The effect of stimulus complexity and verbalizability on associative learning and related memory processes

Kálmán Tót

PhD Thesis

Szeged

2023

University of Szeged
Albert Szent-Györgyi Medical School
Doctoral School of Theoretical Medicine

The effect of stimulus complexity and verbalizability on associative learning and related memory processes

PhD Thesis

Kálmán Tót, MS

Supervisors:
Dr. habil. Attila Nagy, MS, PhD
Dr. Gabriella Eördegh, MS, PhD

Szeged
2023

Table of contents

List of publications providing the basis of the thesis

I. Eördegh, G., Tót, K., Kelemen, A., Kiss, Á., Bodosi, B., Hegedűs, A., Lazsádi, A., Hertelendy, Á., Kéri, S., & Nagy, A. (2022). The Influence of Stimulus Complexity on the Effectiveness of Visual Associative Learning. *Neuroscience*, *487*, 26–34. https://doi.org/10.1016/j.neuroscience.2022.01.022
SJR indicator: Q2
IF: 3.3

II. Tót, K., Eördegh, G., Kiss, Á., Kelemen, A., Braunitzer, G., Kéri, S., Bodosi, B., & Nagy, A. (2022). Visual consequent stimulus complexity affects performance in audiovisual associative learning. *Scientific reports*, *12*(1), 17793. https://doi.org/10.1038/s41598-022-22880-z
SJR indicator: D1
IF: 4.6

Other publications:

I. Eördegh, G., Tót, K., Kiss, Á., Kéri, S., Braunitzer, G., & Nagy, A. (2022). Multisensory stimuli enhance the effectiveness of equivalence learning in healthy children and adolescents. PloS one, 17(7), e0271513. https://doi.org/10.1371/journal.pone.0271513

II. Rosu, A., Tót, K., Godó, G., Kéri, S., Nagy, A., & Eördegh, G. (2022). Visually guided equivalence learning in borderline personality disorder. Heliyon, 8(10), e10823. https://doi.org/10.1016/j.heliyon.2022.e10823

List of abbreviations

MTL: medial temporal lobe

COVIS: competition between verbal and implicit systems

AEL: acquired equivalence learning

RAET: Rutgers acquired equivalence test

SN: substantia nigra

VT: ventral tegmentum

NAT: number of acquisition trials

ALER: acquisition learning error ratio

RER: retrieval error ratio

GER: generalization error ratio

RT: response times

1. Introduction

1.1. Learning and memory

Learning and memory are basic functions of the nervous system in the animal kingdom. It enables an animal to adapt better to the environment, therefore it is important for survival, and gives an evolutionary advantage. The learning is a process of acquiring information about the environment, which results a change in behavior and neural connections, while memory is the process of storing and retrieving these information (Kandel et al., 2014; McGann, 2015). These two functions cannot be fully separated at a functional level. There are different types of memory based on the time length and the nature of the given information, and they are mediated by different brain regions (Kandel et al., 2014; Smith et al., 2003).

According to the multi-store model by Atkinson and Shiffrin (1968), based on the time length, memory can be divided into sensory store, short-term and long-term memory. Although, this model was challenged, as it has now become evident that both short-term and long-term memory is more complicated than previously thought, and they include several components (Baddeley & Hitch, 1974).

The sensory memory stores all the incoming information from the sensory systems, but just for a fraction of a second. This type of memory can be considered as part of the perception (Baddeley et al., 2009). After the information catches our attention it will be transferred to the short-term memory or otherwise known as working memory, which is mediated by the prefrontal and orbitofrontal cortices. This memory storage stores conscious goal-oriented information, which we use for different kind of cognitive operations. It can be considered as a cognitive workspace. However, this is a transient storage and its capacity is limited. After approximately 20-30 seconds it fades away, or through repetitions can be get stored in the long-term memory. The long-term memory is an unlimited storage to our current knowledge. According to the nature of the given information, it can be separated into explicit (declarative) and implicit (non-declarative) memory (Figure 1.; (Hannula et al., 2023; Kandel et al., 2014; Squire & Dede, 2015; Tulving, 1983).

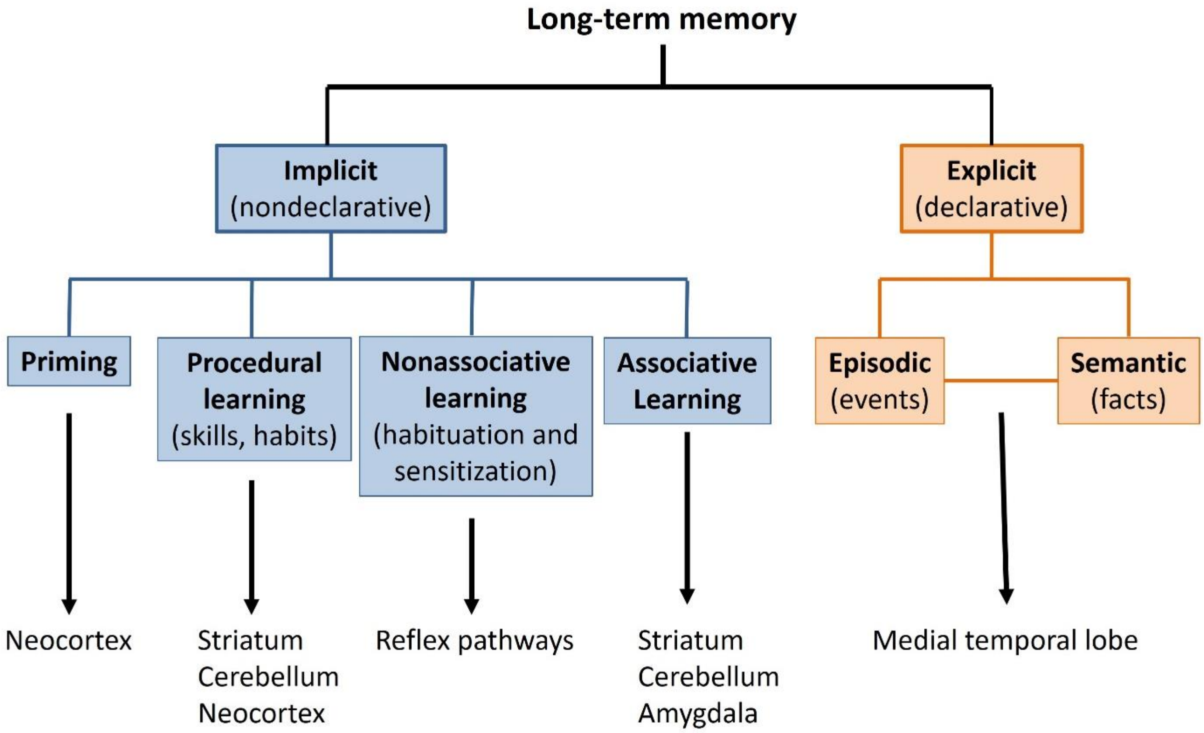

Figure 1. The organization of long-term memory systems and the related brain structures. (Adapted from Kandel et al., 2014)

1.1.1. Explicit learning

The explicit or declarative memory refers to the storage and retrieval of information which can be expressed in words, is conscious and accessible to the awareness (Hannula et al., 2023; Kandel et al., 2014; Squire & Dede, 2015). It provides a way to represent the world around us. These representations are highly flexible, unlike the implicit memory which is tightly connected to the original learning conditions (Squire & Dede, 2015).

The explicit learning depends on the medial temporal lobe (MTL) which includes the hippocampus, perirhinal, entorhinal and parahippocampal cortices (Lech & Suchan, 2013; Squire et al., 2004). This interacts with other cortical areas, from which the most important are the prefrontal cortex and the parietal association cortex (Batterink et al., 2019; Kandel et al., 2014; Shimamura et al., 1991; Tulving & Markowitsch, 1998). Besides the cortical areas, the diencephalic midline, which includes the mammillary nuclei, thalamic nuclei and the mammillothalamic tract, also plays a critical role, since the damage of these structures has the same effect as the damage in the MTL, probably due to the connections between them (Harding et al., 2000; Markowitsch, 1988; Squire & Wixted, 2011).

The explicit memory can be further separated into episodic and semantic memory. The episodic memory is the recollections of our personal experiences or time-and-place-specific events, it can be also called as the autobiographical memory, whereas, the semantic memory contains facts and concepts about the world, and it is not typically associated with the context in which the informations was acquired (Kandel et al., 2014; Squire & Dede, 2015; Tulving, 1983).

The formation of the explicit memory includes four stages or operations: the encoding, storage, consolidation and retrieval. Although, the MTL plays a critical role in explicit memory formation, the memory traces (engram) can be stored in the associative cortices based on different aspects (Kandel et al., 2014).

1.1.2. Implicit learning

Implicit or non-declarative learning and memory is an umbrella term for different kind of learnings which are not dependent on the MTL system (Squire & Dede, 2015), and are expressed through performance without requiring conscious awareness of the memory content, such as priming, procedural learning, associative and non-associative learning. Additionally, compared to the explicit learning, implicit learning happens slowly and gradually (Hannula et al., 2023; Squire & Dede, 2015). These types of learning contrary to the explicit memory are intact in amnesic patients with selective MTL lesion (Kandel et al., 2014; Reber, 2013; Squire & Dede, 2015).

1.1.2.1. Priming

Priming refers to a type of learning when a previously encountered stimulus is re-encountered. It is processed usually quickly and shows also a tendency to pop into mind (Hannula et al., 2023; Kandel et al., 2014; Reber, 2013; Squire & Dede, 2015). There are two types of priming: conceptual priming, which is related to semantic knowledge, and perceptual priming, which occurs within a specific sensory modality (Kandel et al., 2014). In a priming task, probably both types are activated; therefore, there is no sharp distinction between them

(Lee et al., 2020). Neuroimaging studies found that priming is followed by decreased brain activity for the re-encountered stimuli. This reduced activity is simultaneous in the perceptual and prefrontal cortices, and a greater neural synchrony can be observed between them (Ghuman et al., 2008; Reber, 2013; Schacter & Buckner, 1998).

1.1.2.2. Procedural learning

Procedural learning includes the learning of habits and motor, perceptual and cognitive skills. These types of learning are acquired incrementally through practice, which improves performance in a rapid and observable way. The learning usually begins with an explicit phase, where attention is needed, but later it becomes autonomous, where the skill can be executed unconsciously. In motor skill learning besides the neocortex, the basal ganglia and the cerebellum also play a crucial role. Disfunctions of these sturctures cause impaired motor learning (Kandel et al., 2014). Practice leads to enhanced and more broadly distributed activity in motor learning related structures, such as motor cortices, cerebellum and striatum (Reber, 2013; Ungerleider et al., 2002). Furthermore, skilled behavior can elicit morphological changes in the motor cortex, like the expansion of the cortical representation of the fingers in musicians (Elbert et al., 1995).

Habits are characterized by automatized and repetitive behavior. Habit learning in humans have been investigated using the Weather-prediction test. The subjects have to predict the weather (sun or rain) based on four possible cue cards, where each cue has a probabilistic relation to each outcome. In this kind of paradigm, as in habit learning, the basal ganglia play a key role, and patients with basal ganglia dysfunction have a poorer performance in the test (Kandel et al., 2014; Knowlton et al., 1996; Squire & Dede, 2015; Yin & Knowlton, 2006).

1.1.2.3. Non-associative learning

Non-associative learning is a phylogenetically early form of behavioral responses to a stimulus. It results when the subject is exposed to a single type of stimulus. It does not require the association of stimuli, but the properties of the single stimulus will be learned. This type of learning can be a habituation, which is a decreased response to a repeated neutral stimulus, or a sensitization, which is an enhanced response after an intense or painful stimulus (Kandel et al., 2014; Squire & Dede, 2015).

1.1.2.4. Associative learning

Associative learning is the learning of a relationship between two stiumuli (classical conditioning) or between a behavior and its consequence (operant conditioning). The forms of associative learning enable the animals to discriminate the co-occuring events from those that are random. In both types of learning the timing is crucial. The conditioned stimulus or reinforcement must closely follow the unconditioned stimulus or the operant action, respectively (Kandel et al., 2014).

In classical conditioning when a conditioned stimulus is followed by an unconditioned stimulus, the two will be associated with each other, and the conditioned stimulus will elicit an unconditioned response, as can be seen in Pavlov's famous experiment (Mallea et al., 2019).

In operant conditioning the behavior is followed by a positive or negative reinforcement. This type of learning can be also called as trial-and-error learning. Behaviors that are rewarded tend to be repeated, whereas the behavior which is punished or not rewarded are not (Kandel et al., 2014). In reinforcement learning the dopaminergic pathways of the basal ganglia play a central role (Delgado et al., 2000; Schultz, 1997; Shohamy et al., 2004; Squire & Dede, 2015). In both types of conditioning, besides the basal ganglia, the cerebellum and the amygdala are also involved (Kandel et al., 2014).

Other types of the associative learning include implicit category learning (Ashby & Maddox, 2005), statistical learning (Batterink et al., 2019) and last but not least acquired equivalence learning (Myers et al., 2003), which I would highlight as the main topic of this thesis.

1.1.3. The relationship between memory systems

The distinction of the explicit and implicit memory systems is based on MTL dependency and awareness, as described earlier. It was believed that the two systems are working independently and parallel to each other, but this has been challenged, due to the fact, that no general implicit learning system can provide a double dissociation from the explicit system in humans. This means that there is not known selective neural dysfunction, which impairs all the implicit learning forms, but leaves the explicit memory intact (Reber, 2013). Some studies suggest that anatomical dissociation cannot divide the two memory systems clearly, because there are some cases were the implicit learning is also impaired in MTL damage (Chun & Phelps, 1999; Hannula et al., 2023).

Healthy participants in many implicit tasks acquire simultaneously both explicit and implicit memory (Batterink et al., 2019; Reber, 2013; Squire & Dede, 2015). For example, this can be seen in the Serial Reaction Time (SRT) test, in which the participants have to learn a series of motor responses to repeatedly presented visual cue. Patients with anterograde amnesia can learn this test normally, without awareness of the sequence (Nissen, 1987; Reber & Squire, 1994, 1998). However, healthy participants with intact MTL could learn the test without awareness to some degree as well, but they always acquire some explicit knowledge about the sequence too (Shanks & Johnstone, 1998).

According to the COVIS (Competition between Verbal and Implicit Systems) model (Ashby et al., 1998), both memory systems can support the feedback based category learning, in which the application of conscious rules are mainly related to the MTL system and prefrontal cortex, and the implicit learning, where the stimuli are learned gradually is related to the cortico-striatal loops of the brain. It is hypothesized in this model that the basal ganglia play role in both the implicit and explicit category learning, due to the dopaminergic projections, since dopamine is released by positive feedbacks (Schultz, 2002). This can support the learning by strengthening the neural connections (Kerr & Wickens, 2001). The activity of the basal ganglia during category learning was shown in several studies (Nomura et al., 2007; Seger & Cincotta, 2005, 2006).

It was also shown in probabilistic learning that the activation of the two memory systems and the striatum typically contributes to feedback-based learning. The fMRI revealed activity both in the MTL and striatum, but as the learning progressed, the activity decreased in the former, and increased in the latter (Poldrack & Gabrieli, 2001). The tendency to shift to an implicit strategy was also shown in a virtual navigation task (Iaria et al., 2003).

