced memory for dually coded relative to unimodally encoded stimuli. The cognitive theory of multimedia learning (CTML [35]), integrated model of text and picture comprehension [36], and motor trace theory [37] take similar approaches. CTML makes the additional assumption that learners actively select which auditory and visual information to integrate [35]. Theories of grounded and embodied cognition can also be considered multimodal because they propose that thinking about a previously encountered stimulus evokes the simulation of sensory input and movements that occurred during the learning itself [38] and that this supports learning outcomes [39]. The relative distinctiveness account of sensorimotor enrichment proposes that enriched items 'pop out' relative to non-enriched items in post-learning recognition tests due to the additional dimension along which the enriched items are encoded during learning [40].

information into learning. Benefits of enrichment depend on learned between-modality congruencies, which are commonly perceptual and/or semantic.

Multisensory: involving the combination of information across two or more different sensory modalities.

Multisensory predictive coding framework (MPCF): theory of learning in which the brain encodes a multisensory and/or sensorimotor generative model during enriched learning that facilitates subsequent unisensory recognition.

Neurostimulation: alteration of neural activity using electrical currents or magnetic fields.

Sensorimotor: involving the combination of movement-related information and sensory information.

Source localization: identification of where in the brain an electrical or magnetic signal recorded at the scalp was generated.

Supramodal representations: representations that combine multiple perceptual features and are more idealized than sensory-specific representations.

Transcranial direct current stimulation (tDCS): application of a weak electrical current to the scalp, which can be used to modify neural excitability.

Transcranial magnetic stimulation (TMS): use of magnetic fields to activate nerve cells in the brain.
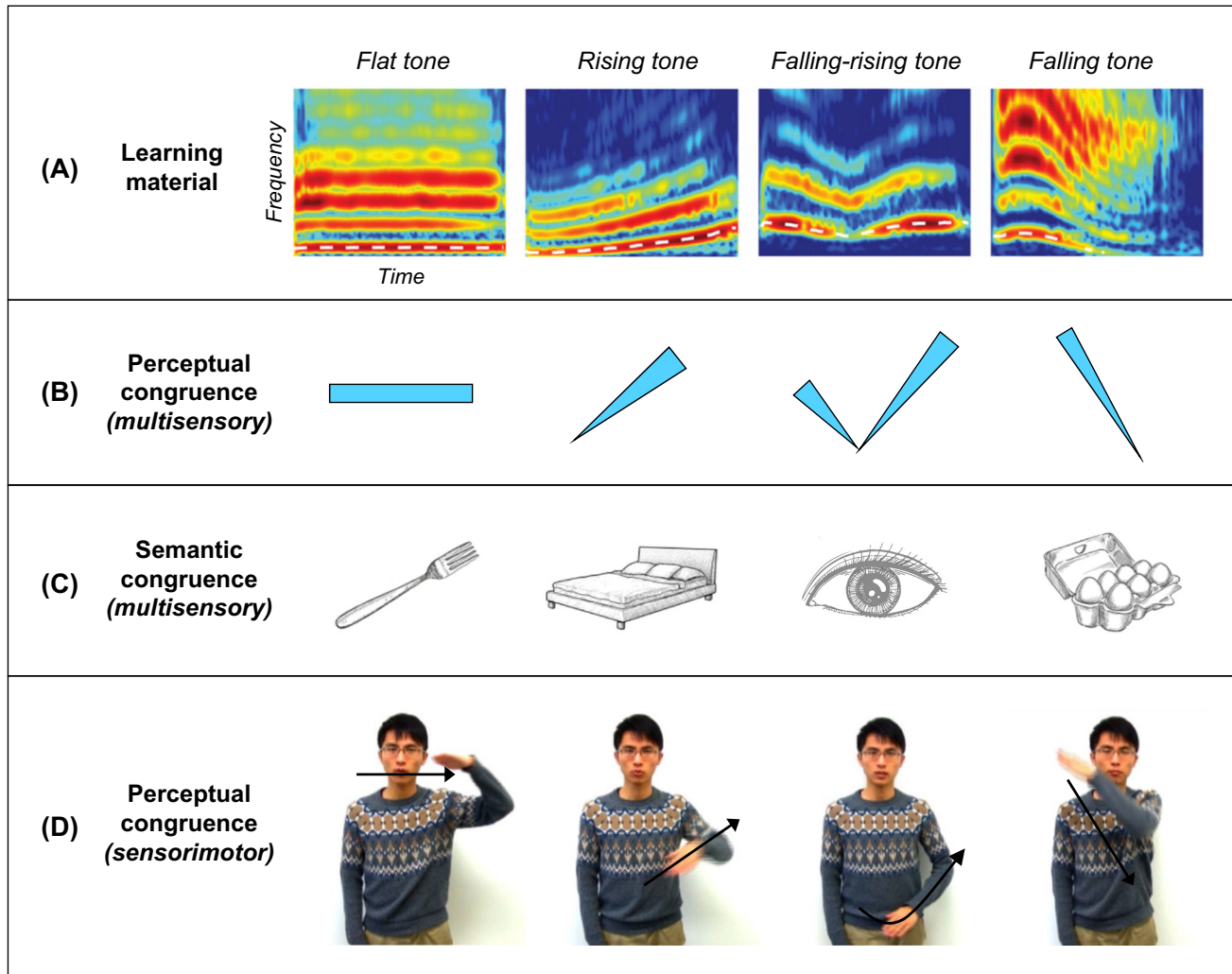
Trends in Cognitive Sciences

Figure 1. Lexical tone-learning material and three types of enrichment material that differ in perceptual and semantic congruency. (A) Learning material. Contours of the four Mandarin lexical tones displayed in auditory spectrograms. Tones are characterized as flat, rising, falling–rising, or falling. Contours within each spectrogram are highlighted by white-broken lines. (B) Visual tone marks that are perceptually congruent with the pitch contour of each tone and that can be used for multisensory perceptual enrichment. (C) Visual representations that are semantically congruent with the meanings of words containing the tones presented in (A) and that are used for multisensory semantic enrichment. (D) Gestural, perceptually congruent, representations of pitch contours. Modified, with permission, from [20] (A) and [19] (D).

Neural

At the level of the brain (Figure 2D), multimodal theories offer two possible explanations for enrichment benefits. The first is that enriched learning induces crossmodal interactions between sensory-specific brain regions during subsequent unisensory tasks. For example, hearing birdsongs that were paired with a complementary visual stimulus during prior learning would be expected to trigger greater visual brain activity ('crossmodal processing'; Figure 2D) compared with birdsongs presented with no accompanying visual stimulus during learning. The **multisensory predictive coding framework** (MPCF [3,30]) proposes that enrichment benefits are supported by crossmodal neural responses as well as by increased **functional connectivity** of relevant sensory-specific cortices (Box 3).

Box 1. Behavioral evidence for multimodal enrichment benefits

Multimodal enrichment benefits the perception of relatively simple visual stimuli used in low-level motion detection and discrimination tasks (e.g., [90]), as well as the learning of more complex, naturalistic stimuli, such as pictures and environmental sounds [54,72], voices [23], speech [29], foreign language vocabulary [3], Morse code [55], and visual sequences [31]. Other lines of work reveal benefits of enrichment that involves a motor component, such as the physical exploration of objects [59], writing text by hand [91,92], tracing [93], performing gestures (i.e., a gesture benefit) [14,22,94], and actively walking through physical environments [95]. The production of auditory stimuli, such as words and melodies, using one's own body, yields enhanced memory for those stimuli than merely perceiving them; this is termed the 'production effect' [62,96].