In some cases, the memory systems can work cooperatively, but in other cases competitively (Knowlton et al., 1996; Schwabe, 2013; Squire & Dede, 2015). Furthermore, some tasks favor one memory system over the another and the prefrontal cortex may play an important role in determining which system gains control over the learning (McDonald & Hong, 2013). Engaging in the less optimal one the performance can be decreases, as seen in skill learning, when trying to use memorization strategy can disrupt the performance (Squire & Dede, 2015).

1.2. Acquired equivalence learning

Acquired equivalence learning (AEL) is a specific kind of associative learning, where two or more independent stimuli are linked together, and they are considered as equivalent with each other if they share the same outcome, and this equivalence can be transferred (transfer or generalization) to new instances (Hall, 1996; Myers et al., 2003).

1.2.1. Animal Studies

AEL is widely investigated in animal research, especially with rats and pigeons, it has been demonstrated by classical conditioning procedures (Hall, 1996; Urcuioli, 2001).

In the experiment conducted by Honey and Hall (1989) rats heard two auditory stimuli, stimuli A and B, which were followed by a food as a reward. On the other hand, the third stimulus, stimulus C was not rewarded. After this, in the second phase Stimulus A was paired with shock, and a conditional response was established and this also transferred (generalization) to stimulus B, but not to stimulus C. It was concluded, that between stimulus A and B an equivalence was formed, due to the common consequence. The experiment was later also repeated by Ward-Robinson and Hall (1999).

Hall and colleagues (1993) carried out an identical experiment to that by Honey and Hall (1989), except the food occurred before the auditory stimuli, and they could observe the same transfer of the conditional response to the other stimuli. This study showed that equivalence can be established not just when two stimuli share a common consequence, but when they share a common antecedent as well. This was supported by another study (Bonardi et al., 1993) which used pigeons, and two pair of stimuli: A-B reinforced by food, and C-D not reinforced. The equivalence was established between the pairs with common training history.

Furthermore, Molet and colleagues (2012) performed an experiment with rats, where cues that shared a common context were also treated equally, which is an evidence of context-mediated equivalence learning in animals. This was also reported in human conditional discrimination (Molet et al., 2011).

Acuired equivalence and generalization can develop between multiple items or stimuli within a category. New outcomes or responses associated to some members of the group should transfer without additional training to other members of the same group (Jitsumori, 2004; Urcuioli, 2001; Zentall et al., 2014). For example, in a study (Vaughan, 1988) pigeons were trained to discriminate slides of trees, which did not share perceptual similarities or dissimilarities. One group of the slides were reinforced with food, but the other group not. After this, the two group were then reversed regarded to reinforcement. The reversal was repeated severel times. After the training, pigeons could generalize the reversals to all stimuli of the group, even if they encountered reversal with just a few stimuli.

Furthermore, it is an interesting founding that equivalence can occur, not just between stimuli within the same group, but even between supraordinate categories as well, as it was shown in an experiment with pigeons (Astley, 1998; Astley & Wasserman, 1999).

1.2.2. Human studies

The forming of acquired equivalence between stimuli, which share a common associate or response can be found in humans too (Goyos, 2000; Meeter et al., 2009; Myers et al., 2003). To investigate this type of learning Myers and colleagues (2003) developed a computer based neurocognitive test, the Rutgers Acquired Equivalence Test (RAET or Face-Fish Test based on the stimuli applied). The task of the test is to learn visual stimulus pairs through trial-and-error learning. The test consists two main phases: the acquisition and the test phase. In the acquisition phase the participant learns the associations (between drawn faces and fishes) gradually, based on the feedback given by the software about the correctness of the answers. Additionally, equivalence will be established between stimuli which share the same consequence. In the test phase, no further feedback is given, and the already learned associations must be recalled (retrieval). Furthermore, previously not shown, but based on the equivalence, predictable associations are tested, too (transfer or generalization).

The main advantage of the test is, that distinct neural structures are taking part in the two phases of the test (Figure 2.). Both the original and further neurocognitive and neuroimaging studies proved that the acquisition phase primarily relies on the integrity of the basal-ganglia-frontal cortex loops, while the test phase primarily relies on the integrity of the hippocampal region (de Araujo Sanchez & Zeithamova, 2023; Gogtay et al., 2006; Larsen & Luna, 2015; Moustafa et al., 2010; Myers et al., 2003; Persson et al., 2014; Porter et al., 2015; Shohamy & Wagner, 2008).

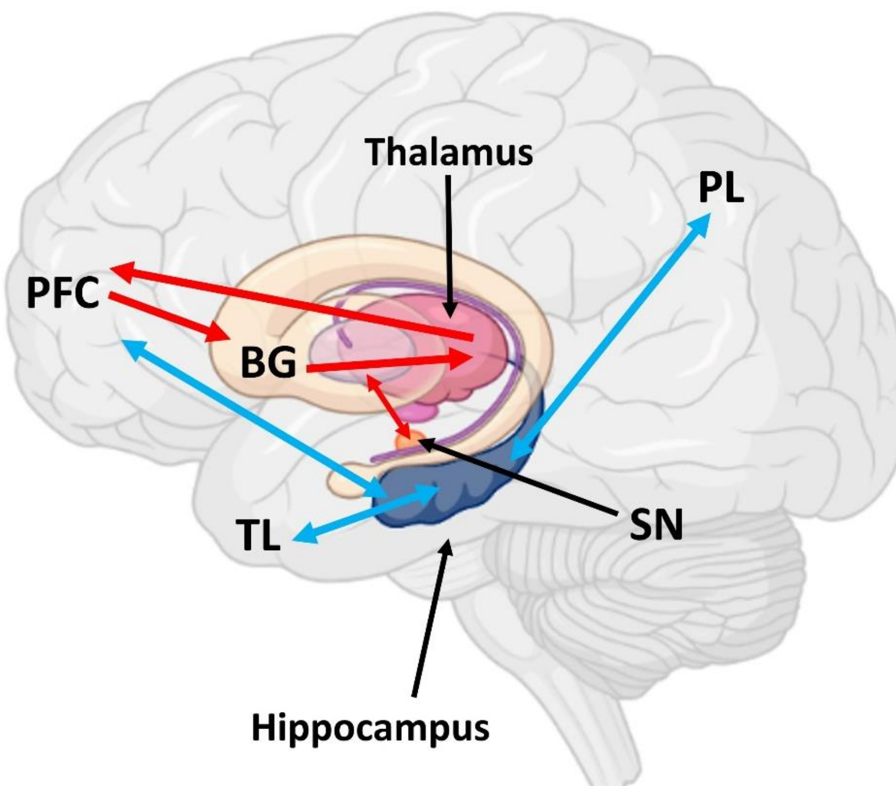

Figure 2. The neural structures involved in acquired equivalence learning.
The learning of the associations are mainly depends on the frontal lobe-basal ganglia loops (red arrows), whereas the retrieval and generalization on the MTL and its connections (blue arrows). PFC: prefrontal cortex; BG: basal ganglia; SN: substantia nigra; TL: temporal lope, PL: Parietal lobe **(***Created with* *BioRender.com.)*

Therefore, it is possible to use it to investigate the acquired equivalence learning in neurological and psychiatric disorders with basal ganglia or hippocampi dysfunctions. In the original study (Myers et al., 2003) patients with Parkinson's disease performed poorer in the acquisition phase, but not in the test phase. On the contrary, performance of patients with hippocampal atrophy was poorer in the test phase but not in the acquisition phase. Besides Parkinson's disease the acquisition was also impaired in borderline personality disorder (Rosu et al., 2022) and Tourette syndrome (Eördegh et al., 2020). The test phase was impaired in Alzheimer's disease (Bódi et al., 2009) and schizophrenia (Kéri et al., 2005). In contrast, no deficit was found in generalized tonic clonic epilepsy (Khalil et al., 2015), children with obsessive–compulsive disorder (Pertich et al., 2020) and children with migraine (Giricz et al., 2021).

1.3. Multisensory integration

The surrounding world and the natural stimuli are usually multimodal. Therefore, combining the information from the various independent sensory modalities into a coherent and unitary representation, could be more informative than a unimodal stimulus from the environment. Consequently, it can help to detect and evaluate the biologically important events more effectively and it has an evolutionary advantage. This process is called multisensory integration (Gibney et al., 2017; Marucci et al., 2021; Meredith & Stein, 1983; Stein et al., 2014).

The responses to a multimodal stimulus are usually different than the sum of the responses to the unimodal stimuli (i.e. superadditive, subadditive effects or multisensory enhancement or inhibition, respectively; Marucci et al., 2021; Stein et al., 2020). Previous studies have proven that responses are faster and more accurate in the case of multimodal stimuli (Frens & Van Opstal, 1998; Giard & Peronnet, 1999; Harrington & Peck, 1998; Lunn et al., 2019; Pannunzi et al., 2015; Patching & Quinlan, 2004). It has been also demonstrated that one modality can enhance the perceived intensity of other modality (Gillmeister & Eimer, 2007;

Stein et al., 1996). Furthermore, greater enhancement can be observed in the case of stimuli with low efficacy (i.e. the principle of inverse effectiveness), which means that the response threshold can be lower to the multisensory information compared to the unimodal ones (Nagy et al., 2006; Stein et al., 2020).

Unimodal stimuli most likely will be integrated if they share a high temporal and spatial concordance (i.e. spatial principle of multisensory integration), which indicates that they are likely to be derived from the same event (Marucci et al., 2021; Stein et al., 2020). In this case the stimuli are mutually reinforcing the event, from independent sources. Congruency is also important in human communication: multisensory neurons in the prefrontal cortex shows sensitivity to facial expression and sound synchrony and context of vocalization. When the different unimodal stimuli are incongruent; therefore, they refer to different events, stimuli are either not integrated or lead to degraded products (Stein et al., 2020).

The principles of multisensory integration can be observed at the cellular (Minciacchi et al., 1987; Nagy et al., 2006; Reig & Silberberg, 2014; Wallace et al., 1998), network/circuit (Macaluso & Driver, 2005; Senkowski et al., 2011; Senkowski et al., 2007), and behavioral levels (Godfroy-Cooper et al., 2015; Lanz et al., 2013; Wallace & Stevenson, 2014).

The multisensory integration is also beneficial for learning and memory (Murray & Shams, 2023; Shams & Seitz, 2008). It was shown that multisensory learning and encoding can lead to a faster learning and improves also the memory performance compared to the unimodal learning. This was demonstrated in perceptual learning tasks (Barakat et al., 2015; Kim et al., 2008), implicit learning tasks (Ernst, 2007; Seitz et al., 2007) and recognition tasks (Fort et al., 2002; Frassinetti et al., 2002; Joassin et al., 2011; Love et al., 2011; Meyerhoff & Huff, 2016; Suied et al., 2009). Several behavioral studies reported that reaction times for multisensory stimuli are faster (Gibney et al., 2017; Hershenson, 1962; Miller, 1982; Patching & Quinlan, 2004; Raab, 1962). Furthermore, crossmodal information in a temporal synchrony can facilitate selective attention, thus further improving the multisensory processing (Gibney et al., 2017; Santangelo & Spence, 2007; Talsma et al., 2010).

It is known from earlier studies that brain structures fundamentally involved in acquired equivalence learning (i.e. the basal ganglia and the hippocampi) are involved in multisensory information processing and multisensory integration, too (Bates & Wolbers, 2014; Chudler et al., 1995; Lee et al., 2017; Nagy et al., 2006; Nagy et al., 2005; Ravassard et al., 2013; Schwarz et al., 1984). Therefore, to investigate the effect of multimodal information on the AEL, our research group has developed an audiovisual test (SoundFace) based on the original RAET. In this test, the antecedent stimuli are distinct sounds, and the consequents are drawn faces (Eördegh et al., 2019). If we compare the learning performances of healthy humans in the two tests (RAET and SoudFace) there is no difference in the effectiveness of the equivalence learning. A possible explanation is that this type of associative learning is a very old and conserved function, where the multisensory information is not superior compared to unimodal modalities, which contribute to the associative learning equally (Eördegh et al., 2019). On the other hand, the retrieval and generalization functions are enhanced in the audiovisual test, which can be clearly seen in the shorter reaction times in the audiovisual paradigm.

1.4. Stimuli complexity and verbalizability

As was mentioned before, some of the previous studies found no deficit in the visually guided AEL among pediatric patients with obsessive-compulsive disorder and migraine (Giricz et al., 2021; Pertich et al., 2020). This raises a question regarding whether there are no differences in the learning abilities between the healthy and patient populations, or the test is not sensitive enough to detect the differences on behavioral level. We argue, that in the RAET, which applies drawn faces and colorful fishes, the stimuli are rich in details, can be verbalized easily, and has a lot of semantic meanings, which can facilitate the learning process, and help to make associative connections between the stimulus pairs. Furthermore, it is possible that the declarative memory with cortical contribution is able to play a compensatory role, even if these implicit learning functions are impaired (Ullman & Pullman, 2015).

Despite this interest, the effect of stimulus complexity and verbalizability is not widely investigated in the field of associative learning, and as far as we know, no one has ever studied it in the acquired equivalence task. It is known from perception research, that stimulus complexity is one of the most important properties, which can influence the performance (Donderi, 2006; Güçlütürk et al., 2018). It was demonstrated that visual stimulus complexity can improve speed and accuracy of shape recognition (Kayaert & Wagemans, 2009). It can influence search time (Donderi, 2005), discrimination (Mavrides & Brown, 1969), recognition tasks (Anderson & Leonard, 1958), perceptual learning (Ahissar & Hochstein, 1997; Bakhtiari, 2020; Maor et al., 2019) and last bot not least the memory (Chai et al., 2010). Furthermore, Güçlütürk and colleagues (2018) observed that more complex visual and auditory stimuli can cause more prominent cortical activation.

Bender and colleagues (2017) found that faces and face-name associations with high visual stimulus complexity and saliency were easier to recognize and required shorter processing time. The higher visual complexity of the items may enhance the distinctiveness and saliency of the stimuli and can be helpful in encoding and recognition of individual items and associations.

The other important factor, which can influence the learning is the verbalizability of the stimulus. The study by Arslan and colleagues (2020), provides an evidence that visual stimuli complexity can influence visual memory performance. They applied stimuli sets with low and high level of verbalizability in a visual serial recall test. If the visual stimuli were more verbalizable, the performance of the participants was better in the task, contrary to the less verbalizable stimuli, which negatively influenced the performance.

Verbalizability is also a discussed topic in category learning. According to the literature both humans and animals can learn categories by means of explicit or implicit way (Ashby & Ell, 2001; Maddox & Ashby, 2004; McLaren, 2018; Smith et al., 2012; Wasserman et al., 2023). The COVIS model states (Ashby et al., 1998; Ashby & Maddox, 2005, 2011) that both memory systems are contributing to category learning simultaneously, but competitively. Categories, where the organizing rules can be easily described verbally are processed predominantly by an explicit way. For this reason, the person must be able to attend selectively to each relevant stimulus feature, and a semantic label must be corresponding to each of these features. On the other hand, when the organizing rules are difficult to verbalize, the categories will be learned gradually, in an implicit way (Ashby & Maddox, 2005).

However, little research has been conducted to investigate specifically the effects of feature type (easy to verbalize vs difficult to verbalize) on category learning. The work by Zettersten and Lupyan (2020) have applied colors and shapes which were easy or difficult to verbalize, and they found that the learning rate was faster in the case of easily verbalizable features, but the study did note address the categorization strategy used by the participants. However, in another study by Brashears and Minda (2020, Preprint) conducted a category learning task in an interactive video game environment by using stimuli with features, which were easy or difficult to describe. The number of trials was less, but not significantly in the case of easily verbalizable stimuli; however, the accuracy was significantly higher compared to the not-easily verbalizable stimuli. Furthermore, they found that participants presented with easily verbalizable stimuli were more likely to use the explicit way of learning, and in the case of not-easily verbalizable features the implicit learning was more likely applied.

Based on these findings we ask the question, whether the complexity and verbalizability could influence the visually guided AEL and the connected memory processes the same way as in the category learning. In fact, AEL can be viewed as a specific kind of category learning (Urcuioli, 2001). In order to investigate this, we developed a new test (Polygon), where we simplified the visual stimuli. We applied two-dimensional geometric shapes (grayscaled circles as antecedents; a triangle, a square, a rhombus, and a concave deltoid as consequents) instead of colored drawn faces and fishes. We argue that these geometric shapes are less complex, have less semantic meanings and are more difficult to verbalize them. We hypothesized that because of the application of the new simplified stimuli set, and the absence of any supporting information (i.e. semantic meaning, feature verbalizability) the learning will be more implicit, therefore, the test will be more sensible to the basal ganglia-frontal cortex functions.

As the next step, we also made a new audiovisual test with a reduced visual stimuli complexity and verbalizability. We kept the auditory stimuli from the SoundFace test (a sound of a cat, a guitar accord, a vehicle and a woman) as antecedents, and we changed the colored drawn faces to the two-dimensional geometric shapes from the Polygon test as consequents. The reason behind this is that, although there was no significant differences in the performances of the acquisition phase between the visual and audiovisual tests (RAET vs SoundFace) when we compared them with each other (Eördegh et al., 2019), the semantic meanings of the stumuli could also influence the results. In the SoundFace test the applied visual stimuli are the same drawn faces from the RAET. Furthermore, the auditory stimuli are ordinary sounds from the daily life. As a consequent, these can facilitate the learning process and the memory performance, because it is easier to make connections between the complex stimuli, due to previously learned associations or the semantic congruency (Lauzon et al., 2022).

It has been shown in several studies, that in audiovisual integration tasks, semantically matching combinations of auditory and visual stimuli result better memory performance, than mismatching stimuli (Amedi et al., 2005; Chen & Spence, 2010; Grassi & Casco, 2010; Heikkilä et al., 2015; Meyerhoff & Huff, 2016; Steinweg & Mast, 2017). Furthermore, adults tend to use previously learned associations in the multisensory processing. The stored memory representations for these prior experiences can include semantic, affective and relational cues, which can enhance the integration process (Lewkowicz, 2014; Ten Oever et al., 2016; Ten Oever et al., 2013).

As semantic congruency, and previously learned associations can modulate the multisensory integration; therefore, we ask the question how can affect the reduced stimulus complexity and verbalizability of the visual stimuli the participant's performances in the audiovisually guided AEL test.

2. Aims of the study

The aim of our studies was to investigate the visual and audiovisual associative learning in healthy adults. We aimed to validate our newly developed associative learning tests with reduced visual stimulus complexity, verbalizability and semantic meaning, and to examine how can this influence the performance of healthy adult participants in acquired equivalence learning and the connected memory processes. Therefore, we compare the performance in tests with more and less complex visual stimuli, both in a visually and an audiovisually guided acquired equivalence learning.

The specific objectives of our studies were:

- to compare the performances in the two visually guided paradigms, the one with more complex and verbalizable visual stimuli (RAET) with the other with less complex and verbalizable stimuli (Polygon). Special attention was focused on whether these possible differences are similar or different in the learning and the test phases of the learning task.
- to compare the performances in the two audiovisually guided paradigms (SoundFace vs SoundPolygon), which were different only in visual consequent stimulus complexity. We ask whether the same phenomena can be observed in the acquisition and the test phases of multisensory paradigm as in unimodal visual associative learning.

3. Methods

3.1. Participants

In our first study, where we compared the two visual tests (REAT vs Polygon), altogether 55 person participated (26 women and 29 men; mean age: 35.11 ± 13.925 years, range: 18–65 years), and this is true also to the second study, where we compared the two audiovisual (SoundFace vs SoundPolygon) tests (27 women and 28 men; mean age: 31.36 ± 14.56 years, range: 18–69 years). The estimated minimum sample size was 47, assuming $p < 0.05$, $1 - \beta = 0.95$ and an effect size of 0.5. The sample size estimation was performed in G*Power 3.1.9.2 (Düsseldorf, Germany).

The participants were recruited on a voluntary basis, no compensation was given, and they were free to quit any time without any consequence. The potential participants were informed about the goals and procedures of the research, and were asked about their health. All the participants were healthy adults without any neurological or psychiatric disorder. The subjects were tested by Ishihara plates to exclude color blindness. Those who decided to volunteer signed an informed consent form, that they took cognizance about the earlier mentioned informations. The study protocol followed the tenets of the Declaration of Helsinki in all respects, and it was approved by the Regional Research Ethics Committee for Medical Research at the University of Szeged, Hungary (27/2020-SZTE).

3.2. The applied visual tests

The tests were run on laptops (Lenovo ThinkBook 15 IIL, Lenovo, China). In the case of the audiovisual tests circumaural headphones were used (Sennheiser HD439, Sennheiser Germany). The testing sections took place in a silent room, where the participants could solve the paradigms without any distraction. They were tested one-by-one. No forced quick responses were expected from them to avoid time pressure. The participants executed both tests (RAET and Polygon, or SoundFace and SoundPolygon) one after another in a pseudorandom order, to statistically minimalize the carry-over effect.

We used two visual tests, one with more complex stimuli (modified RAET), and one with a simplified stimuli (Polygon). The original RAET was developed by Myers and colleagues (2003). The test was translated to Hungarian, got written in Assembly for Windows, and the stimuli were slightly modified (Eördegh et al., 2022; Eördegh et al., 2019; Öze et al., 2017).

The main goal of the tests is to learn associations between four antecedent stimuli and four consequent stimuli. In the RAET the antecedents were different drawn faces: a man, a woman, a boy and a girl. All of them possess three binary features: age (adult vs child), gender (male vs female) and hair color (brown vs black), hence they can compose pairs. The consequents were different colored drawn fishes, with the same shape: red, green, blue and yellow (Figure 3.). Each antecedent is associated with two consequents, which makes eight possible pairings. The rule of association, i.e. which stimulus feature (age, gender or hair color) is used to link the antecedent-consequent pairs, was randomly generated by the software for each participant. The subjects were unaware of this pairing rule.

In the case of the Polygon test, where we applied simplified stimuli. The antecedents were grayscale shaded circles: white, light gray, dark gray and black. Contrary to the faces in the RAET, they differed only in one feature. The consequents were different two-dimensional blank, geometric shapes: a triangle, a square, a rhombus and a concave deltoid (Figure 4.).

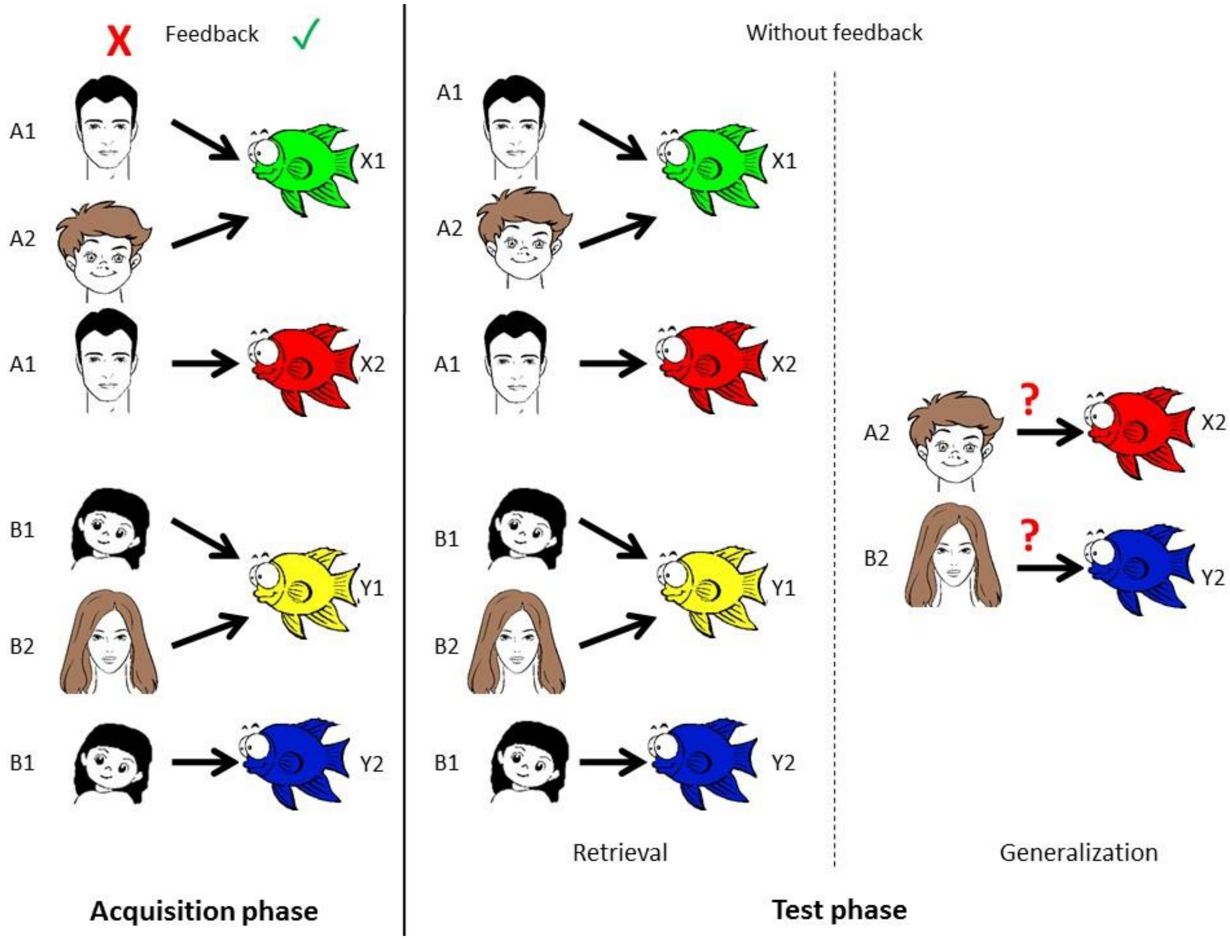

Figure 3. The overview structure of the Rutgers Acquired Equivalence Test

In this test the antecedents are drawn face, and the consequents different colored drawn fishes. The tests have two main phases. In the acquisition phase the participant learns the association pairs gradually, based on the feedback given by the software. To each fish two antecedents is associated; therefore, acquired equivalence will form between them. In the test phase no further feedback is given. The participant had to recall the already learned associations, and not seen pairs are presented, which are predictable by the acquired equivalence.

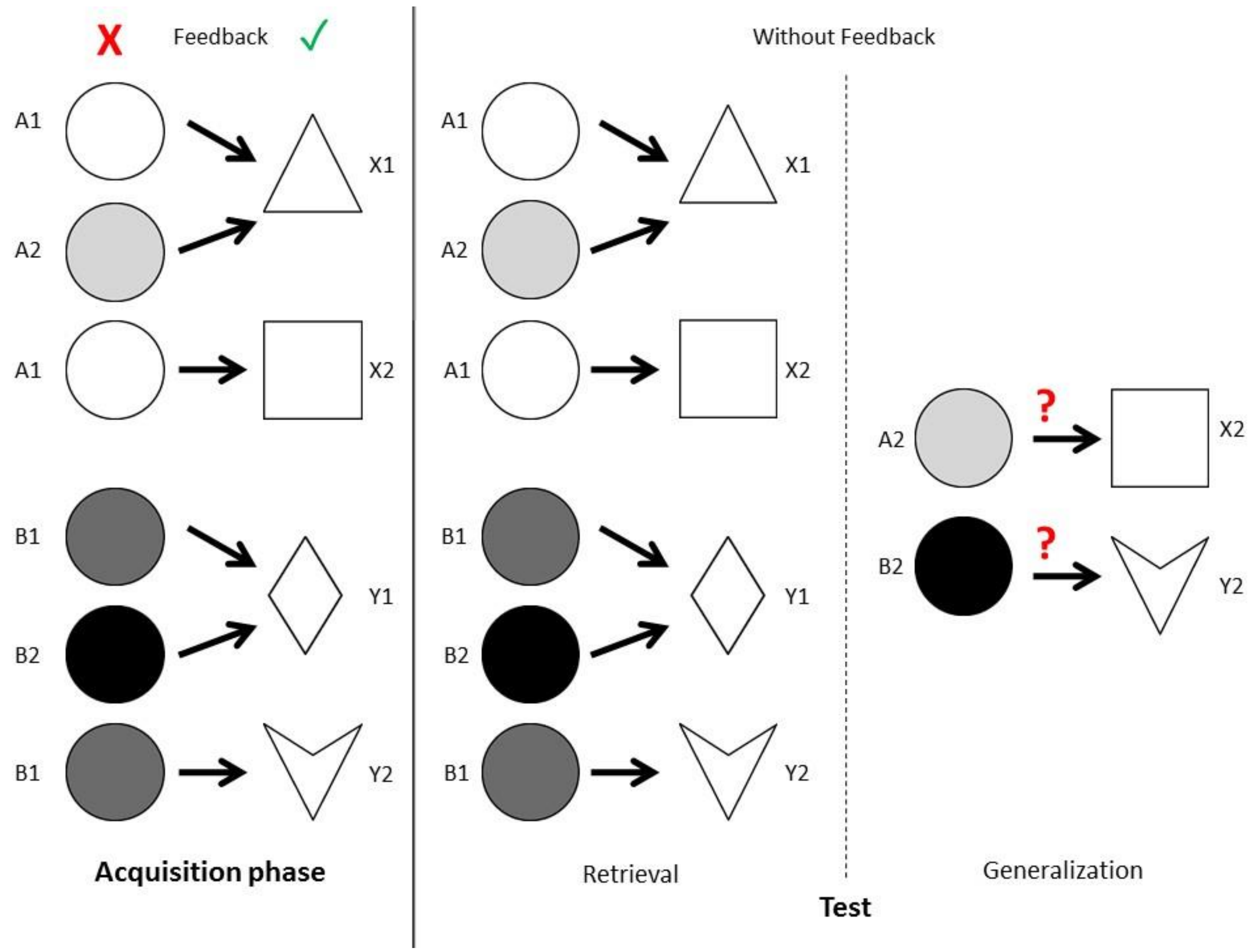

Figure 4. The overview structure of the Polygon test

The structure is the same as in the RAET. In this test the antecedents are grayscale shaded circles, and the consequents different two-dimensional geometric shapes.

Both tests have the same structure, which is divided into two main phases: acquisition and test phase. In the acquisition phase the participants learn the associations between the antecedents and consequents through trial and error. In each trial, an antecedent stimulus appears in the middle of the screen, and two consequents on the left and right sides below it. The participant had to guess which consequent is linked to the given antecedent, by pressing either the “left” or the “right” button corresponded to the consequents on each side of the monitor. The visual stimuli lasted until the participant pressed the button. The participant learns the associations, pair by pair, gradually, based on the feedback about the correctness of the responses given by the program. If the response is correct a green check mark appears with the word “Helyes!” (“Correct!”), and a big red X with the word “Helytelen!” (“Incorrect!”), when it is incorrect (Figure 5-6.).