The translation of laboratory findings to the classroom has yielded promising results. Primary and secondary schoolchildren who receive multisensory enriched reading instruction show enhanced post-learning literacy skills relative to control learners [97,98]. The combination of spoken words with pictures facilitates schoolchildren's language learning [16,99], and audiovisual animations improve the learning of science, technology, engineering, and mathematics (STEM) concepts [13], consistent with the **multimedia principle** [100]. Gesture-based enrichment interventions facilitate the learning of language and mathematics [16,18,101,102].

Do benefits of enrichment emerge immediately during learning or take time to materialize? The point at which enrichment benefits begin to appear depends on factors that are not yet fully understood. Benefits in the recognition of auditory or visual objects can emerge following a single instance of audiovisual exposure [103]. The face benefit requires a bit more time [104]; a single instance is not sufficient, but 2 min of audiovisual voice–face exposure can induce robust behavioral facilitation [23]. The performance of **deictic gestures**, such as pointing, during math instruction can also benefit learning quickly [14]. By contrast, the benefits of performing **iconic gestures** during word learning can take several hours to emerge [3,22,105]. These differences may hinge on factors such as the baseline level of learning needed to perform specific tasks, the concepts that are learned, or the task used to assess benefits. The field currently lacks a systematic assessment of general principles regarding requisite amounts of training.

A second possible multimodal explanation for enrichment benefits is that brain regions known as **convergence zones** generate **supramodal representations** of enriched, but not non-enriched, stimuli. This supramodal approach is taken by the multisensory version of the reverse hierarchy theory of perceptual learning [41]. According to supramodal frameworks, hearing birdsongs that were previously encountered along with a complementary visual stimulus should trigger altered processing in visuomotor convergence zones in the brain ('supramodal processing'; Figure 2D) compared with hearing non-enriched birdsongs. Whereas the MPCF proposes that sensory-specific brain regions

Box 2. Enrichment, treatment, and rehabilitation

Multimodal enrichment is a feature of many interventions intended to support the learning of students who may display learning difficulties, such as students learning in a non-native language and students with neurodevelopmental disorders. In fact, many enriched approaches to teaching and learning were originally designed to aid struggling learners rather than typical learners, because the use of multimodal input might compensate for deficits in learning associated with disabilities or disorders [106]. For example, developmental dyslexia is characterized by difficulties in reading and/or spelling [107]. Programs such as the Wilson Reading System, which pairs spoken phonemes with finger movements, aim to overcome deficits associated with dyslexia by providing multiple routes for audiovisual decoding of written words [108,109]. Unfortunately, the use of these programs is not currently evidence based: the relatively few studies evaluating them do not fulfill important experimental design criteria [110], such as the inclusion of a control group [108,109], or they find no evidence of significant benefits, potentially due to the use of small samples [111]. Many interventions are unlikely to maximally reinforce sound–symbol associations due to low levels of crossmodal congruency.

Deficits in neural functions can also be caused by brain injury, and multimodal enrichment is increasingly being used in rehabilitative therapies. For example, one of the most frequent language-related functional impairments is heightened difficulty retrieving words from memory (anomic aphasia [112]). Traditional therapeutic approaches to anomic aphasia have relied on the use of auditory cues for facilitating retrieval recovery [113]. Although some studies suggest that sensorimotor strategies, such as the observation and execution of gestures associated with the meanings of specific words, also facilitate word retrieval in patients with anomic aphasia [114], other studies report no gesture advantage [115]. These mixed findings are likely due to the heterogeneity of patient populations [116,117], as well as potential differences in task difficulty between studies [118]. Thus, while multimodal enrichment may ultimately provide an important tool in cases of disability and recovery, more rigorous research is needed to unlock potential benefits.

Key figure

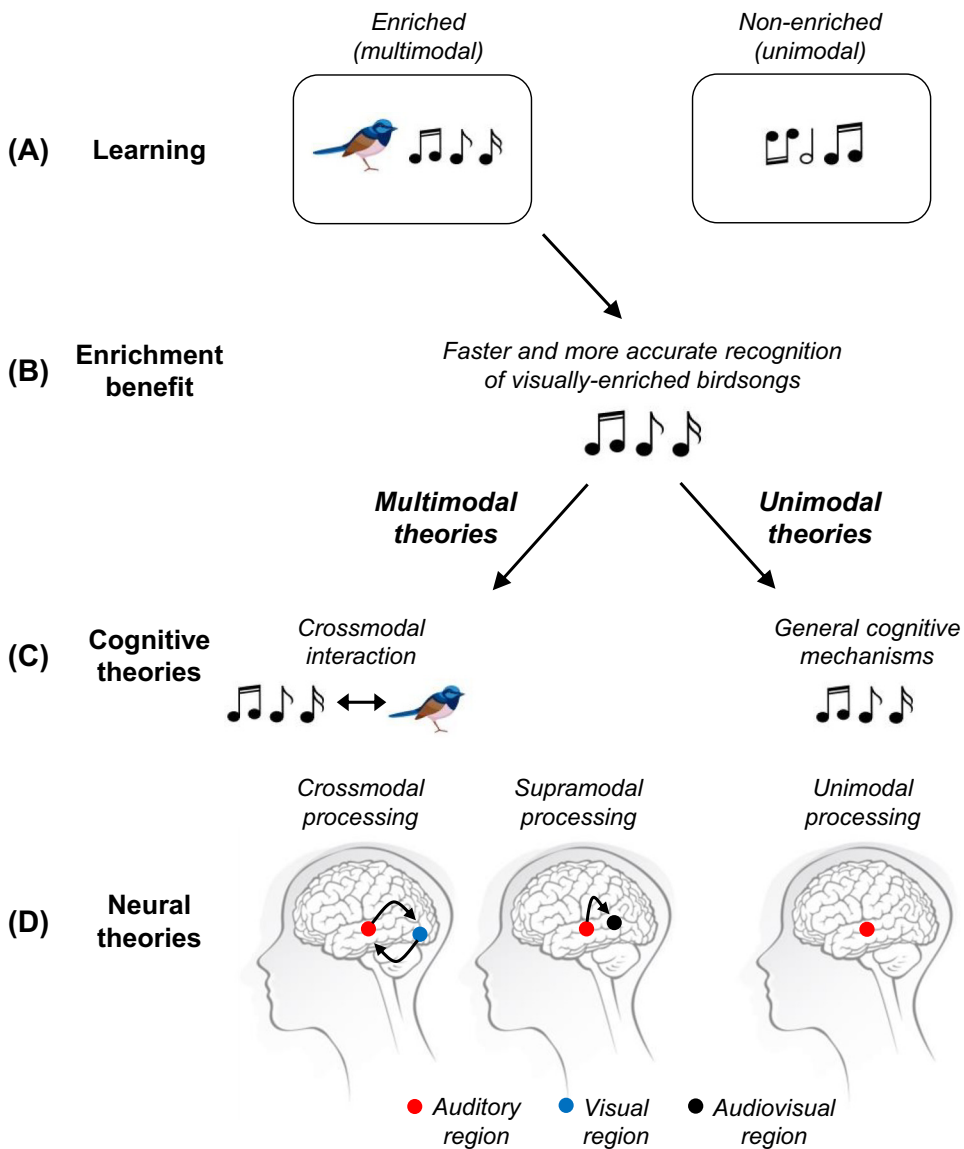Cognitive and neural mechanisms of enriched learning

Figure 2. (A) In this example, multisensory enrichment is used: birdsong is learned under two conditions. In an enriched learning condition, each song is paired with a picture of a specific bird species. In a non-enriched learning condition, each song is learned with no accompanying visual stimulus. (B) Following learning, enriched birdsongs are recognized faster and more accurately compared with non-enriched birdsongs (enrichment benefit). Multimodal and unimodal theories offer diverging theoretical accounts of enrichment benefits. (C) Cognitive theories. Multimodal theories propose that the recognition of a familiar birdsong can activate its corresponding picture, thereby facilitating the recognition of the birdsong. Unimodal theories propose that general cognitive mechanisms, such as greater attention or the processing of the meaning of an item rather than its surface-level features, during enriched learning, yield enhanced recognition of enriched

(Figure legend continued at the bottom of the next page.)

communicate directly with each other, supramodal accounts assume that higher-level convergence zones mediate connections between sensory-specific cortices.