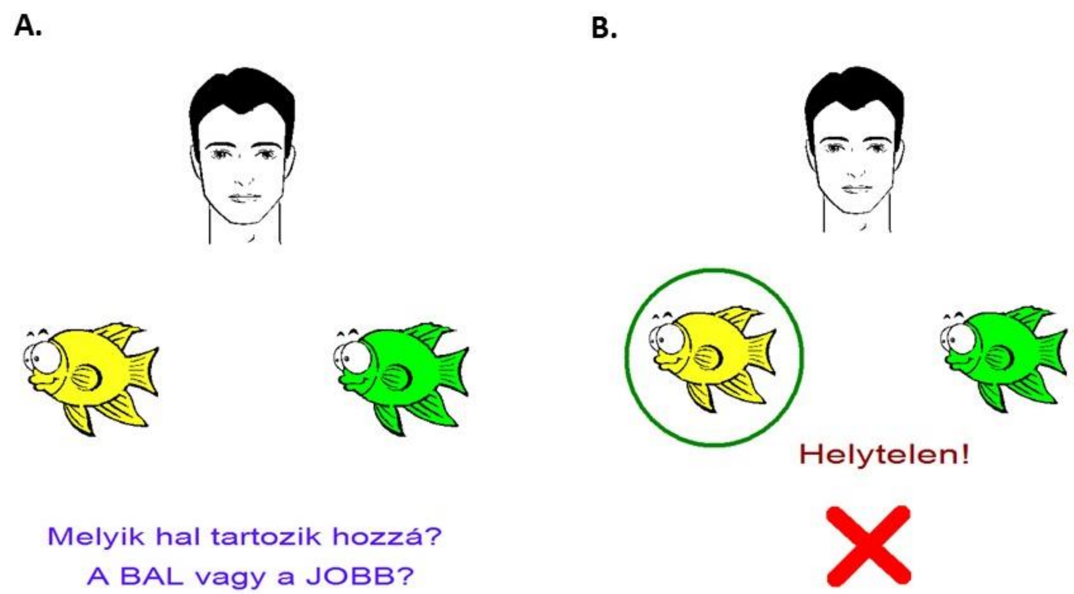

Figure 5. One possible trial in the RAET

A) *One drawn face appears in the middle of the screen (antecedent), and two fishes with different color (consequents) on the left and right sides below it. The participant makes a choice by pressing either the “left” or the “right” button corresponded to the consequents on each side of the monitor. The visual stimuli lasted until the participant pressed the button.*

B) *After the choice immediate feedback was given. In this case the response was incorrect, therefore a big red X appeared on the screen with the word “Helytelen!” (“Incorrect!”).*

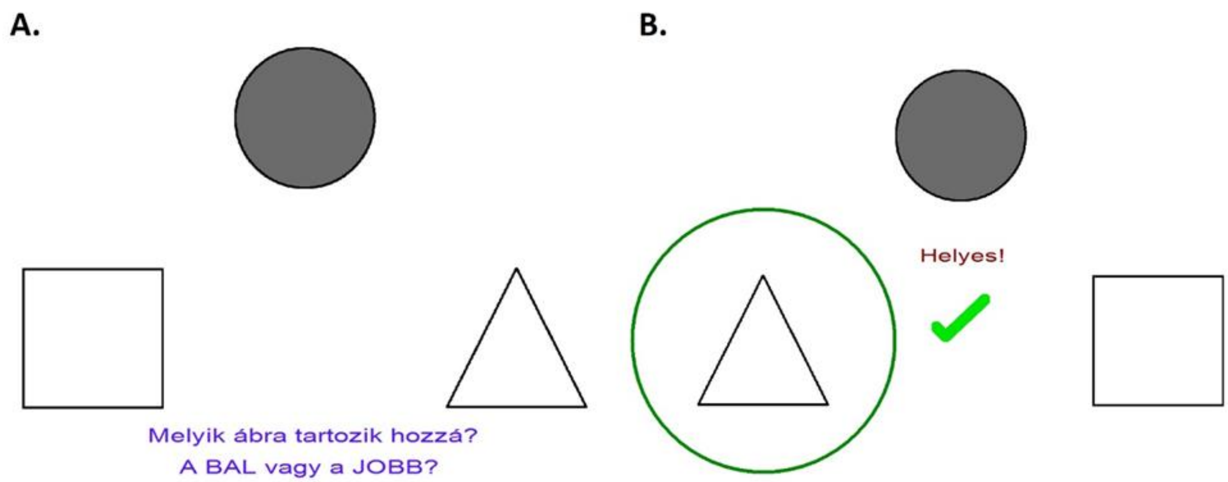

Figure 6. One possible trial in the Polygon test

A) *One circle (antecedent) appears in the middle of the screen, and two geometric shape (consequents) on the left and right sides below it. The participant had to choose which shape is linked to the given shaded circle.*
B) *In this case the response was correct; therefore, a green check mark appeared on the screen with the word "Helyes!" ("Correct!").*

After each new introduced stimulus pair, the participant must give a certain number of subsequent correct answers (4, 6, 8, 10, 12 after each new association, respectively) to proceed further in the test, and finally accomplish the acquisition phase. As consequence, the number of trials in this phase was varied between the participants, depending on learning performance. In the initial step the participants had to memorize the first antecedent-consequent pairs (A1: X1 and B1: Y1). In the next step two new antecedents were introduced (A2 and B2), which are linked with the same consequents as the previous ones, which means they are equivalent with each other. This step is the equivalence learning. After this, a new pair of consequents are introduced (X2 and Y2). This way A1 and B1 antecedents have two more additional consequents (A1: X1, X2 and B1: Y1, Y2). At this point, from the eight possible pairs six were presented, and if the certain number of correct answers was given the acquisition phase is accomplished.

In the test phase, no further feedback was given about the correctness of the responses. In this phase, the participant had to recall the already learned associations (retrieval). Furthermore, the two remaining pairs (A2: X2 and B2: Y2) were presented, and if the equivalence forming was successful, the participant could conclude the correct answers (generalization), even if he or she has not seen these pairs in the acquisition phase. About the new pairs in the test phase the participants were not informed. Altogether, there are 48 trials in the test phase, of which 36 is retrieval and 12 is generalization. Contrary to the acquisition, the number of trials in this phase are fixed. The order of the two types of trials are randomly mixed in the test phase. The overview structure of the tests can be seen in the Table 1.

ACQUISITION			**TEST**	
Shaping	**Equivalence training**	**New consequents**	**Retrieval**	**Generalization**
A1 -> X1	A1 -> X1 A2 -> X1	A1 -> X1 A2 -> X1 A1 -> X2	A1 -> X1 A2 -> X1 A1 -> X2	A2 -> X2
B1 -> Y1	B1 -> Y1 B2 -> Y1	B1 -> Y1 B2 -> Y1 B1 -> Y2	B1 -> Y1 B2 -> Y1 B1 -> Y2	B2 -> Y2

Table 1. A summary of the associative learning paradigms.

The tests have two main phases. In the acquisition phase the participant learns the first antecedent (A or B) and consequent (X or Y) pairs. The next step is the introduction of a new antecedent which shares the same consequent (equivalence learning). In the final step one more consequent is introduced to one of the antecedents. In the test phase the participant had to recall the already learned associations, and the remaining, not seen pairs are presented, which are predictable by the acquired equivalence.

3.3. The applied audiovisual tests

The audiovisual tests have the same structure as the visual tests, but the antecedents were clearly distinguishable distant sounds: a cat, a guitar accord, a sound of a vehicle and a woman saying "Hello". These stimuli where shown to the participants before the testing session to make sure they could hear them correctly. In the SoundFace test the consequents are the same drawn faces mentioned in the RAET (Figure 7.). In the SoundPolygon they are identical to those geometric shapes, which we applied in the Polygon test (Figure 8.). As mentioned above, the consequents of the Polygon test have less feature as the drawn faces.

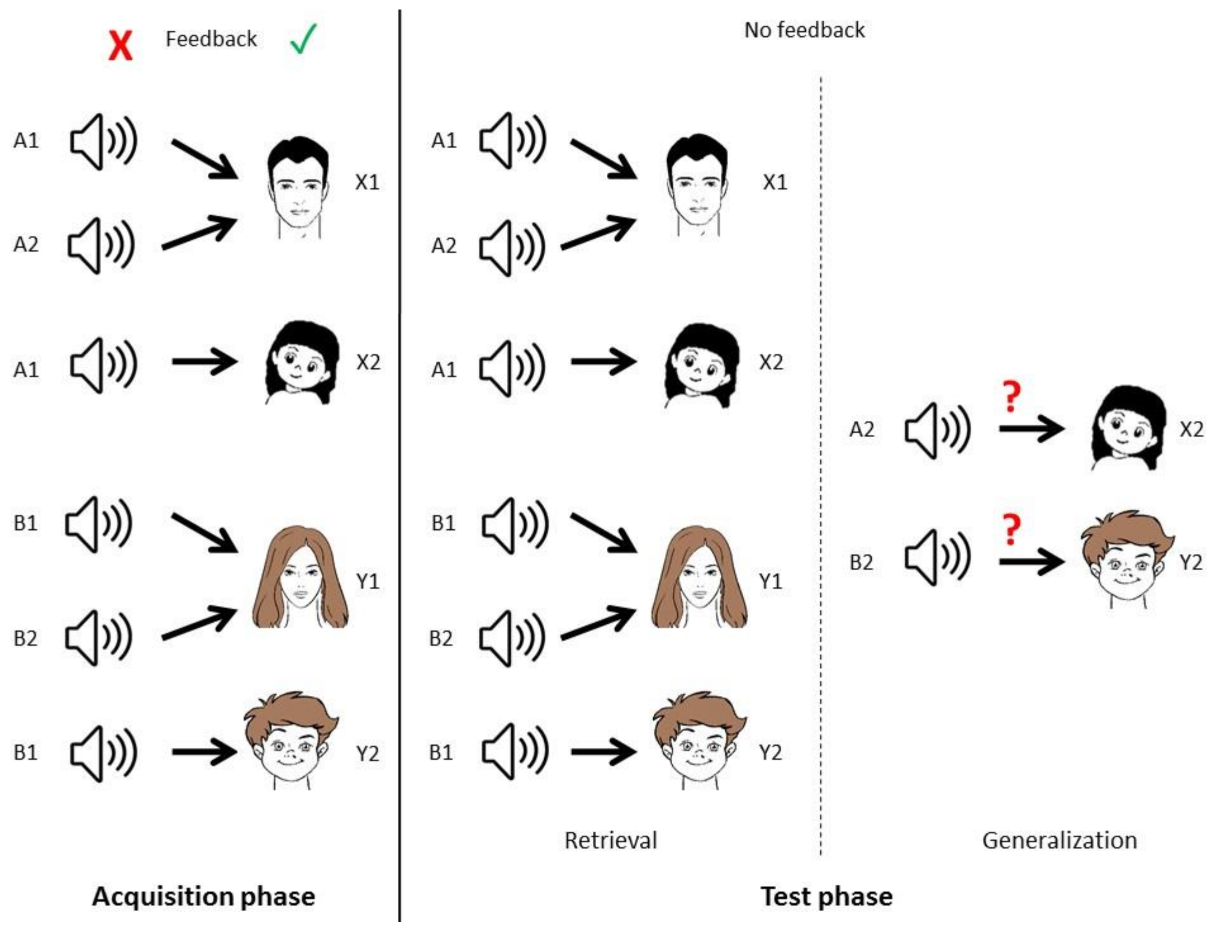

Figure 7. The overview structure of the SoundFace test

The structure is the same as in the RAET. In this test the antecedents are distinct sounds (a cat, a guitar accord, a sound of a vehicle and a woman saying "Hello") and the consequents different drawn faces.

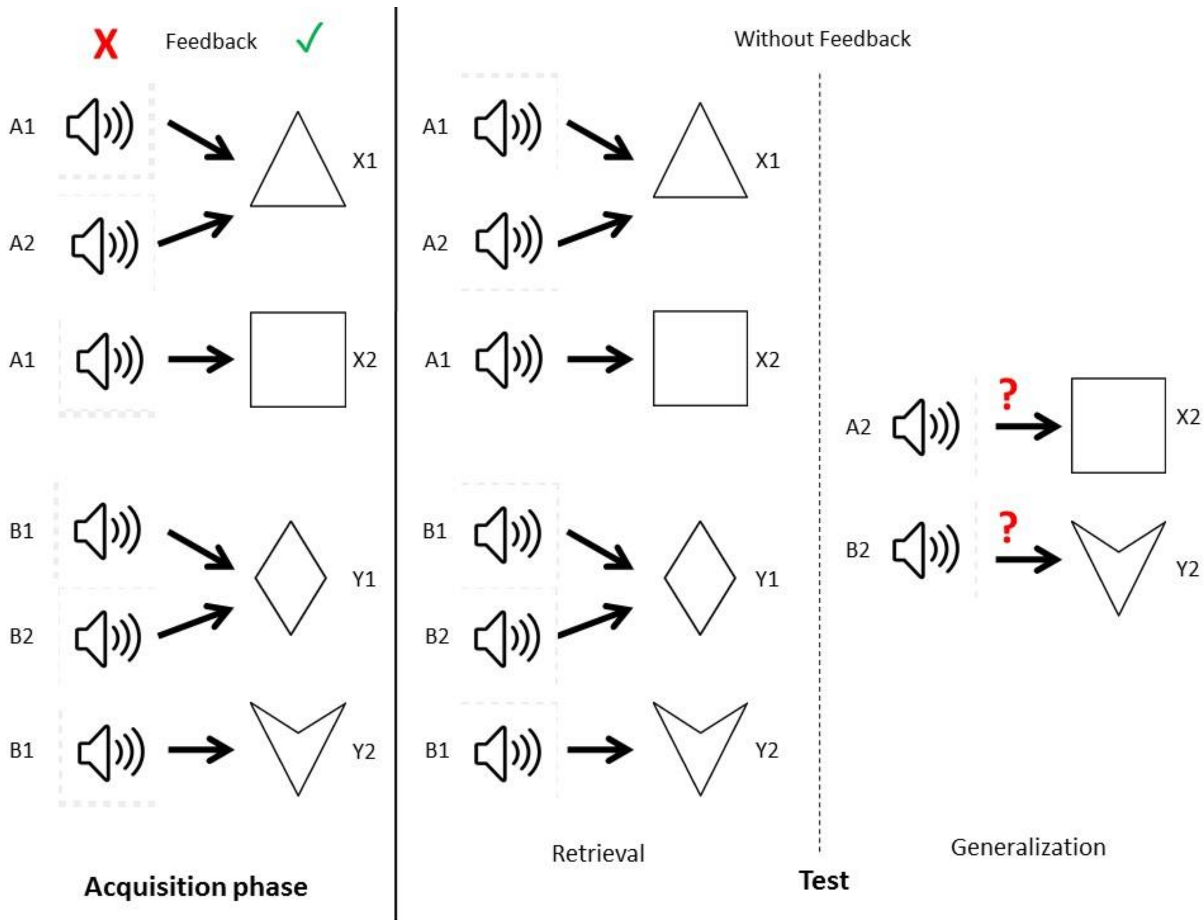

Figure 8. The overview structure of the SoundPolygon test

The structure is the same as in the RAET. In this test the antecedents are the same sounds from the SoundFace test, and the consequents are two-dimension geometric shapes.

In each trial the participants simultaneously hear a sound and see two faces/geometric shapes on the left and right side of the screen, respectively (Figure 9-10.). They have to choose which face belongs to the given sound with the same procedure described earlier. The duration of the auditory stimuli was fixed to 1.5 s, and visual stimuli lasted until the participant made a decision, by pressing down the "left" or "right" button corresponded to the consequents on each side of the monitor.

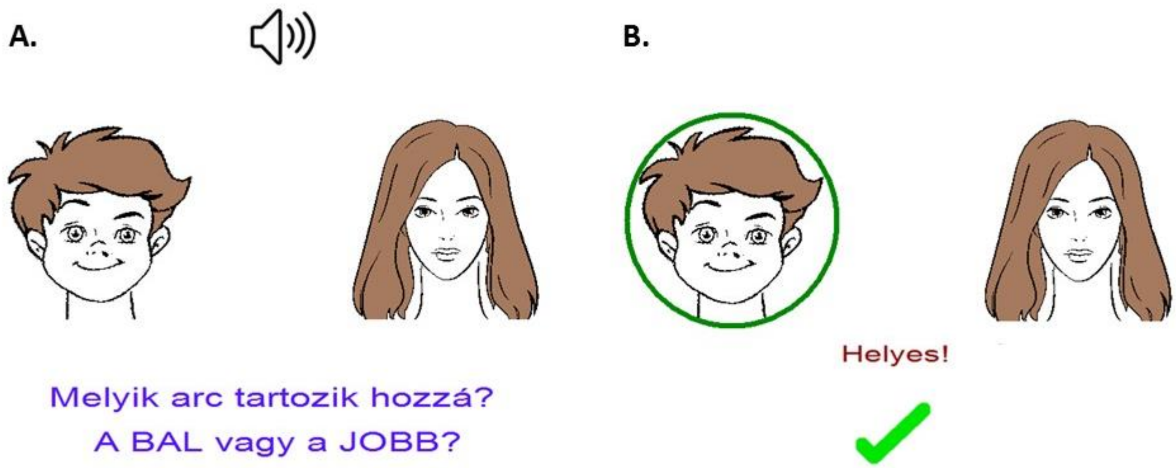

Figure 9. One possible trial from the SoundFace test

A) *A sound is presented as antecedent, simultaneously with two visual consequents (faces) on the left and right sides of the screen. The participant makes a choice by pressing either the "left" or the "right" button corresponded to the consequents on each side of the monitor. The duration of auditory stimuli was 1.5 s, but the visual stimuli lasted until the participant pressed the button.*

B) *In this case the response was correct; therefore, a green check mark appeared on the screen with the word "Helyes!" ("Correct!").*

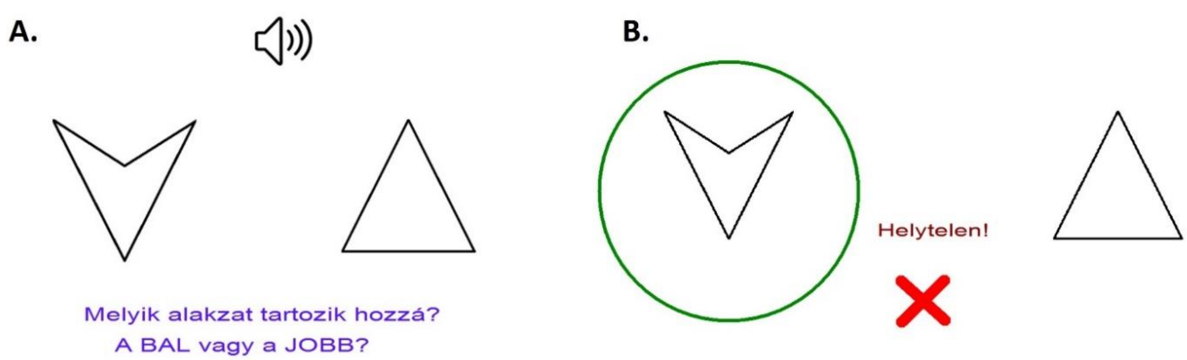

Figure 10. One possible trial from the SoundPolygon test

A) *A sound is presented as antecedent, simultaneously with two visual consequents (two-dimensional geometric shapes) on the left and right sides of the screen. The participant makes a choice by pressing either the "left" or the "right" button corresponded to the* *consequents on each side of the monitor. The duration of auditory stimuli was 1.5 s, but the visual stimuli lasted until the participant pressed the button.*

B) *After the choice immediate feedback was given. In this case the response was incorrect, therefore a big red X appeared on the screen with the word "Helytelen!" ("Incorrect!").*

3.4. Data analysis

The following cognitive learning parameters were calculated and analysed in each test: the number of trials (NAT), the error ratios in the acquisition (ALER), retrieval (RER) and generalization (GER) sections. The first two parameters are the indicators of the learning performance, and the last two are the indicators of the test phase. The error ratios were calculated by dividing the number of incorrect answers by the number of trials of the given section. The number of trials (NAT) in the acquisition phase was variable and performance dependent, but it was fixed in the case of retrieval (36) and generalization parts of the test phase (12).

We also recorded the response times (RT) in millisecond accuracy, and calculated the average value in the acquisition, retrieval and generalization. The response times are the time duration between the appearance of the stimuli and the participant's response. Values greater than 3SD were excluded.

The statistical analysis was performed in Statistica 14.0.0.15 (TIBCO Software Inc.). The Shapiro-Wilk normality test was applied to decide the data whether, or not has a normal distribution. Due to nonparametric distribution the Wilcoxon Matched-Pairs test was used to compare the cognitive learning perfromances in the different tests (RAET vs Polygon and SoundFace vs SoundPolygon).

4. Results

4.1. The comparison of performances in the visual tests

Altogether 55 healthy adult volunteers participated in this study. Exception of one participant (who was unable to complete the Polygon test but had no problem with the completion of the RAET) all of them could complete both the RAET and the Polygon tests. Therefore, the statistical comparison of the performances of 54 volunteers will be presented in detail. The results are presented according to the two main phases of the test paradigm (acquisition and test).

4.1.1. Acquisition phase

In the RAET the median of the NAT was 54.0 (range: 43.0-136.0), but in the Polygon it was 68.5 (range: 42.0-213.0). In the Polygon test, significantly more trials were necessary to learn the associations than in the RAET (Wilcoxon Matched-Pairs Test $Z = 3.731$, $p = 0.0002$; Figure 11.). In the RAET the median of the ALER was 0.052 (range: 0.0-0.240), and in the case of the Polygon test it was 0.096 (range: 0.0-0.340). Similarly, to the NAT, the difference in ALER was also strongly significant between the two tests (Wilcoxon Matched-Pairs Test $Z = 3.939$, $p = 0.00008$; Figure 11.).

Furthermore, the response times (RT) were significantly longer in the Polygon test (Wilcoxon Matched-Pairs Test $Z = 2.983$, $p = 0.003$). The median RT in the RAET was 1606.22 ms (range: 1004.74-3052.88 ms) and it was 1802.06 ms in the Polygon test (range: 888.49-4618.79 ms).

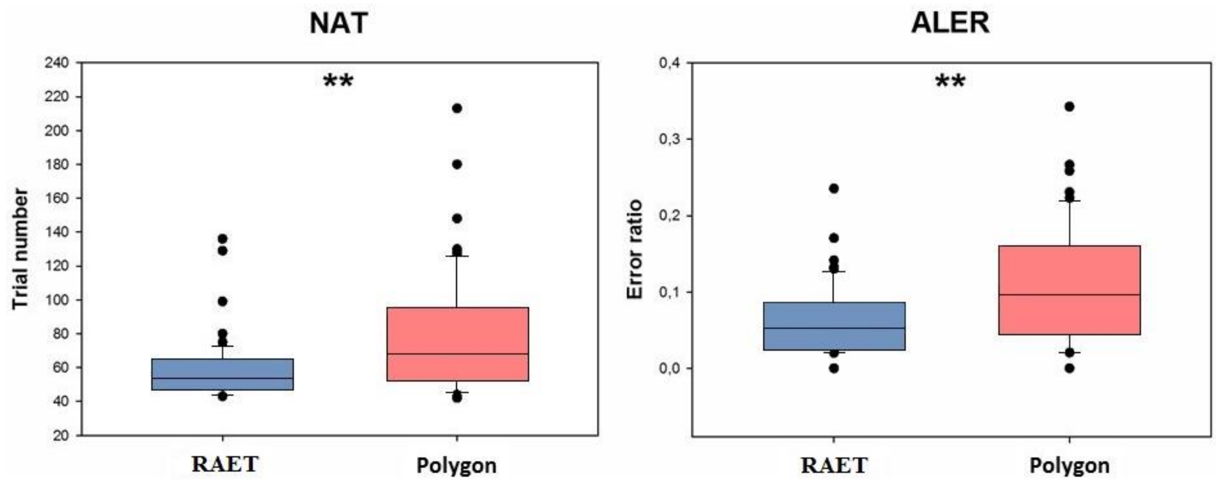

Figure 11. The performances in the acquisition phase in the two visual associative learning tests

NAT indicate the number of the necessary trials to complete the acquisition phase; ALER shows the error ratios in the acquisition phase. The lower margin of the boxes marks the 25th percentile; the line within the boxes the median; and the upper margin the 75th percentile. The error bars (whiskers) above and below the boxes indicate the 90th and 10th percentiles, respectively. The dots over and under the whiskers show the extreme outliers. The two black stars indicate highly significant differences ($p < 0.01$).

4.1.2. Test phase

In contrast to the acquisition phase, there were no significant differences in the performance between the RAET and the Polygon tests, both in retrieval (Wilcoxon Matched-Pairs Test Z = 0.739, p = 0,460; Figure 12.) and generalization (Wilcoxon Matched-Pairs Test Z = 1.624, p = 0.104; Figure 12.) parts of the test phase. In the RAET the median of RER was 0.028 (range: 0.00-0.31) and in the Polygon test it was 0.00 (range: 0.00-0.42). The median of GER in the RAET was 0.083 (range: 0.00-1.00) and it was 0.00 (range: 0.00-1.00) in Polygon test.

The response times were also not significantly different in the retrieval (Wilcoxon Matched-Pairs Test Z = 0.667, p = 0,505), and in the generalization (Wilcoxon Matched-Pairs

Test Z = 0.595, p = 0,552) parts of the test phase between the two paradigms. The median RT of retrieval in the RAET was 1840.71 ms (range: 1156.26-4046.87 ms), and it was 1763.73 ms in the Polygon test (range: 934.56-4036.71 ms). The median RT in the generalization part of the RAET was 2127.58 ms (range: 1300.67-13075.50 ms) and it was 2450.20 ms in the Polygon test (range: 1048.25-8230.50 ms).

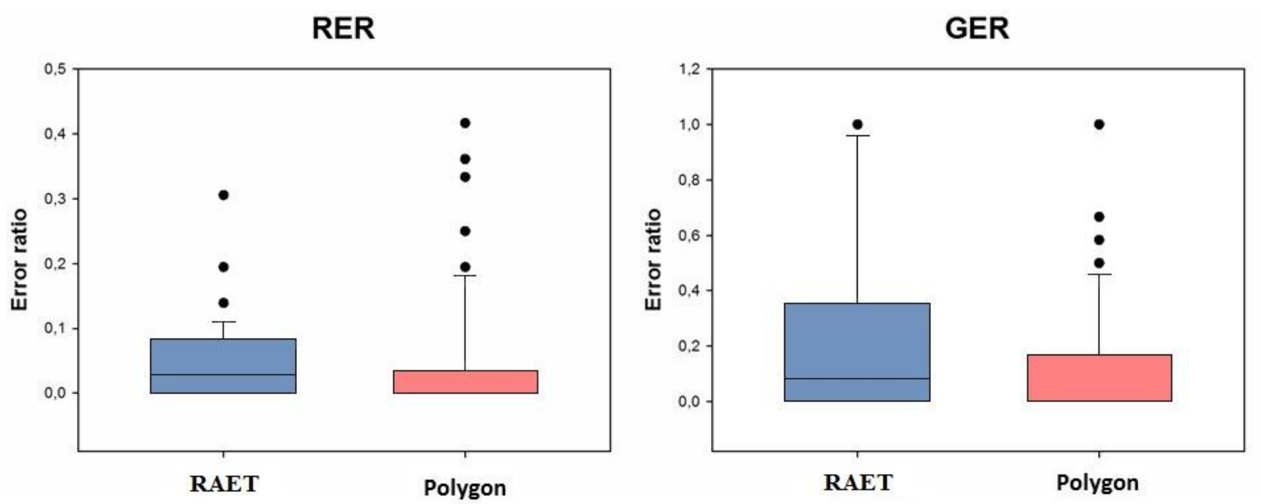

Figure 12. The performances in the test phase in the two visual associative learning tests

RER shows the error ratio of in retrieval and GER in the generalization parts of the test phase. Other notations are the same as on Figure 11.

4.2. The comparison of performances in the audiovisual tests

Alltogether 55 volunteers participated in the research. All of them could complete both tests (SoundFace and SoudPolygon). The analysis of their performance data is presented according to the two main phases of the test paradigm (acquisition and test).

4.2.1. Acquisition phase

The median number of trials in acquisition (NAT) in SoundFace was 47 (range: 41-113), and in the SoundPolygon it was 53 (range: 42-149). The participants needed significantly more trials to learn the associations in SoundPolygon ($Z = 2.417$, $p = 0.016$; Figure 13.).

In SoundFace, the median of error ratios in the acquisition (ALER) was 0.038 (range: 0.00-0.19), and in SoundPolygon, it was 0.058 (range: 0.00-0.28). The difference between the two tests was significant ($Z = 2.213$, $p = 0.027$; Figure 13.). The median response time (RT) for the acquisition trials in SoundFace was 1611.619 ms (range: 1095.022-4016.42 ms), and in SoundPolygon, it was 1834.810 ms (range: 1204.867-4762.33 ms). The difference was significant ($Z = 3.703$, $p = 0.0002$).

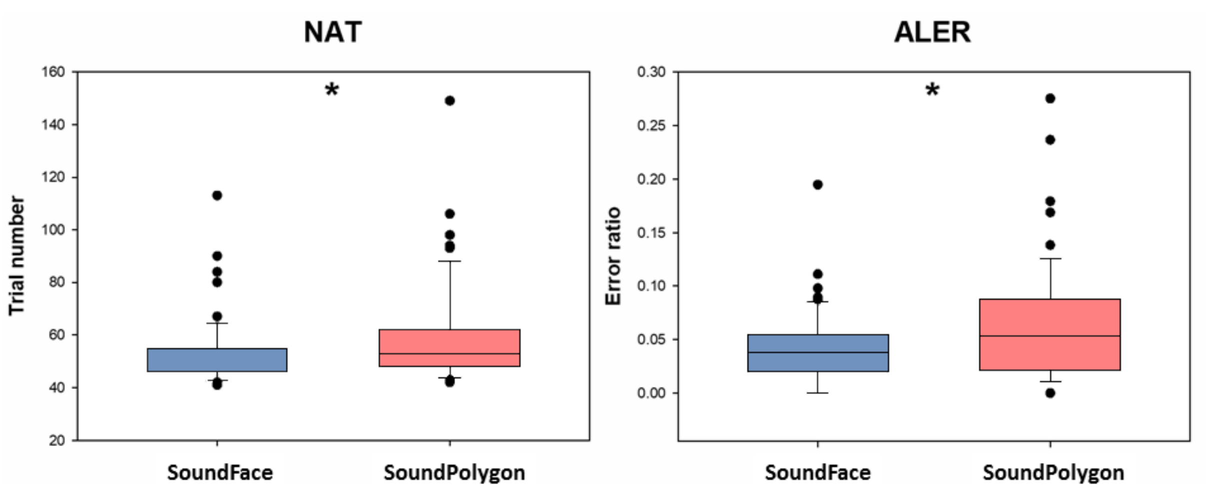

Figure 13. The performances in the acquisition phase in the two audiovisual associative learning tests

The notations are the same as on Figure 11. The black stars indicate the significant differences ($p < 0.05$).