Computational

Multimodal computational models of enriched learning rely heavily on supramodal processing mechanisms to explain enrichment benefits. One such model, the oscillatory neural network model [7], assumes that convergence zones mediate activity within sensory-specific brain regions. The model configures sensory-specific regions as not communicating directly, but instead transferring information via higher-level convergence regions based on sparse spatiotemporal encoding of multisensory input. Another model of multisensory perception [42] similarly predicts enrichment benefits based on interactions between multisensory and sensory-specific representations using a probabilistic Bayesian framework. Other computational models also suggest that modality-independent representations optimally account for enrichment benefits [43,44].

Unimodal theories
Cognitive

Another common explanation for enrichment benefits is that enriched stimuli are processed more deeply or with greater attention ('general cognitive mechanisms'; Figure 2C) compared with non-enriched stimuli during encoding, leading to more robust representations. For example, the levels-of-processing theory proposes that deeper, as opposed to shallower, processing boosts memory for a given stimulus [45]. Shallow processing may entail the encoding of stimulus appearance, while deeper processing may entail the encoding of stimulus meaning [46]. Levels-of-processing approaches have faced some criticism, because depth is a difficult concept to define, and deeper processing could also simply involve greater effort or attention [47]. Indeed, learners may pay more attention under enriched relative to non-enriched learning conditions because enriched learning may be more attention-grabbing than non-enriched learning [48]. However, a single attentional mechanism is often not able to explain patterns of enrichment benefits across different types of enriched learning in both typically developed participants [3] and participants with neurodevelopmental disorders who display deficits in enriched learning for some tasks but not others [29].

Neural

According to unimodal neuroscience theories, sensory and motor areas of the brain undergo changes during enriched learning that enhance their sensitivity during subsequent unisensory tasks. Critically, unlike multimodal theories, unimodal theories do not expect these unisensory tasks to trigger crossmodal brain responses. For the example of birdsong learning, this means that the auditory cortex, but not the visual cortex, responds with greater sensitivity to birdsongs after visually enriched learning ('unimodal processing'; Figure 2D). One unisensory model, the common framework for perceptual learning, proposes that enrichment boosts learning via factors such as increased attention or reinforcement, which enhance the sensitivity of unimodal brain responses following learning [33].

birdsongs. (D) Neural theories. Multimodal theories come in two possible flavors: crossmodal and supramodal. During crossmodal processing, recognition of an auditory-only presented birdsong after visually enriched learning triggers responses within auditory brain regions (red circle) and crossmodally in visual brain regions (blue circle), which interact. By contrast, during supramodal processing, recognizing the visually enriched birdsong triggers responses within auditory brain regions and audiovisual convergence zones (black circle). According to unimodal theories, enriched birdsong is processed solely by auditory brain regions even after enriched learning. Theoretically, crossmodal, supramodal, and unimodal mechanisms are not mutually exclusive and could operate in parallel.

Computational

To our knowledge, no existing computational models rely on unimodal mechanisms to explain multimodal enrichment benefits.

Neuroscience evidence for multimodal and unimodal theories

Although behavioral experiments have provided an abundance of new evidence in recent years for the effectiveness of enrichment, they cannot adjudicate between multimodal and unimodal theories that explain why enrichment is effective. However, neuroscience has recently made large strides in solidifying evidence for explaining enrichment benefits. The bulk of the evidence (Figure 2) provides support for multimodal models of enrichment and specifically for crossmodal mechanisms (Figure 3).

Uncovering this evidence has required the use of several converging neuroscience techniques. **Magnetic resonance imaging (MRI)** reveals which brain regions respond to enriched stimuli and whether they are functionally and structurally connected. **Neurostimulation** demonstrates the causal relevance of specific brain regions for behavioral enrichment benefits. **Electroencephalography (EEG)** and **magnetoencephalography (MEG)** characterize the timing of brain responses to enriched versus non-enriched stimuli.

Crossmodal mechanisms

Crossmodal neural theories of enrichment (Figure 2D) generate several unique predictions. First, items learned under enriched conditions should elicit crossmodal neural responses when those items are subsequently encountered under non-enriched conditions. Second, crossmodal responses should causally contribute to enrichment benefits. Third, connectivity between crossmodal and sensory-specific regions should increase for enrichment benefits. Fourth,

---

Box 3. Predictive coding frameworks of multimodal enrichment

Sensory brain regions demonstrate functional specialization in the processing of input arising from specific sensory modalities [119]. Why, then, might regions specialized in the processing of input from one modality (e.g., visual) participate in a task presented in another modality (e.g., auditory) following enriched learning, as some multimodal theories propose? Furthermore, why would such crossmodal brain responses benefit task performance? According to the MPCF [3,30], crossmodal brain responses constrain possible interpretations of incoming sensory information, which makes perception faster and more accurate. Crossmodal responses indicate that the brain uses multisensory internal generative models, which are encoded during enriched learning and deployed during subsequent unisensory perception (Figure I).

As an example, suppose that one encounters an unfamiliar person in a face-to-face meeting. During the meeting, one acquires an internal generative model for the unfamiliar person that includes, among other details, their face and voice characteristics. Some days later, the same person is encountered in a unisensory context, such as a phone conversation, where only voice information is available. The MPCF proposes that, during the phone conversation, the brain uses its previously encoded internal model to infer the absent visual face information, which effectively constrains the interpretation of the unimodal stimulus, thereby facilitating comprehension of the incoming speech and identification of the speaker [29]. This means that auditory perception is based not only on the incoming auditory signal and internally generated auditory predictions, but also on the face–voice correspondences encoded during the initial face-to-face encounter. Thus, inference about the missing visual features serves as an additional source of knowledge that can be used to construct a more accurate percept during the auditory-only encounter. The MPCF assumes that the same general mechanism occurs also for other types of enriched stimuli and enrichment material such as novel vocabulary paired with gestures [3].