4.2.2. Test phase

The median of retrieval error ratio (RER) in SoundFace was 0.00 (range: 0.00-0.25), and in SoundPolygon, it was 0.028 (range: 0.00-0.39). The difference was significant (Z = 2.727, p = 0.0064; Figure 14.). As for the response times, the median RT for the retrieval trials in SoundFace was 1586.39 ms (range: 980.056-3802.11 ms), and in SoundPolygon, it was 2000.42 ms (range: 1222.33-4790.44 ms). The difference was significant (Z = 4.994, p = 0.000001).

The median of generalization error ratio (GER) in SoundFace was 0.00 (range: 0.00-0.58), and in SoundPolygon, it was 0.083 (range: 0.00-1.00). The difference was significant (Z = 3.085, p = 0.002; Figure 14.). The median RT for the generalization trials in SoundFace was 1769.17 ms (range: 1145.75-5722.83 ms), and in SoundPolygon, it was 2544.25 ms (range: 1282.58-18381.00 ms). The difference was significant (Z = 3.938, p = 0.00008).

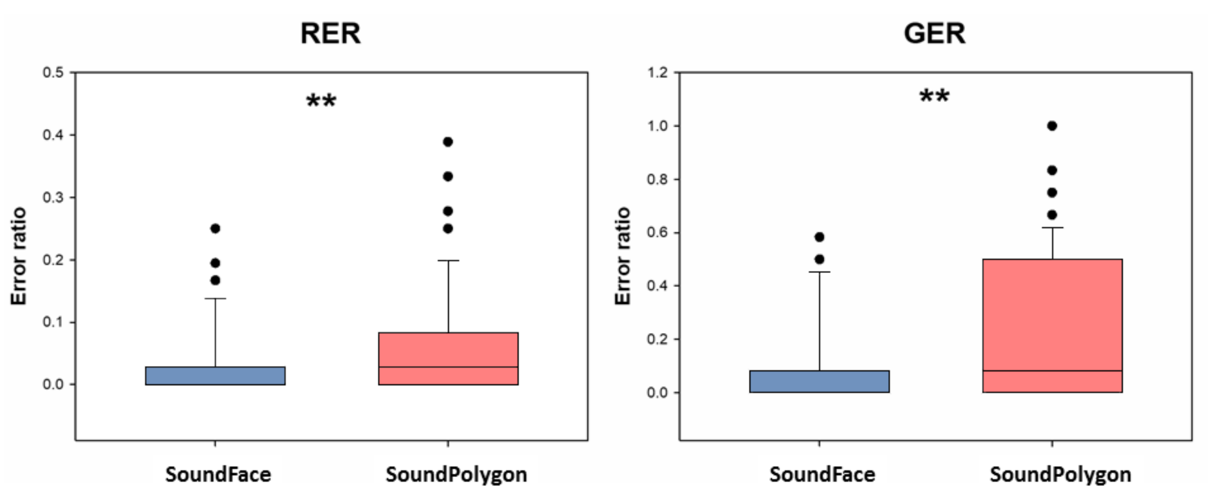

Figure 14. The performances in the acquisition phase in the two audiovisual associative learning tests

The notations are the same as on Figure 11. The two black stars indicate highly significant differences (p < 0.01).

5. Discussion

The Rutgers Acquired Equivalence Test (RAET) is a visually guided associative learning test, which built up from two phases, and in each phase primarily distinct brain structures are involved. The performance in the acquisition phase provides informations about the functions of the frontal-striatal loops, whereas the performance in the test phase provides informations about the functions of the hippocampus and medial temporal lobe (Moustafa et al., 2009; Myers et al., 2003). Because of this dissociation, the test is beneficial to investigate psychiatric and neurological disorders where the given brain structures are concerned. Our laboratory used the modified RAET to investigate clinical populations, and in some conditions altered performance has been found, but not in others (Bódi et al., 2009; Eördegh et al., 2020; Giricz et al., 2021; Khalil et al., 2015; Kéri et al., 2005; Myers et al., 2003; Pertich et al., 2020; Rosu et al., 2022) Therefore, the sensitivity of this test can be questionable in the case of mild changes in psychophysical performance. Having recognized this, we aimed to improve the sensibility for basal ganglia-frontal cortex functions and developed a new visually guided equivalence learning test (Polygon). We argue that the applied stimuli in the RAET (i.e drawn faces and drawn fishes with different color) are complex stimuli, which can have semantic and verbal information, and they could serve as an extra clue for cognitive associations, and this gives a possibility to compensatory mechanisms involving other cortical areas (Ullman & Pullman, 2015).

For this reason, in the new test (Polygon) we reduced the visual stimuli. We have applied grayscale shaded circles instead of drawn faces as antecedents, and simple two-dimensional blank geometric shapes instead of colored drawn fishes as consequents. The primary aim of this changes was to reduce the supporting information (semantic features, verbalizability and color information), make the test more implicit, and due to this to reduce cortical contribution (Puszta et al., 2019). Therefore, the test can be more sensible regarding to basal ganglia-frontal loop dysfunctions, and it could be suitable to detect slight differences, which were not detectable with the RAET (Giricz et al., 2021; Khalil et al., 2015; Pertich et al., 2020). The test also enables us to investigate the associative learning in healthy population and broaden our general knowledge about learning and memory, and how stimulus complexity influences them. To our knowledge, this is the first study which investigates stimulus complexity in visually guided equivalence learning in healthy adult population.

The participants' performances were significantly poorer in the acquisition phase of the Polygon test, compared to those of the RAET, as indicated by the significantly higher error ratios and the number of acquisition trials, even though the response times were significantly longer. Our results suggest that acquired equivalence learning, which is primary linked to intact basal ganglia functions, is strongly influenced by the complexity and verbalizability of the applied visual stimuli. This is in line with earlier findings, which demonstrated that stimulus complexity could influence memory performance in a visual serial recall test (Arslan et al., 2020) and less verbalizable stimuli can lead to a poorer performance. The same can be found in the study by Zettersten and Lupyan (2020) where the learning rate was faster for categories, when the features of the stimuli were easily verbalizable. Furthermore, another category learning study by Brashears and Minda (2020, Prepint) is consistent with these findings. In this study the partipicants learned categories in an interactive video game environment with easily verbalizable or not-easily verbalizable stimulus features and had a better performance in the case of easily verbalizable stimuli.

As for the test phase, the error ratios in the retrieval and generalization parts of the test phase did not differ between the RAET and Polygon tests, and neither did the response times. In other words, stimulus complexity and verbalizability had no effect on the retrieval of the previously learned associations and on the transfer of the acquired equivalence to previously unseen stimulus pairs.

In summary, our results suggest that the stimulus complexity and verbalizability are important factors for visual AEL and can have a considerable influence on the learning performance, but they have none or only a weak effect on the connected memory processes. We conclude that due to the simplified stimuli set in the Polygon, participants tend to shift to a more implicit learning, with less cortical contribution by the declarative memore system. The results also suggest that the Polygon test is more sensitive regarding to basal ganglia-frontal loop functions, than the original RAET.

It is well known from earlier studies that both brain structures fundamentally involved in visual associative learning, the basal ganglia and the hippocampi receive not only visual but also multisensory information (Bates & Wolbers, 2014; Chudler et al., 1995; Lee et al., 2017;

Nagy et al., 2006; Nagy et al., 2005; Ravassard et al., 2013; Schwarz et al., 1984). A bimodal or multimodal information could be more informative in its complexity than a unimodal stimulus from the environment. In our second study we compared the learning perfromances in two audiovisual tests (SoundFace vs SoundPolygon), which applied the same auditory stimuli with different (simple or complex) visual stimuli set. The visual stimuli set differed in their feature richness and verbalizability. To our knowledge, this is the first study to investigate the effect of visual stimulus complexity and verbalizability on multisensory AEL.

Both audiovisual tests had the same structure as the original visual RAET (Myers et al., 2003). The antecedent stimuli were distinct sounds (a cat, a guitar note, a sound of a vehicle and a woman saying “Hello”), and the consequents are the same drawn faces from the RAET. This test was compared with the unimodal visual test in a healthy (Eördegh et al., 2019) and a psychiatric population (Pertich et al., 2020). The results from these comparisons suggest that the fact that the task had become multisensory did not have a significant effect on the participant’s performances, which led us to the conclusion that multimodality alone does not interfere with effectiveness in associative equivalence learning. However, it can influence the retrieval and generalization parts of the test phase, which can be clearly observed with the shorter response times in the SoundFace test.

The applied drawn faces as consequents in the SoundFace test are the same as the antecedents in the RAET. These visual stimuli consist semantical meanings and they are rich in features. Furthermore, the auditory stimuli are also complex, common in the daily life. Consequently, this can help make connections between the stimulus pairs, and help expressing an organizing rule verbally. These explicit properties can also involve various cortical areas to enhance the learning (Puszta et al., 2018; Puszta et al., 2020; Puszta et al., 2019).

As we demonstrated in the previous study, where we compared two visually guided paradigms with different visual stimuli set, complexity and verbalizability has indeed an effect on the associative equivalence learning, but not on retrieval and generalization. The next logical step was to investigate wether it has the same effect on a multimodal AEL or has an effect at all. For this reason, we developed a new test (SoundPolygon), combining the auditory stimuli from the SoundFace with the two-dimensional geometric shapes from the Polygon.

Comparing the two audiovisual tests, the results revealed that contrary to what was found when the two visual tests (RAET vs Polygon) were compared, in the case of the multimodal tests the reduced stimulus complexity and verbalizability influenced not only the performances in the acquisition phase, but in the entire test phase, including the retrieval and generalization. Based on these results, it seems that when the learning is unimodal (visual), the reduced stimulus complexity makes the learning difficult, but if it has been successful, retrieval and generalization are spared. Such a sparing does not seem to occur in the case of a multimodal (audiovisual) learning task. While it is not surprising (in fact, it is somehow intuitive) that stimulus complexity and verbalizability influences the effectiveness of associative learning (Arslan et al., 2020; Bender et al., 2017; Brashears and Minda, 2020), it is hard to explain why it influences the performance in all phases of the audiovisual version of the test, whereas in the visual version only the acquisition phase is concerned. We argue that the neural background of this specific paradigm might be best explained by the integrative encoding mechanism of associative learning (Bowman et al., 2021; de Araujo Sanchez & Zeithamova, 2023; Eichenbaum, 2000; Shohamy & Wagner, 2008). The idea of the mechanism is that the memory representations of overlapping events will be constructed into an integrated memory at the time of learning (Richter et al., 2016; Schapiro et al., 2017; Shohamy & Wagner, 2008; Zeithamova et al., 2012). Supposedly, this happens through the dopaminergic connections between the midbrain and the hippocampus, which involves the substantia nigra (SN)-striatum loop and the ventral tegmentum (VT)-hippocampus loop. These loops are activated simultaneously and working cooperatively. While the SN-striatum loop supports the reward-based learning, the VT-hippocampus loop transfers information to the hippocampus, which will be resonponsible for crossevent integration and encoding these memory representations neuronal networks (Balleine et al., 2007; Daniel & Pollmann, 2014; Hiebert et al., 2014; Scimeca & Badre, 2012; Shohamy & Wagner, 2008; Tulving & Markowitsch, 1998). Afterwards, this hippocampal network will be activated during retrieval and generalization in the test phase. Based on this model, it is possible that the hippocampus, which is usually considered as the key structure of explicit memory (Eichenbaum, 2000; Tulving & Markowitsch, 1998), can support implicit functions as well. This is the reason, why we argued in some of our earlier studies (Braunitzer et al., 2017; Öze et al., 2017) that children can generalize at a high level, even when the acquisition and retrieval is poor. This indicates that they have the information, and they can use it as long as no conscious effort is involved. However, it is important to note that integrative encoding is based on visual learning, and we cannot be sure that it can be applied to multimodal learning as well.

Beside the poorer perfomances, response times were also significantly longer in SoundPolygon. This is in line with results by Bender and colleagues (2017), who found that in face-name association task, the participants' reaction times were significantly shorter to more complex, colored faces expressing emotions compared to less distinctive grayscaled faces. The higher visual complexity may enhance the distinctiveness and saliency of the stimuli, and can be helpful in encoding and recognition of individual items and associations.

Based on these, one possible hypothesis can be for explaining why reduced stimulus complexity and verbalizability influences all the phases of the multimodal tests (but only the acquisition phase in the visual test), that hippocampal compensation is either specific to the visual modality or it works in unimodal learning. The other possible explanation is that in the case of the SoundFace, both the auditory stimuli (cat, human, vehicle, guitar) and visual stimuli (drawn faces) are rich in semantic meanings and can be easily described verbally; therefore, it is easier to make associative connections between them, and form a verbalized rule for the equivalence. Consequently, this can make the hippocampus more active (as the part of the declarative memory) and make the encoding more efficient (integrating the events more efficiently). On the other hand, it is much more difficult to make such a connection between the sounds and the geometric shapes, which can negatively influence not only the encoding but the recall and transfer functions, too.

6. Summary

The main aim of the present studies was to investigate the influential effect of visual stimulus complexity and verbalizability on the associative equivalence learning and the connected memory processes (retrieval and generalization) in visually guided and audiovisually guided learning paradigms.

In the case of the unimodal visual acquired equivalence learning, we found that, it has a profound effect on the acquisition phase, which is primary mediated by the basal ganglia-frontal lobe loops. The simplified visual stimuli decreased the participants' learning performances. However, it did not have any or only a weak effect on retrieval and generalization, which are primary mediated by the hippocampus-medial temporal lobe. We conclude that due to the reduced stimulus complexity and verbalizability the participants tend to shift to a more implicit strategy, with less cortical compensatory contribution. The results also indicate that contrary to the previously used Rutgers Ecquired Equivalence Test (RAET), the Polygon test could be more sensitive to measure and analyze the basal ganglia-frontal loop functions.

In the case of the audiovisual acquired equivalence learning, we found that, applying visual stimuli with reduced stimulus complexity and verbalizability, decrease the participants' performance in both phases of the audiovisual learning paradigm (i.e. acquisition phase, retrieval and generalization parts of the test phase). One explanation of it can be that the hippocampus is dominated by unimodal visual encoding and less sensitive to the audiovisual multisensory inputs. The second explanation could be that it is more difficult to make associative connections between the less verbalizable visual stimuli and the applied sounds through semantic congruency or previously learned associations.