Predictive coding is one instantiation of the Bayesian brain hypothesis [120,121], a widely accepted framework of brain function, which assumes that the most fundamental computational principle of the brain is to infer probable causes of incoming sensory information. Internal generative models and predictive coding mechanisms are also featured prominently in explanations of the binding problem [122]. Neurocomputational accounts, such as the MPCF, move beyond earlier, more qualitative explanations of enrichment benefits by providing frameworks for more detailed and testable hypotheses related to neurobiologically plausible computational mechanisms of enrichment benefits. For example, facilitatory neurostimulation of visual brain regions following visually enriched auditory learning should enhance post-learning auditory task performance [60].
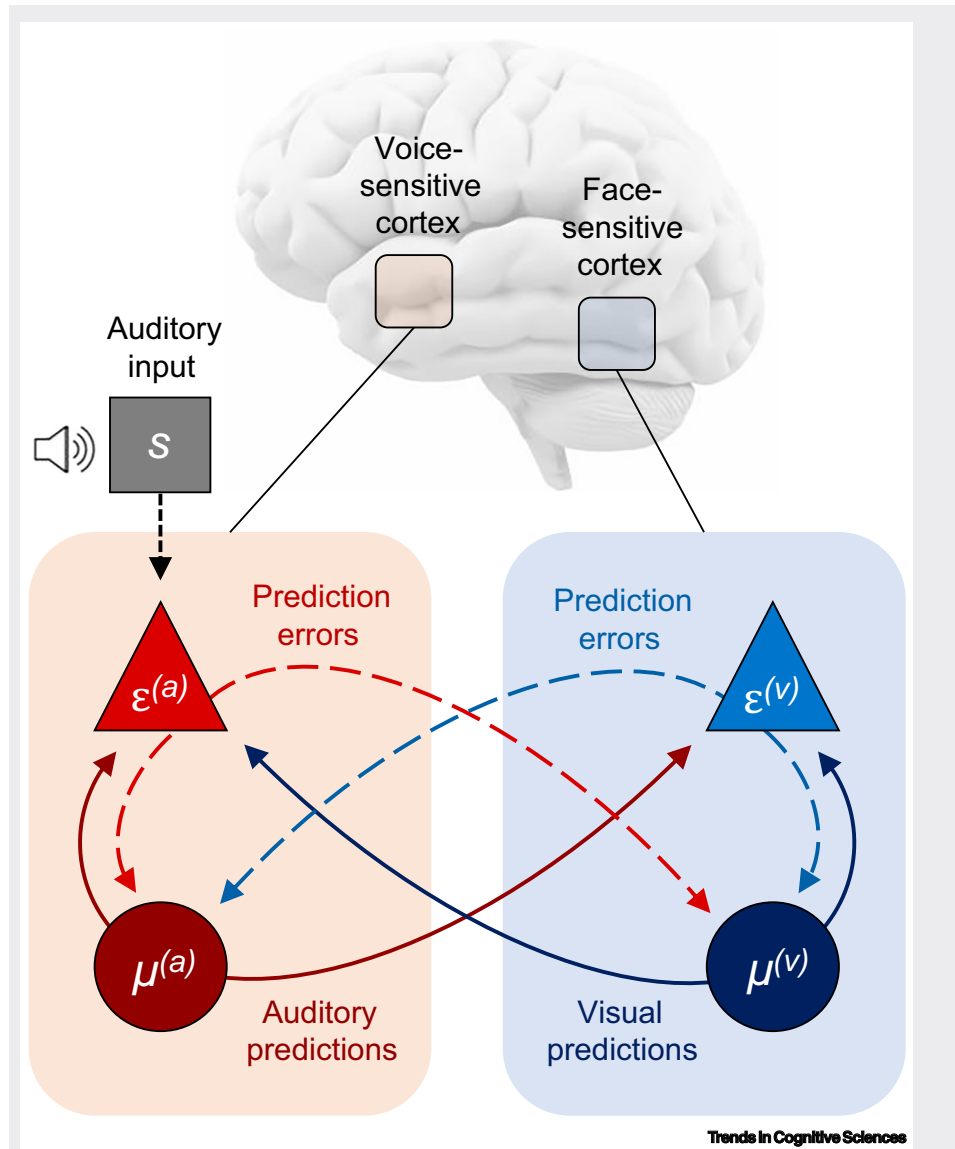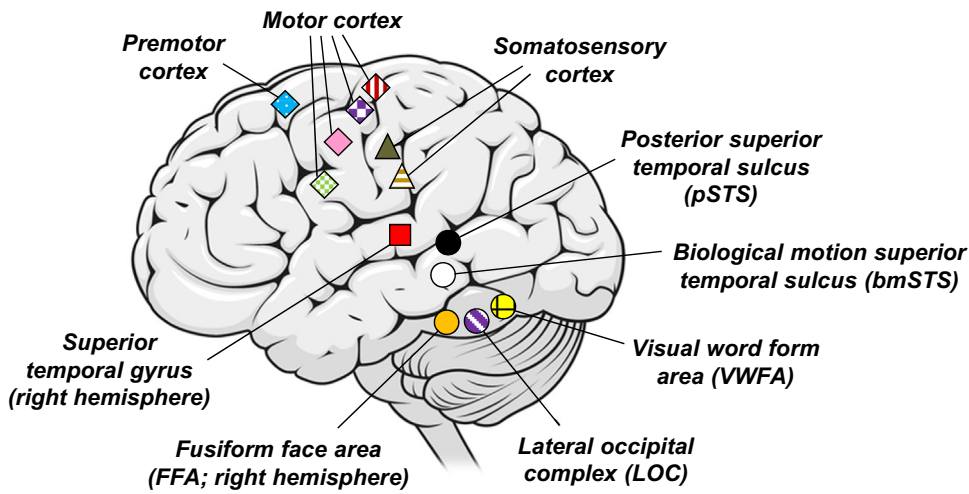

Figure I. Predictive coding of enrichment benefits in the example of voice-identity recognition. According to the multisensory predictive coding framework (MPCF), crossmodal responses following enriched learning are implemented using predictive coding [6,123,124]. Predictive coding involves the generation of **forward models** across multiple levels of processing in the brain, which are used to predict sensory input based on learned parameters. Forward models attempt to minimize the difference between predictions and sensory input, referred to as prediction error. In this simplified example, forward models [$\mu^{(a)}$] generated by voice-sensitive auditory brain regions (red [23,125]) are used to form predictions (unbroken arrows) about incoming auditory signals (s), in this case, a speaker's voice. These auditory predictions are compared with incoming auditory sensory information by error coding units [$\varepsilon^{(a)}$]. Any mismatch between predictions and sensory input generates a prediction error signal (dashed arrows). Thus, prediction error is computed as the difference between forward model predictions and incoming sensory input ($\varepsilon = s - \mu$) [124]. Auditory predictions and prediction errors are transmitted to the face-sensitive visual areas of the cortex, thereby engaging visual forward models [$\mu^{(v)}$]. In turn, visual error coding units [$\varepsilon^{(v)}$] transmit predictions and prediction errors back to voice-processing brain areas. This results in a decreased system-wide level of prediction error. Enrichment benefits can be attributed in this system to the additional information about the auditory signal provided by visual forward models. For simplification purposes, we do not display here inhibitory self-connections that modulate the precision of forward models and error coding units by up- or downregulating the weights of auditory and visual predictions and error signals.

Figure 3. Crossmodal brain responses following multisensory and sensorimotor enriched learning. A common finding across several recent neuroscience studies is that crossmodal visual, auditory, somatosensory, and motor brain regions show enhanced responses to stimuli learned under enriched learning conditions compared with non-enriched conditions. The schematic of the brain shown here specifies locations of maximal crossmodal responses based on neural coordinates reported in individual studies [3,4,29,52,54,55,58,59,61,62]. For example, in one study, participants learned auditorily-presented Morse code under two conditions [55]. In one condition, auditory Morse code sequences were presented simultaneously with perceptually congruent vibrotactile stimulation applied to the hand. In a second, non-enriched condition, participants learned Morse code sequences only by listening with no accompanying vibrotactile stimulation. Following learning, participants heard Morse code sequences that they had learned while MRI was used to measure their brain responses. A site within the somatosensory cortex (filled triangle), a brain region that processes tactile sensory information, responded more to auditorily presented sequences that had been paired with vibrotactile stimulation during learning than to sequences learned only by listening.

crossmodal responses should occur at a time point that precedes recognition to have an impact on the beneficial effects of enrichment. Current evidence has confirmed each of these predictions.

Crossmodal visual brain responses following multisensory enriched learning

Several studies report crossmodal responses in visual brain regions following audiovisual learning (Figure 3). The fusiform face area (FFA), a region involved in processing face identity, responds when participants recognize the identities of voices that are paired with corresponding faces during prior exposure [28,29]. Similarly, the posterior superior temporal sulcus (pSTS), a region associated with the visual perception of facial movements, responds when participants recognize speech produced by voices that have previously been paired with corresponding faces. Stimulation of the pSTS using **transcranial direct current stimulation (tDCS)** decreases the face benefit for speech recognition [49].

Similarly, patterns of responses within the lateral occipital complex (LOC), a visual brain region associated with the recognition of objects, can predict whether novel vocabulary has previously been learned only by listening or by listening while viewing related pictures [3].

Recent work also shows that visuotactile training of the Braille alphabet can lead to responses within the visual word form area (VWFA) of the brain during blindfolded tactile reading in sighted individuals, as well as to substantial gray and white matter reorganization in early visual cortices [50–52]. The application of **transcranial magnetic stimulation (TMS)** to the VWFA decreased Braille reading accuracy following learning, suggesting the causal relevance of the VWFA for tactile reading following visuotactile learning [52].

Crossmodal auditory and somatosensory brain responses following multisensory enriched learning

Crossmodal auditory and somatosensory responses (Figure 3) are less investigated than are crossmodal visual responses. Two studies found that viewing pictures that were paired during prior learning with congruent sounds evoked auditory neural responses [53,54]. In addition, the learning of Morse code presented simultaneously in auditory and tactile modalities yielded enhanced somatosensory cortex responses to auditory Morse code following learning [55]. Neurostimulation evidence for the causal relevance of somatosensory and auditory crossmodal responses has not yet been reported.

Crossmodal motor cortex responses following sensorimotor enriched learning

Crossmodal motor cortex responses to auditory and visual stimuli following sensorimotor-enriched learning are widely observed (Figure 3). The primary motor cortices respond more strongly to symbols that have previously been learned by self-writing as opposed to viewing another person write the same symbols [4,56,57]. The motor cortex responds more to audiovisual stimuli learned through self-generated actions than to stimuli learned by viewing the same actions generated by someone else [58].

Crossmodal motor cortex responses are also associated with gesture enrichment benefits. For example, solving mathematical equivalence problems while speaking and gesturing, rather than only speaking, leads to a subsequent increase in motor and somatosensory cortices during passive problem-solving [59]. Gesture-enriched and nongesture-enriched stimuli can be accurately dissociated based on patterns of motor cortex responses [3]. Inhibitory TMS of the motor cortices selectively disrupts the auditory translation of foreign language words learned by performing gestures [5], suggesting a causal role of crossmodal motor cortex responses in gesture enrichment benefits. The application of inhibitory TMS to a visual brain region known as the biological motion STS also reduced gesture enrichment benefits [60], suggesting a mediating role of both motor and sensory cortices.

Sensorimotor-enriched learning of music, relative to auditory-only learning, is also characterized by enhanced post-learning auditory–motor interactions in which motor cortex areas show increased responses to auditory stimuli [61–63].

Crossmodal functional connectivity following enriched learning

Enriched learning can change how sensory-specific brain regions work together following learning. For example, the face benefit for voice identity is associated with enhanced functional connectivity between the FFA and auditory voice-sensitive areas [23]. Conversely, the face benefit for speech is associated with enhanced functional connectivity between the pSTS and auditory language areas [64]. Learning letters by handwriting increases the functional connectivity between visual and motor cortices [65], as does the sensorimotor-enriched learning of audiovisual objects [58]. Training in tactile reading enhances functional connectivity between the VWFA and somatosensory regions [52], and music training is associated with enhanced functional connectivity between auditory and motor regions [66,67]. Several studies also report correlations of behavioral enrichment benefits with crossmodal functional connectivity and/or crossmodal responses [3,28,29], suggesting that crossmodal connectivity and responses are task relevant and not merely epiphenomenal.

Increases in functional connectivity may be supported by direct structural connections between sensory-specific brain regions, such as auditory and visual motion cortices [68], and regions related to face and voice processing, that is, the FFA and temporal voice-sensitive area [69,70].

Early timing of crossmodal brain responses following enriched learning

Crossmodal responses can only benefit behavior if they precede stimulus recognition. Several studies report crossmodal responses occurring early during the perception of enriched stimuli. FFA responses associated with the face benefit occur only 100 ms after voice onsets [71]. This early FFA response is followed by an increase in auditory cortex activity that peaks ~200 ms post-auditory onset and correlates with voice recognition performance, indicating that crossmodal responses are evoked during early stages of sensory processing. In addition, sounds that have been paired with pictures during prior learning elicit a distinct EEG time course as early as 35 ms post-auditory onset relative to sounds learned without pictures [72]. **Source localization** suggests that these differences are associated with increased responses in visual brain regions. Motor contributions to auditory processing following sensorimotor-enriched music learning occur within 200 ms of auditory onsets [62,63].

Supramodal mechanisms

The potential contributions of neural convergence zones to enrichment benefits are largely unknown, because convergence zone responses have almost exclusively been investigated in paradigms that investigate the **binding problem** without manipulating learning experience [73,74]. One MEG learning study found that responses within putative STS convergence regions correlated with behavioral performance in a visual task following audiovisual perceptual learning, but not following visual-only learning or visual learning paired with auditory noise [75]. Neuromodulation studies have not yet taken up the question of whether convergence zones contribute to enrichment benefits. One difficulty is to anatomically differentiate convergence zones in the STS from nearby pSTS visual movement regions [3].

Unimodal mechanisms

While it is well established that learning can alter sensory-specific cortices [76], only a few studies have reported greater alterations following enriched learning than following non-enriched learning. For example, children show greater VWFA responses to letters that they have previously

written by hand compared with letters that they have viewed another person write [56]. The viewing of tool-related actions elicits more pronounced alpha oscillatory desynchronization over the visual cortex if the actions were learned through self-performance rather than through viewing [77]. Similarly, brain responses to pictures that have been learned with accompanying sounds are associated with enhanced activity in the LOC occurring within 60 ms of picture onsets [78,79]. However, other studies report no altered responses of unisensory cortices after enriched learning (e.g., [23]).

Integrating cognitive, neuroscience, and computational theories of enrichment benefits

How do the neuroscience findings reviewed in the previous section inform cognitive and computational theories?

Convergence of cognitive, computational, and neuroscientific theories

Current cognitive, computational, and neural theories of enriched learning converge on some broader principles. First, the notion that enriched learning involves the formation of multimodal representations is expressed by several neuroscience and computational theories and by several cognitive theories, including the cognitive multimodal theories mentioned earlier (dual coding, CTML, motor trace theory, and grounded cognition theories). Second, these theories assume that learned information, such as visual and auditory bird features, is represented in multiple modalities that interact following learning. Third, these interactions are beneficial for remembering the learned material. The studies reviewed above provide evidence in support of all three of these principles.

Updating cognitive and computational approaches

Nevertheless, neuroscience findings from the past decade could be used to update some assumptions of cognitive theories. For instance, dual coding theory posits that nonverbal information, such as the shape and sound of a stimulus, is represented within the same system [34]. However, differing sensory-specific regions specialize in the processing of these two features. For example, birdsong is likely represented in auditory association cortices [80,81] and visual bird representations in the LOC [82]. In addition, findings that enrichment benefits can be supported by interactions between sensory and motor cortices (e.g., [59]) are difficult to explain when assuming that enrichment benefits rely on interactions between verbal and nonverbal codes. Thus, despite its multimodal framing, the translation of the assumptions of dual coding theory to the brain is not entirely clear, which is often the case when comparing cognitive and neuroscience frameworks [83].

The theoretical divide between neuroscientific and computational approaches appears smaller than the divide between neuroscientific and cognitive approaches. However, computational accounts rely on supramodal mechanisms, the contributions of which to learning have not yet been investigated systematically outside of low-level visual perceptual learning tasks [41]. To our knowledge, no computational model exists that implements enrichment benefits using crossmodal mechanisms, despite strong neuroscientific evidence for such a mechanism and a mathematical framework that could explain enrichment benefits (Box 3).