Based on these, a question arises, whether the modality has any effect when we compare the visual and audiovisual tests with less complex and verbalizable visual stimuli? Furthermore, stimulus modality or stimulus complexity has the primary influential effect on learning and memory? We are currently investigating these questions and we are positive that the application of further associative learning tests could give answers to these questions. Hopefully, future work will also include EEG recordings and neuroimaging, which could be complete the behavioral studies, and give better insights into the neural processes.

Beyond the scope of memory research, further application is possible in clinical research. The tests could be possible tools to detect cognitive disfunctions in psychiatric and neurological disorders connected to the dysfunction of the basal ganglia and the hippocampus. These equivalence learning tests could probably be monitoring tools in neurorehabilitation, too.

7. Acknowledgements

First of all, I would like to express my deepest gratitude to my mentors and supervisors, Dr. Attila Nagy and Dr. Gabriella Eördegh for the patient guidence through my time as their student. I would like to thank for the encouragment, the advices, the answers to my scientific questions they gave me, and for helping me trough difficult situations. I have been extremely lucky to have supervisors like them.

I would also like to thank my gratitude to Professor Gábor Jancsó and Professor Ferenc Bari for giving me the chance to participate in the Neuroscience PhD program and I am also grateful to Professor Gyula Sáry for allowing me to do my research in the Department of Physiology.

I am also grateful to my colleagues in the laboratory, namely Dr. Balázs Bodosi for his help and funny jokes, Dr. Gábor Braunitzer for reviewing and correcting the articles and sharing scientific ideas, and Dr. Marietta Orosz-Ábel for her kindness.

I would especially like to thank Noémi Harcsa-Pintér and Ádám Kiss, my fellow Phd students, who were by my side every day and made my days better. I would like to thank them for the coworking, the support, the help and the conversations. I could always count on them.

Furthermore, I would like to thank Dr. Anett Rosu and Dr. András Kelemen for their professional cooperation, and all of the students who participated somehow in our research, namely Olívia Huszár, Bátor Brassó, Márton Gaszner, András Hegedűs, Anna Lazsádi, Ábel Hertelendy, Márk Horváth, Bulcsú Szűcs and Krisztián Szemcsuk.

I would like to express my deepest gratitude to my mother Edit Tot, who was always there for me when I needed her.

And finally, I would like to thank everyone who inspired or helped me in any way: colleagues, friends and students.

This PhD work was supported by grants of SZTE-SZAOK-KKA (no. 2019/270-62-2 and no. 2023/5S479), and ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the Source of the National Research, Development and Innovation Fund.

8. References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401-406. https://doi.org/10.1038/387401a0

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res*, *166*(3-4), 559-571. https://doi.org/10.1007/s00221-005-2396-5

Anderson, N. S., & Leonard, J. A. (1958). The recognition, naming, and reconstruction of visual figures as a function of contour redundancy. *J Exp Psychol*, *56*(3), 262-270. https://doi.org/10.1037/h0044893

Arslan, S., Broc, L., & Mathy, F. (2020). Lower verbalizability of visual stimuli modulates differences in estimates of working memory capacity between children with and without developmental language disorders. *Autism Dev Lang Impair*, *5*, 2396941520945519. https://doi.org/10.1177/2396941520945519

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol Rev*, *105*(3), 442-481. https://doi.org/10.1037/0033-295x.105.3.442

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends Cogn Sci*, *5*(5), 204-210. https://doi.org/10.1016/s1364-6613(00)01624-7

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu Rev Psychol*, *56*, 149-178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Ann N Y Acad Sci*, *1224*, 147-161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Astley, S. a. W. E. (1998). Novelty and functional equivalence in superordinate categorization by pigeons. *Anim. Learn. Behav.*, *26*, 125-138. https://doi.org/10.3758/BF03199206

Astley, S. L., & Wasserman, E. A. (1999). Superordinate category formation in pigeons: association with a common delay or probability of food reinforcement makes perceptually dissimilar stimuli functionally equivalent. *J Exp Psychol Anim Behav Process*, *25*(4), 415-432.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In W. S. a. J. T. S. Kenneth (Ed.), (Vol. 2, pp. 89-195). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A., Eysenck, M., & Anderson, M. (2009). *Memory*. https://doi.org/10.4324/9780429449642

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In H. B. Gordon (Ed.), (Vol. 8, pp. 47-89). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1

Bakhtiari, S. a. A. A. a. P. C. (2020). Influence of stimulus complexity on the specificity of visual perceptual learning. *Journal of Vision*, *20*. https://doi.org/10.1167/jov.20.6.13

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J Neurosci*, *27*(31), 8161-8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Curr Biol*, *25*(2), R60-R61. https://doi.org/10.1016/j.cub.2014.12.011

Bates, S. L., & Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiol Aging*, *35*(12), 2761-2769. https://doi.org/10.1016/j.neurobiolaging.2014.04.003

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the Neural Bases of Implicit and Statistical Learning. *Top Cogn Sci*, *11*(3), 482-503. https://doi.org/10.1111/tops.12420

Bender, A. R., Naveh-Benjamin, M., Amann, K., & Raz, N. (2017). The role of stimulus complexity and salience in memory for face-name associations in healthy adults: Friend or foe? *Psychol Aging*, *32*(5), 489-505. https://doi.org/10.1037/pag0000185

Bonardi, C., Rey, V., Richmond, M., & Hall, G. (1993). Acquired equivalence of cues in pigeon autoshaping: Effects of training with common consequences and with common antecedents. *Animal Learning & Behavior*, *21*(4), 369-376. https://doi.org/10.3758/BF03198003

Bowman, C. R., de Araujo Sanchez, M. A., Hou, W., Rubin, S., & Zeithamova, D. (2021). Generalization and False Memory in an Acquired Equivalence Paradigm: The Influence of Physical Resemblance Across Related Episodes. *Front Psychol*, *12*, 669481. https://doi.org/10.3389/fpsyg.2021.669481

Brashears, B. a. M. J. (2020). The Effects of Feature Verbalizablity on Category Learning [Psyarxiv, Preprint]. https://doi.org/10.31234/osf.io/gfh4a

Braunitzer, G., Őze, A., Eördegh, G., Pihokker, A., Rózsa, P., Kasik, L., . . . Nagy, A. (2017). The development of acquired equivalence from childhood to adulthood-A cross-sectional study of 265 subjects. *PLoS One*, *12*(6), e0179525. https://doi.org/10.1371/journal.pone.0179525

Bódi, N., Csibri, E., Myers, C. E., Gluck, M. A., & Kéri, S. (2009). Associative learning, acquired equivalence, and flexible generalization of knowledge in mild Alzheimer disease. *Cogn Behav Neurol*, *22*(2), 89-94. https://doi.org/10.1097/WNN.0b013e318192ccf0

Chai, X. J., Ofen, N., Jacobs, L. F., & Gabrieli, J. D. (2010). Scene complexity: influence on perception, memory, and development in the medial temporal lobe. *Front Hum Neurosci*, *4*, 21. https://doi.org/10.3389/fnhum.2010.00021

Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389-404. https://doi.org/10.1016/j.cognition.2009.10.012

Chudler, E. H., Sugiyama, K., & Dong, W. K. (1995). Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Res*, *674*(1), 33-45. https://doi.org/10.1016/0006-8993(94)01427-j

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat Neurosci*, *2*(9), 844-847. https://doi.org/10.1038/12222

Daniel, R., & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: evidence from human studies. *Neurobiol Learn Mem*, *114*, 90-100. https://doi.org/10.1016/j.nlm.2014.05.002

de Araujo Sanchez, M. A., & Zeithamova, D. (2023). Generalization and false memory in acquired equivalence. *Cognition*, *234*, 105385. https://doi.org/10.1016/j.cognition.2023.105385

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol*, *84*(6), 3072-3077. https://doi.org/10.1152/jn.2000.84.6.3072

Donderi, D. a. M. S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays*, *26*, 71-78. https://doi.org/10.1016/j.displa.2005.02.002

Donderi, D. C. (2006). Visual complexity: a review. *Psychol Bull*, *132*(1), 73-97. https://doi.org/10.1037/0033-2909.132.1.73

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci*, *1*(1), 41-50. https://doi.org/10.1038/35036213

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*(5234), 305-307. https://doi.org/10.1126/science.270.5234.305

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J Vis*, *7*(5), 7.1-14. https://doi.org/10.1167/7.5.7

Eördegh, G., Pertich, Á., Tárnok, Z., Nagy, P., Bodosi, B., Giricz, Z., . . . Nagy, A. (2020). Impairment of visually guided associative learning in children with Tourette syndrome. *PLoS One*, *15*(6), e0234724. https://doi.org/10.1371/journal.pone.0234724

Eördegh, G., Tót, K., Kelemen, A., Kiss, Á., Bodosi, B., Hegedűs, A., . . . Nagy, A. (2022). The Influence of Stimulus Complexity on the Effectiveness of Visual Associative Learning. *Neuroscience*, *487*, 26-34. https://doi.org/10.1016/j.neuroscience.2022.01.022

Eördegh, G., Őze, A., Bodosi, B., Puszta, A., Pertich, Á., Rosu, A., . . . Nagy, A. (2019). Multisensory guided associative learning in healthy humans. *PLoS One*, *14*(3), e0213094. https://doi.org/10.1371/journal.pone.0213094

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex*, *12*(10), 1031-1039. https://doi.org/10.1093/cercor/12.10.1031

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res*, *147*(3), 332-343. https://doi.org/10.1007/s00221-002-1262-y

Frens, M. A., & Van Opstal, A. J. (1998). Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res Bull*, *46*(3), 211-224. https://doi.org/10.1016/s0361-9230(98)00007-0

Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proc Natl Acad Sci U S A*, *105*(24), 8405-8409. https://doi.org/10.1073/pnas.0710674105

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*, *11*(5), 473-490. https://doi.org/10.1162/089892999563544

Gibney, K. D., Aligbe, E., Eggleston, B. A., Nunes, S. R., Kerkhoff, W. G., Dean, C. L., & Kwakye, L. D. (2017). Visual Distractors Disrupt Audiovisual Integration Regardless of Stimulus Complexity. *Front Integr Neurosci*, *11*, 1. https://doi.org/10.3389/fnint.2017.00001

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Res*, *1160*, 58-68. https://doi.org/10.1016/j.brainres.2007.03.041

Giricz, Z., Pertich, Á., Őze, A., Puszta, A., Fehér, Á., Eördegh, G., . . . Nagy, A. (2021). Visually guided associative learning in pediatric and adult migraine without aura. *Cephalalgia*, *41*(2), 176-184. https://doi.org/10.1177/0333102420958388

Godfroy-Cooper, M., Sandor, P. M., Miller, J. D., & Welch, R. B. (2015). The interaction of vision and audition in two-dimensional space. *Front Neurosci*, *9*, 311. https://doi.org/10.3389/fnins.2015.00311

Gogtay, N., Nugent, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., . . . Thompson, P. M. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus*, *16*(8), 664-672. https://doi.org/10.1002/hipo.20193

Goyos, C. (2000). Equivalence Class Formation Via Common Reinforcers Among Preschool Children. *Psychological Record*, *50*, 629-654. https://doi.org/10.1007/BF03395375

Grassi, M., & Casco, C. (2010). Audiovisual bounce-inducing effect: when sound congruence affects grouping in vision. *Atten Percept Psychophys*, *72*(2), 378-386. https://doi.org/10.3758/APP.72.2.378

Güçlütürk, Y., Güçlü, U., van Gerven, M., & van Lier, R. (2018). Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Sci Rep*, *8*(1), 3439. https://doi.org/10.1038/s41598-018-21636-y

Hall, G. (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior*, *24*(3), 233-255. https://doi.org/10.3758/BF03198973

Hall, G., Ray, E., & Bonardi, C. (1993). Acquired equivalence between cues trained with a common antecedent. *J Exp Psychol Anim Behav Process*, *19*(4), 391-399. https://doi.org/10.1037//0097-7403.19.4.391

Hannula, D. E., Minor, G. N., & Slabbekoorn, D. (2023). Conscious awareness and memory systems in the brain. *Wiley Interdiscip Rev Cogn Sci*, e1648. https://doi.org/10.1002/wcs.1648

Harding, A., Halliday, G., Caine, D., & Kril, J. (2000). Degeneration of anterior thalamic nuclei differentiates alcoholics with amnesia. *Brain*, *123 (Pt 1)*, 141-154. https://doi.org/10.1093/brain/123.1.141

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res*, *122*(2), 247-252. https://doi.org/10.1007/s002210050512

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Exp Psychol*, *62*(2), 123-130. https://doi.org/10.1027/1618-3169/a000279

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *J Exp Psychol*, *63*, 289-293. https://doi.org/10.1037/h0039516

Hiebert, N. M., Vo, A., Hampshire, A., Owen, A. M., Seergobin, K. N., & MacDonald, P. A. (2014). Striatum in stimulus-response learning via feedback and in decision making. *Neuroimage*, *101*, 448-457. https://doi.org/10.1016/j.neuroimage.2014.07.013

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *J Exp Psychol Anim Behav Process*, *15*(4), 338-346.

Iaria, G., Petrides, M., Dagher, A., Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J Neurosci*, *23*(13), 5945-5952. https://doi.org/10.1523/JNEUROSCI.23-13-05945.2003

Jitsumori, M. (2004). Categorization and formation of equivalence classes in animals: Studies in Japan on the background of contemporary developments. *Japanese Psychological Research*, *46*, 182 - 194. https://doi.org/10.1111/j.1468-5584.2004.00250.x

Joassin, F., Pesenti, M., Maurage, P., Verreckt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex*, *47*(3), 367-376. https://doi.org/10.1016/j.cortex.2010.03.003

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., Hudspeth, A. J., & (Eds.), S. M. (2014). *Principles of Neural Science* (5th ed.). McGraw Hill.