Neuroscience findings challenge unimodal accounts

Unimodal accounts in cognitive science and neuroscience share an emphasis on the role of general cognitive mechanisms, such as attention, that drive enriched learning beyond a threshold, below which neural activation is not sufficient [33] or processing is not deep enough [45] for the learning materials to be robustly encoded. Unimodal neuroscience theories find some support

in studies of audiovisual learning demonstrating enhanced visual cortex responses to visual stimuli following learning [56]. However, the wealth of neuroscience findings makes it clear that enrichment benefits within the variety of learning domains and paradigms reviewed here are not based on purely unimodal mechanisms. Whether unimodal and supramodal mechanisms operate in parallel with crossmodal mechanisms is, to date, unclear.

Interactions between research in cognitive neuroscience, computational models, and education

Connections between education, psychology, and neuroscience are actively explored and debated. Although there are some prominent critics (e.g., [84]), there is broad consensus that neuroscience can be used to constrain mechanisms featured in cognitive theories, which, in turn, suggest educational strategies [85]. Enhancing teachers' knowledge of neuroscience concepts can also directly impact how they aim to engage students' everyday learning [86].

One criticism of educational neuroscience research is that studies are so highly controlled that their findings cannot be applied to natural learning situations. The research reviewed here has taken steps to overcome this limitation by examining enriched learning in more naturalistic ways, such as relying on group learning situations to simulate classroom learning environments [3], and by investigating the learning of more realistic materials, such as foreign language words [87], written text [65], and melodies [63].

While neuroscience studies of multimodal enrichment are becoming more naturalistic, studies conducted by educational researchers are also becoming increasingly controlled. Recent research makes use of random assignment, within-participant designs, and non-enriched control conditions, permitting more robust conclusions about the effects of enrichment [16,88,89]. Therefore, neuroscientific and educational approaches to researching multisensory enrichment are beginning to converge.

We envision recent neuroscience findings as informing an interdisciplinary neuroscience–psychology–education approach to optimizing teaching strategies [86]. For example, further development of computational models in a neurobiologically plausible fashion could permit the simulation of learning outcomes and multimodal inputs for which crossmodal processing is optimized. Model predictions could then be tested behaviorally in the laboratory and, ultimately, in classroom environments. Greater engagement between neuroscience, education, and computational fields will lead to greater insights into how to best implement enriched learning practices.

Concluding remarks

Studying how humans can optimally learn from the wealth of sensory information that surrounds them as well as from their own movements will be key to developing pedagogical techniques that can transform classrooms into enriched learning environments. Sensory and motor brain regions respond crossmodally to unisensory input following enriched learning, and these crossmodal responses are instrumental in triggering enrichment benefits, in support of multimodal models of enriched learning. These recent findings contribute to our understanding of why several longstanding educational strategies, such as the Montessori Method, are effective. The findings may also help to update cognitive and computational theories of enrichment and provide new hypotheses regarding optimal learning strategies that might be tested in behavioral and educational experiments. Of course, many important questions remain (see Outstanding questions). We hope that this review inspires further exchange between investigators working in the fields of cognitive and computational science, neuroscience, education, and rehabilitation.

Outstanding questions

How robust are the plastic changes in crossmodal brain regions induced by enriched learning? In some studies, crossmodal responses persist for at least 6 months. Do crossmodal mechanisms support benefits of enriched learning that are maintained over even longer timescales?

Is there a role for supramodal brain regions in learning? If yes, what mechanisms determine the involvement of supramodal versus crossmodal responses in supporting enrichment benefits?

Can brain responses or patterns recorded during enriched learning be used to predict learning outcomes? This question has not yet been answered because few neuroscience studies of enriched learning have assessed brain responses during learning itself.

What factors contribute to individual differences in benefits achieved from enriched learning?

Which features of a stimulus or task determine the time at which benefits of enriched learning begin to appear?

How long would an arbitrary pairing of stimuli need to be trained to induce enrichment benefits? Does explicit enriched learning yield benefits more quickly than learning by passive exposure?

How do benefits of enriched learning in naturalistic contexts compare with those achieved through the use of virtual reality or mental imagery?

Trends in Cognitive Sciences

Resources

ᶦM. Macedonia, et al., Gesture video corpus, *figshare*, 2021, https://doi.org/10.6084/m9.figshare.c.5296396.v1

References

1. Marshall, C. (2017) Montessori education: a review of the evidence base. *NPJ Sci. Learn.* 2, 1–9
2. Mavilidi, M.F. *et al.* (2022) Embodiment as a pedagogical tool to enhance learning. In *The Body, Embodiment, and Education: An Interdisciplinary Approach* (Stolz, S.A., ed.), pp. 183–203, Routledge
3. Mayer, K.M. *et al.* (2015) Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* 25, 530–535
4. Kersey, A.J. and James, K.H. (2013) Brain activation patterns resulting from learning letter forms through active self-production and passive observation in young children. *Front. Psychol.* 4, 567
5. Mathias, B. *et al.* (2021) Motor cortex causally contributes to vocabulary translation following sensorimotor-enriched training. *J. Neurosci.* 41, 8618–8631
6. Keller, G.B. and Mrsic-Flogel, T.D. (2018) Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435
7. Rao, A.R. (2018) An oscillatory neural network model that demonstrates the benefits of multisensory learning. *Cogn. Neurodyn.* 12, 481–499
8. Greenberg, K. (2021) Individual differences in visuospatial working memory capacity influence the modality effect. *J. Comput. Assist. Learn.* 37, 735–744
9. Rau, P.L.P. *et al.* (2020) Distractive effect of multimodal information in multisensory learning. *Comput. Educ.* 144, 103699
10. Macrine, S.L. and Fugate, J.M., eds (2022) *Movement Matters: How Embodied Cognition Informs Teaching and Learning*, MIT Press
11. Shams, L. and Seitz, A.R. (2008) Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417
12. Barakat, B. *et al.* (2015) Visual rhythm perception improves through auditory but not visual training. *Curr. Biol.* 25, R60–R61
13. Johnson-Glenberg, M.C. and Megowan-Romanowicz, C. (2017) Embodied science and mixed reality: how gesture and motion capture affect physics education. *Cognit. Res. Principles Implications* 2, 1–28
14. Wakefield, E. *et al.* (2018) Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21, e12664
15. Liu, Y. *et al.* (2011) Learning a tonal language by attending to the tone: an in vivo experiment. *Lang. Learn.* 61, 1119–1141
16. Andrä, C. *et al.* (2020) Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educ. Psychol. Rev.* 32, 815–850
17. Cooper, A. and Wang, Y. (2012) The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.* 131, 4756–4769
18. Mathias, B. *et al.* (2022) Twelve- and fourteen-year-old school children differentially benefit from sensorimotor- and multisensory-enriched vocabulary training. *Educ. Psychol. Rev.* 34, 1739–1770
19. Baills, F. *et al.* (2019) Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Stud. Second. Lang. Acquis.* 41, 33–58
20. Zhen, A. *et al.* (2019) Manual directional gestures facilitate cross-modal perceptual learning. *Cognition* 187, 178–187
21. Lingnau, A. (2018) *Pantomimes, Emblems, and Meaningless Gestures*, Royal Holloway, University of London
22. García-Gámez, A.B. and Macizo, P. (2019) Learning nouns and verbs in a foreign language: The role of gestures. *Appl. Psycholinguist.* 40, 473–507
23. von Kriegstein, K. and Giraud, A.L. (2006) Implicit multisensory associations influence voice recognition. *PLoS Biol.* 4, e326
24. Ernst, M.O. (2007) Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 7
25. Shams, L. *et al.* (2011) Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2, 264
26. Seitz, A.R. *et al.* (2006) Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427
27. Maguinness, C. *et al.* (2021) Prior audio-visual learning facilitates auditory-only speech and voice-identity recognition in noisy listening conditions. *PsyArXiv* Published online February 1, 2021. http://dx.doi.org/10.31234/osf.io/gc4xa
28. Maguinness, C. and von Kriegstein, K. (2021) Visual mechanisms for voice-identity recognition flexibly adjust to auditory noise level. *Hum. Brain Mapp.* 42, 3963–3982
29. von Kriegstein, K. *et al.* (2008) Simulation of talking faces in the human brain improves auditory speech recognition. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6747–6752
30. von Kriegstein, K. (2012) A multisensory perspective on human auditory communication. In *The Neural Bases of Multisensory Processes* (Murray, M.M. and Wallace, M.T., eds), pp. 683–700, CRC Press/Taylor & Francis
31. Pahor, A. *et al.* (2021) Multisensory facilitation of working memory training. *J. Cognit. Enhanc.* 5, 386–395
32. Clark, J.M. and Paivio, A. (1991) Dual coding theory and education. *Educ. Psychol. Rev.* 3, 149–210
33. Seitz, A.R. and Dinse, H.R. (2007) A common framework for perceptual learning. *Curr. Opin. Neurobiol.* 17, 148–153
34. Sadoski, M. and Paivio, A. (2013) *Imagery and Text: A Dual Coding Theory of Reading and Writing*, Routledge
35. Mayer, R.E. (2005) Cognitive theory of multimedia learning. In *The Cambridge Handbook of Multimedia Learning* (Mayer, R.E., ed.), pp. 31–48, Cambridge University Press
36. Schnotz, W. (2014) Integrated model of text and picture comprehension. In *The Cambridge Handbook of Multimedia Learning* (Mayer, R.E., ed.), pp. 72–103, Cambridge University Press
37. Engelkamp, J. (2001) Action memory: a systems-oriented approach. In *Memory for Action: A Distinct Form of Episodic Memory?* (Zimmer, H.D. *et al.*, eds), pp. 49–96, Oxford University Press
38. Barsalou, L.W. (2020) Challenges and opportunities for grounding cognition. *J. Cogn.* 3, 31
39. Kiefer, M. and Trumpp, N.M. (2012) Embodiment theory and education: the foundations of cognition in perception and action. *Trends Neurosci. Educ.* 1, 15–20
40. Zhou, Y. and MacLeod, C.M. (2021) Production between and within: distinctiveness and the relative magnitude of the production effect. *Memory* 29, 168–179
41. Proulx, M.J. *et al.* (2014) Multisensory integration, sensory substitution and visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 1–2
42. Yildirim, I. and Jacobs, R.A. (2012) A rational analysis of the acquisition of multisensory representations. *Cogn. Sci.* 36, 305–332

43. Erdogan, G. *et al.* (2015) From sensory signals to modality-independent conceptual representations: a probabilistic language of thought approach. *PLoS Comput. Biol.* 11, e1004610

44. Jacobs, R.A. and Xu, C. (2019) Can multisensory training aid visual learning? A computational investigation. *J. Vis.* 19, 1

45. Craik, F.I. (2002) Levels of processing: past, present... and future? *Memory* 10, 305–318

46. Galli, G. (2014) What makes deeply encoded items memorable? Insights into the levels of processing framework from neuroimaging and neuromodulation. *Front. Psychiatry* 5, 61

47. Eysenck, M.W. (2014) Depth, elaboration, and distinctiveness. In *Levels of Processing in Human Memory* (Cermak, L.S. and Craik, F.I.M., eds), pp. 89–119, Psychology Press

48. Santangelo, V. and Spence, C. (2007) Multisensory cues capture spatial attention regardless of perceptual load. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1311–1321

49. Riedel, P. *et al.* (2015) Visual face-movement sensitive cortex is relevant for auditory-only speech recognition. *Cortex* 68, 86–99

50. Bola, Ł. *et al.* (2017) Task-specific reorganization of the auditory cortex in deaf humans. *Proc. Natl. Acad. Sci. U. S. A.* 114, E600–E609

51. Matuszewski, J. *et al.* (2021) Brain plasticity dynamics during tactile Braille learning in sighted subjects: Multi-contrast MRI approach. *NeuroImage* 227, 117613

52. Siuda-Krzywicka, K. *et al.* (2016) Massive cortical reorganization in sighted Braille readers. *eLife* 5, e10762

53. Nyberg, L. *et al.* (2000) Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11120–11124

54. Thelen, A. *et al.* (2012) Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *NeuroImage* 62, 1478–1488

55. Junker, F.B. *et al.* (2021) Impact of multisensory learning on perceptual and lexical processing of unisensory Morse code. *Brain Res.* 1755, 147259

56. James, K.H. (2010) Sensori-motor experience leads to changes in visual processing in the developing brain. *Dev. Sci.* 13, 279–288

57. James, K.H. and Atwood, T.P. (2009) The role of sensorimotor learning in the perception of letter-like forms: tracking the causes of neural specialization for letters. *Cognit. Neuropsychol.* 26, 91–110

58. Butler, A.J. and James, K.H. (2013) Active learning of novel sound-producing objects: motor reactivation and enhancement of visuo-motor connectivity. *J. Cogn. Neurosci.* 25, 203–218

59. Wakefield, E.M. *et al.* (2019) Learning math by hand: the neural effects of gesture-based instruction in 8-year-old children. *Atten. Percept. Psychophys.* 81, 2343–2353

60. Mathias, B. *et al.* (2021) Visual sensory cortices causally contribute to auditory word recognition following sensorimotor-enriched vocabulary training. *Cereb. Cortex* 31, 513–528

61. Fujioka, T. and Ross, B. (2017) Beta-band oscillations during passive listening to metronome sounds reflect improved timing representation after short-term musical training in healthy older adults. *Eur. J. Neurosci.* 46, 2339–2354

62. Mathias, B. *et al.* (2015) Sensorimotor learning enhances expectations during auditory perception. *Cereb. Cortex* 25, 2238–2254

63. Mathias, B. *et al.* (2016) Sensory, cognitive, and sensorimotor learning effects in recognition memory for music. *J. Cogn. Neurosci.* 28, 1111–1126

64. Schall, S. and von Kriegstein, K. (2014) Functional connectivity between face-movement and speech-intelligibility areas during auditory-only speech perception. *PLoS One* 9, e86325

65. Vinci-Booher, S. *et al.* (2021) Visual-motor contingency during symbol production contributes to short-term changes in the functional connectivity during symbol perception and long-term gains in symbol recognition. *NeuroImage* 227, 117554

66. Lee, H. and Noppeney, U. (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U. S. A.* 108, E1441–E1450

67. Palomar-García, M.Á. *et al.* (2017) Modulation of functional connectivity in auditory–motor networks in musicians compared with nonmusicians. *Cereb. Cortex* 27, 2768–2778

68. Gurtubay-Antolin, A. *et al.* (2021) Direct structural connections between auditory and visual motion-selective regions in humans. *J. Neurosci.* 41, 2393–2405

69. Benetti, S. *et al.* (2018) White matter connectivity between occipital and temporal regions involved in face and voice processing in hearing and early deaf individuals. *NeuroImage* 179, 263–274

70. Blank, H. *et al.* (2011) Direct structural connections between voice-and face-recognition areas. *J. Neurosci.* 31, 12906–12915

71. Schall, S. *et al.* (2013) Early auditory sensory processing of voices is facilitated by visual mechanisms. *NeuroImage* 77, 237–245

72. Thelen, A. *et al.* (2015) Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition* 138, 148–160

73. Alsius, A. *et al.* (2018) Forty years after hearing lips and seeing voices: the McGurk effect revisited. *Multisens. Res.* 31, 111–144

74. Spence, C. *et al.* (2009) Crossmodal processing. *Exp. Brain Res.* 198, 107–111

75. Zilber, N. *et al.* (2014) Supramodal processing optimizes visual perceptual learning and plasticity. *NeuroImage* 93, 32–46

76. LeMessurier, A.M. and Feldman, D.E. (2018) Plasticity of population coding in primary sensory cortex. *Curr. Opin. Neurobiol.* 53, 50–56

77. Bryant, L.J. and Cuevas, K. (2019) Effects of active and observational experience on EEG activity during early childhood. *Psychophysiology* 56, e13360

78. Murray, M.M. (2005) The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage* 27, 473–478

79. Murray, M.M. *et al.* (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage* 21, 125–135

80. Muhammed, L. *et al.* (2018) Agnosia for bird calls. *Neuropsychologia* 113, 61–67

81. He, C. *et al.* (2020) Roles of category, shape, and spatial frequency in shaping animal and tool selectivity in the occipitotemporal cortex. *J. Neurosci.* 40, 5644–5657

82. Prather, J.F. *et al.* (2017) Brains for birds and babies: neural parallels between birdsong and speech acquisition. *Neurosci. Biobehav. Rev.* 81, 225–237

83. Price, C.J. and Friston, K.J. (2005) Functional ontologies for cognition: the systematic definition of structure and function. *Cognit. Neuropsychol.* 22, 262–275

84. Bowers, J.S. (2016) The practical and principled problems with educational neuroscience. *Psychol. Rev.* 123, 600–612

85. Thomas, M.S. *et al.* (2019) Annual research review: Educational neuroscience: progress and prospects. *J. Child Psychol. Psychiatry* 60, 477–492

86. Dubinsky, J.M. *et al.* (2019) Contributions of neuroscience knowledge to teachers and their practice. *Neuroscientist* 25, 394–407

87. Macedonia, M. and Mueller, K. (2016) Exploring the neural representation of novel words learned through enactment in a word recognition task. *Front. Psychol.* 7, 953

88. Mavilidi, M.F. *et al.* (2015) Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educ. Psychol. Rev.* 27, 413–426

89. Schmidt, M. *et al.* (2019) Embodied learning in the classroom: effects on primary school children's attention and foreign language vocabulary learning. *Psychol. Sport Exerc.* 43, 45–54

90. Zweig, L.J. *et al.* (2013) Arbitrary sounds facilitate visual search for congruent objects. *J. Vis.* 13, 1083

91. Longcamp, M. *et al.* (2008) Learning through hand-or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *J. Cogn. Neurosci.* 20, 802–815

92. Zemlock, D. *et al.* (2018) Visual–motor symbol production facilitates letter recognition in young children. *Read. Writ.* 31, 1255–1271

93. Hu, F.T. *et al.* (2015) Getting the point: tracing worked examples enhances learning. *Learn. Instr.* 35, 85–93

94. Bower, C. *et al.* (2020) Piecing together the role of a spatial assembly intervention in preschoolers' spatial and mathematics learning: Influences of gesture, spatial language, and socioeconomic status. *Dev. Psychol.* 56, 686

95. Schomaker, J. and Wittmann, B.C. (2021) Effects of active exploration on novelty-related declarative memory enhancement. *Neurobiol. Learn. Mem.* 179, 107403

96. MacLeod, C.M. and Bodner, G.E. (2017) The production effect in memory. *Curr. Dir. Psychol. Sci.* 26, 390–395

97. Scheffel, D.L. *et al.* (2008) The efficacy of a supplementary multisensory reading program for first-grade students. *J. Read. Improv.* 45, 139–152

98. Warnick, K. and Caldarella, P. (2016) Using multisensory phonics to foster reading skills of adolescent delinquents. *Read. Writ. Q.* 32, 317–335

99. Hadley, E.B. *et al.* (2016) Examining the acquisition of vocabulary knowledge depth among preschool students. *Read. Res. Q.* 51, 181–198

100. Mayer, R.E. (2020) *Multimedia Learning*, Cambridge University Press

101. Levine, S.C. *et al.* (2018) Mental transformation skill in young children: the role of concrete and abstract motor training. *Cogn. Sci.* 42, 1207–1228

102. Rueckert, L. *et al.* (2017) Gesture enhances learning of a complex statistical concept. *Cognit. Res. Principles Implic.* 2, 1–6

103. Thelen, A. and Murray, M.M. (2013) The efficacy of single-trial multisensory memories. *Multisens. Res.* 26, 483–502

104. Cook, S. and Wilding, J. (2001) Earwitness testimony: effects of exposure and attention on the Face Overshadowing Effect. *Br. J. Psychol.* 92, 617–629

105. Macedonia, M. and Knösche, T.R. (2011) Body in mind: how gestures empower foreign language learning. *Mind Brain Educ.* 5, 196–211

106. Hale, J.B. *et al.* (2016) Reconciling individual differences with collective needs: the juxtaposition of sociopolitical and neuroscience perspectives on remediation and compensation of student skill deficits. *Trends Neurosci. Educ.* 5, 41–51

107. Snowling, M.J. *et al.* (2020) Defining and understanding dyslexia: past, present and future. *Oxf. Rev. Educ.* 46, 501–513

108. Duff, D. *et al.* (2016) Using curriculum-based measurement data to monitor the effectiveness of the Wilson Reading System for students with disabilities: an exploratory study. *Int. J. Disability Hum. Develop.* 15, 93–100

109. Stebbins, M.S. *et al.* (2012) Monitoring the effectiveness of the Wilson reading system for students with disabilities: one district's example. *Exceptionality* 20, 58–70

110. Jacoby, N. and Ahissar, M. (2013) What does it take to show that a cognitive training procedure is useful? A critical evaluation. *Prog. Brain Res.* 207, 121–140

111. Stevens, E.A. *et al.* (2021) Current state of the evidence: examining the effects of Orton-Gillingham reading interventions for students with or at risk for word-level reading disabilities. *Except. Child.* 87, 397–417

112. Postman-Caucheteux, W.A. *et al.* (2010) Single-trial fMRI shows contralesional activity linked to overt naming errors in chronic aphasic patients. *J. Cogn. Neurosci.* 22, 1299–1318

113. Meteyard, L. and Bose, A. (2018) What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *J. Speech Lang. Hear. Res.* 61, 658–674

114. Durand, E. *et al.* (2018) The neural and behavioral correlates of anomia recovery following personalized observation, execution, and mental imagery therapy: a proof of concept. *Neural Plast.* 2018, 5943759

115. Boo, M. and Rose, M.L. (2011) The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology* 25, 154–175

116. Clough, S. and Duff, M.C. (2020) The role of gesture in communication and cognition: Implications for understanding and treating neurogenic communication disorders. *Front. Hum. Neurosci.* 14, 323

117. Kroenke, K.M. *et al.* (2013) Lexical learning in mild aphasia: gesture benefit depends on patholinguistic profile and lesion pattern. *Cortex* 49, 2637–2649

118. Kroenke, K.M. *et al.* (2013) Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex* 49, 2553–2568

119. Houdé, O. *et al.* (2010) Mapping numerical processing, reading, and executive functions in the developing brain: an fMRI meta-analysis of 52 studies including 842 children. *Dev. Sci.* 13, 876–885

120. Aitchison, L. and Lengyel, M. (2017) With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* 46, 219–227

121. Friston, K. (2012) The history of the future of the Bayesian brain. *NeuroImage* 62, 1230–1233

122. Noppeney, U. and Lee, H.L. (2018) Causal inference and temporal predictions in audiovisual perception of speech and music. *Ann. N. Y. Acad. Sci.* 1423, 102–116

123. Friston, K. (2018) Does predictive coding have a future? *Nat. Neurosci.* 21, 1019–1021

124. Friston, K.J. (2019) Waves of prediction. *PLoS Biol.* 17, e3000426

125. Pernet, C.R. *et al.* (2015) The human voice areas: spatial organization and inter-individual variability in temporal and extratemporal cortices. *Neuroimage* 119, 164–174