Kayaert, G., & Wagemans, J. (2009). Delayed shape matching benefits from simplicity and symmetry. *Vision Res*, *49*(7), 708-717. https://doi.org/10.1016/j.visres.2009.01.002

Kerr, J. N., & Wickens, J. R. (2001). Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *J Neurophysiol*, *85*(1), 117-124. https://doi.org/10.1152/jn.2001.85.1.117

Khalil, R., Abo Elfetoh, N., Moftah, M. Z., & Khedr, E. M. (2015). Acquired equivalence associative learning in GTC epileptic patients: experimental and computational study. *Front Cell Neurosci*, *9*, 418. https://doi.org/10.3389/fncel.2015.00418

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, *3*(1), e1532. https://doi.org/10.1371/journal.pone.0001532

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399-1402. https://doi.org/10.1126/science.273.5280.1399

Kéri, S., Nagy, O., Kelemen, O., Myers, C. E., & Gluck, M. A. (2005). Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. *Schizophr Res*, *77*(2-3), 321-328. https://doi.org/10.1016/j.schres.2005.03.024

Lanz, F., Moret, V., Rouiller, E. M., & Loquet, G. (2013). Multisensory Integration in Non-Human Primates during a Sensory-Motor Task. *Front Hum Neurosci*, *7*, 799. https://doi.org/10.3389/fnhum.2013.00799

Larsen, B., & Luna, B. (2015). In vivo evidence of neurophysiological maturation of the human adolescent striatum. *Dev Cogn Neurosci*, *12*, 74-85. https://doi.org/10.1016/j.dcn.2014.12.003

Lauzon, S. A., Abraham, A. E., Curcin, K., Butler, B. E., & Stevenson, R. A. (2022). The relationship between multisensory associative learning and multisensory integration. *Neuropsychologia*, *174*, 108336. https://doi.org/10.1016/j.neuropsychologia.2022.108336

Lech, R. K., & Suchan, B. (2013). The medial temporal lobe: memory and beyond. *Behav Brain Res*, *254*, 45-49. https://doi.org/10.1016/j.bbr.2013.06.009

Lee, H., Stirnberg, R., Stöcker, T., & Axmacher, N. (2017). Audiovisual integration supports face-name associative memory formation. *Cogn Neurosci*, *8*(4), 177-192. https://doi.org/10.1080/17588928.2017.1327426

Lee, S. M., Henson, R. N., & Lin, C. Y. (2020). Neural Correlates of Repetition Priming: A Coordinate-Based Meta-Analysis of fMRI Studies. *Front Hum Neurosci*, *14*, 565114. https://doi.org/10.3389/fnhum.2020.565114

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Dev Psychobiol*, *56*(2), 292-315. https://doi.org/10.1002/dev.21197

Love, S. A., Pollick, F. E., & Latinus, M. (2011). Cerebral correlates and statistical criteria of cross-modal face and voice integration. *Seeing Perceiving*, *24*(4), 351-367. https://doi.org/10.1163/187847511X584452

Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition*, *187*, 38-49. https://doi.org/10.1016/j.cognition.2019.02.008

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci*, *28*(5), 264-271. https://doi.org/10.1016/j.tins.2005.03.008

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behav Processes*, *66*(3), 309-332. https://doi.org/10.1016/j.beproc.2004.03.011

Mallea, J., Bustamante, J., Miguez, G., & Laborda, M. A. (2019). Classical Conditioning. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior* (pp. 1-16). Springer International Publishing. https://doi.org/10.1007/978-3-319-47829-6_1214-1

Maor, I., Shwartz-Ziv, R., Feigin, L., Elyada, Y., Sompolinsky, H., & Mizrahi, A. (2019). Neural Correlates of Learning Pure Tones or Natural Sounds in the Auditory Cortex. *Front Neural Circuits*, *13*, 82. https://doi.org/10.3389/fncir.2019.00082

Markowitsch, H. J. (1988). Diencephalic amnesia: a reorientation towards tracts? *Brain Res*, *472*(4), 351-370. https://doi.org/10.1016/0006-8993(88)91226-7

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., . . . Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Sci Rep*, *11*(1), 4831. https://doi.org/10.1038/s41598-021-84196-8

Mavrides, C. M., & Brown, D. R. (1969). Discrimination and reproduction of patterns: Feature measures and constraint redundancy as predictors. *Perception & Psychophysics*, *6*(5), 276-280. https://doi.org/10.3758/BF03210098

McDonald, R. J., & Hong, N. S. (2013). How does a specific learning and memory system in the mammalian brain gain control of behavior? *Hippocampus*, *23*(11), 1084-1102. https://doi.org/10.1002/hipo.22177

McGann, J. P. (2015). Associative learning and sensory neuroplasticity: how does it happen and what is it good for? *Learn Mem*, *22*(11), 567-576. https://doi.org/10.1101/lm.039636.115

McLaren, I. P. L. a. M. A. a. A. K. a. M. R. a. F. C. a. B. W. a. M. S. a. V. F. (2018). Mackintosh Lecture. Association and Cognition: Two Processes, One System.

Meeter, M., Shohamy, D., & Myers, C. E. (2009). Acquired equivalence changes stimulus representations. *J Exp Anal Behav*, *91*(1), 127-141. https://doi.org/10.1901/jeab.2009.91-127

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389-391. https://doi.org/10.1126/science.6867718

Meyerhoff, H. S., & Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Mem Cognit*, *44*(3), 390-402. https://doi.org/10.3758/s13421-015-0575-6

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol*, *14*(2), 247-279. https://doi.org/10.1016/0010-0285(82)90010-x

Minciacchi, D., Tassinari, G., & Antonini, A. (1987). Visual and somatosensory integration in the anterior ectosylvian cortex of the cat. *Brain Res*, *410*(1), 21-31. https://doi.org/10.1016/s0006-8993(87)80016-1

Molet, M., Miller, H., & Zentall, T. R. (2012). Acquired equivalence of cues by presentation in a common context in rats. *Anim Cogn*, *15*(1), 143-147. https://doi.org/10.1007/s10071-011-0431-4

Molet, M., Miller, H. C., & Zentall, T. R. (2011). Acquired equivalence between stimuli trained in the same context. *Psychon Bull Rev*, *18*(3), 618-623. https://doi.org/10.3758/s13423-011-0087-8

Moustafa, A. A., Keri, S., Herzallah, M. M., Myers, C. E., & Gluck, M. A. (2010). A neural model of hippocampal-striatal interactions in associative learning and transfer generalization in various neurological and psychiatric patients. *Brain Cogn*, *74*(2), 132-144. https://doi.org/10.1016/j.bandc.2010.07.013

Moustafa, A. A., Myers, C. E., & Gluck, M. A. (2009). A neurocomputational model of classical conditioning phenomena: a putative role for the hippocampal region in associative learning. *Brain Res*, *1276*, 180-195. https://doi.org/10.1016/j.brainres.2009.04.020

Murray, C. A., & Shams, L. (2023). Crossmodal interactions in human learning and memory. *Front Hum Neurosci*, *17*, 1181760. https://doi.org/10.3389/fnhum.2023.1181760

Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., . . . Schwartz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J Cogn Neurosci*, *15*(2), 185-193. https://doi.org/10.1162/089892903321208123

Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z., & Benedek, G. (2006). Multisensory integration in the basal ganglia. *Eur J Neurosci*, *24*(3), 917-924. https://doi.org/10.1111/j.1460-9568.2006.04942.x

Nagy, A., Paróczy, Z., Norita, M., & Benedek, G. (2005). Multisensory responses and receptive field properties of neurons in the substantia nigra and in the caudate nucleus. *Eur J Neurosci*, *22*(2), 419-424. https://doi.org/10.1111/j.1460-9568.2005.04211.x

Nissen, M. a. B. P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., . . . Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cereb Cortex*, *17*(1), 37-43. https://doi.org/10.1093/cercor/bhj122

Pannunzi, M., Pérez-Bellido, A., Pereda-Baños, A., López-Moliner, J., Deco, G., & Soto-Faraco, S. (2015). Deconstructing multisensory enhancement in detection. *J Neurophysiol*, *113*(6), 1800-1818. https://doi.org/10.1152/jn.00341.2014

Patching, G. R., & Quinlan, P. T. (2004). Cross-modal integration of simple auditory and visual events. *Percept Psychophys*, *66*(1), 131-140. https://doi.org/10.3758/bf03194867

Persson, J., Spreng, R. N., Turner, G., Herlitz, A., Morell, A., Stening, E., . . . Söderlund, H. (2014). Sex differences in volume and structural covariance of the anterior and posterior hippocampus. *Neuroimage*, *99*, 215-225. https://doi.org/10.1016/j.neuroimage.2014.05.038

Pertich, Á., Eördegh, G., Németh, L., Hegedüs, O., Öri, D., Puszta, A., . . . Nagy, A. (2020). Maintained Visual-, Auditory-, and Multisensory-Guided Associative Learning Functions in Children With Obsessive-Compulsive Disorder. *Front Psychiatry*, *11*, 571053. https://doi.org/10.3389/fpsyt.2020.571053

Poldrack, R. A., & Gabrieli, J. D. (2001). Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. *Brain*, *124*(Pt 1), 67-82. https://doi.org/10.1093/brain/124.1.67

Porter, J. N., Roy, A. K., Benson, B., Carlisi, C., Collins, P. F., Leibenluft, E., . . . Ernst, M. (2015). Age-related changes in the intrinsic functional connectivity of the human ventral vs. dorsal striatum from childhood to middle age. *Dev Cogn Neurosci*, *11*, 83-95. https://doi.org/10.1016/j.dcn.2014.08.011

Puszta, A., Katona, X., Bodosi, B., Pertich, Á., Nyujtó, D., Braunitzer, G., & Nagy, A. (2018). Cortical Power-Density Changes of Different Frequency Bands in Visually Guided Associative Learning: A Human EEG-Study. *Front Hum Neurosci*, *12*, 188. https://doi.org/10.3389/fnhum.2018.00188

Puszta, A., Pertich, Á., Giricz, Z., Nyujtó, D., Bodosi, B., Eördegh, G., & Nagy, A. (2020). Predicting Stimulus Modality and Working Memory Load During Visual- and Audiovisual-Acquired Equivalence Learning. *Front Hum Neurosci*, *14*, 569142. https://doi.org/10.3389/fnhum.2020.569142

Puszta, A., Pertich, Á., Katona, X., Bodosi, B., Nyujtó, D., Giricz, Z., . . . Nagy, A. (2019). Power-spectra and cross-frequency coupling changes in visual and Audio-visual acquired equivalence learning. *Sci Rep*, *9*(1), 9444. https://doi.org/10.1038/s41598-019-45978-3

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Trans N Y Acad Sci*, *24*, 574-590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D. A., Cushman, J., . . . Mehta, M. R. (2013). Multisensory control of hippocampal spatiotemporal selectivity. *Science*, *340*(6138), 1342-1346. https://doi.org/10.1126/science.1232655

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*(10), 2026-2042. https://doi.org/10.1016/j.neuropsychologia.2013.06.019

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learn Mem*, *1*(4), 217-229.

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *J Cogn Neurosci*, *10*(2), 248-263. https://doi.org/10.1162/089892998562681

Reig, R., & Silberberg, G. (2014). Multisensory integration in the mouse striatum. *Neuron*, *83*(5), 1200-1212. https://doi.org/10.1016/j.neuron.2014.07.033

Richter, F. R., Chanales, A. J. H., & Kuhl, B. A. (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *Neuroimage*, *124*(Pt A), 323-335. https://doi.org/10.1016/j.neuroimage.2015.08.051

Rosu, A., Tót, K., Godó, G., Kéri, S., Nagy, A., & Eördegh, G. (2022). Visually guided equivalence learning in borderline personality disorder. *Heliyon*, *8*(10), e10823. https://doi.org/10.1016/j.heliyon.2022.e10823

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *J Exp Psychol Hum Percept Perform*, *33*(6), 1311-1321. https://doi.org/10.1037/0096-1523.33.6.1311

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*(2), 185-195. https://doi.org/10.1016/s0896-6273(00)80448-1

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philos Trans R Soc Lond B Biol Sci*, *372*(1711). https://doi.org/10.1098/rstb.2016.0049

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Curr Opin Neurobiol*, *7*(2), 191-197. https://doi.org/10.1016/s0959-4388(97)80007-4

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241-263. https://doi.org/10.1016/s0896-6273(02)00967-4

Schwabe, L. (2013). Stress and the engagement of multiple memory systems: integration of animal and human studies. *Hippocampus*, *23*(11), 1035-1043. https://doi.org/10.1002/hipo.22175

Schwarz, M., Sontag, K. H., & Wand, P. (1984). Sensory-motor processing in substantia nigra pars reticulata in conscious cats. *J Physiol*, *347*, 129-147. https://doi.org/10.1113/jphysiol.1984.sp015057

Scimeca, J. M., & Badre, D. (2012). Striatal contributions to declarative memory retrieval. *Neuron*, *75*(3), 380-392. https://doi.org/10.1016/j.neuron.2012.07.014

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *J Neurosci*, *25*(11), 2941-2951. https://doi.org/10.1523/JNEUROSCI.3401-04.2005

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb Cortex*, *16*(11), 1546-1555. https://doi.org/10.1093/cercor/bhj092

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10), 1445-1453. https://doi.org/10.1068/p5843

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage*, *56*(4), 2200-2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, *45*(3), 561-571. https://doi.org/10.1016/j.neuropsychologia.2006.01.013

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn Sci*, *12*(11), 411-417. https://doi.org/10.1016/j.tics.2008.07.006

Shanks, D. R., & Johnstone, T. (1998). Implicit knowledge in sequential learning tasks. In S. M. A. & F. P. A. (Eds.), *Handbook of implicit learning* (pp. 533–572). Sage Publications, Inc.

Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1991). What is the role of frontal lobe damage in memory disorders? In H. S. Levin, H. M. Eisenberg, & A. L. Bento (Eds.), *Frontal lobe function and dysfunction* (pp. 173–195). Oxford University Press.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain*, *127*(Pt 4), 851-859. https://doi.org/10.1093/brain/awh100

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378-389. https://doi.org/10.1016/j.neuron.2008.09.023

Smith, E. E., Nolen-Hoeksema, S., Fredrickson, B., & Loftus, G. (2003). *Atkinson and Hilgard's Introduction to Psychology* (14th ed.). Wadsworth Publishing.

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., . . . Grace, R. C. (2012). Implicit and explicit categorization: a tale of four species. *Neurosci Biobehav Rev*, *36*(10), 2355-2369. https://doi.org/10.1016/j.neubiorev.2012.09.003

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harb Perspect Biol*, *7*(3), a021667. https://doi.org/10.1101/cshperspect.a021667

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu Rev Neurosci*, *27*, 279-306. https://doi.org/10.1146/annurev.neuro.27.070203.144130

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annu Rev Neurosci*, *34*, 259-288. https://doi.org/10.1146/annurev-neuro-061010-113720

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J Cogn Neurosci*, *8*(6), 497-506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nat Rev Neurosci*, *15*(8), 520-535. https://doi.org/10.1038/nrn3742

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory Integration and the Society for Neuroscience: Then and Now. *J Neurosci*, *40*(1), 3-11. https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Steinweg, B., & Mast, F. W. (2017). Semantic incongruity influences response caution in audio-visual integration. *Exp Brain Res*, *235*(1), 349-363. https://doi.org/10.1007/s00221-016-4796-0

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Exp Brain Res*, *194*(1), 91-102. https://doi.org/10.1007/s00221-008-1672-6

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci*, *14*(9), 400-410. https://doi.org/10.1016/j.tics.2010.06.008

Ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2016). The COGs (context, object, and goals) in multisensory processing. *Exp Brain Res*, *234*(5), 1307-1323. https://doi.org/10.1007/s00221-016-4590-z

Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Front Psychol*, *4*, 331. https://doi.org/10.3389/fpsyg.2013.00331

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford University Press.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, *8*(3), 198-204. https://doi.org/10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.0.CO;2-G

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neurosci Biobehav Rev*, *51*, 205-222. https://doi.org/10.1016/j.neubiorev.2015.01.008

Ungerleider, L. G., Doyon, J., & Karni, A. (2002). Imaging brain plasticity during motor skill learning. *Neurobiol Learn Mem*, *78*(3), 553-564. https://doi.org/10.1006/nlme.2002.4091

Urcuioli, P. (2001). Categorization & acquired equivalence. In C. R.G. (Ed.), *Avian visual cognition [online]*.

Vaughan, W. (1988). Formation of Equivalence Sets in Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 36-42. https://doi.org/10.1037/0097-7403.14.1.36

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol*, *80*(2), 1006-1010. https://doi.org/10.1152/jn.1998.80.2.1006

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Ward-Robinson, J., & Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *Q J Exp Psychol B*, *52*(4), 335-350. https://doi.org/10.1080/027249999393031

Wasserman, E. A., Kain, A. G., & O'Donoghue, E. M. (2023). Resolving the associative learning paradox by category learning in pigeons. *Curr Biol*, *33*(6), 1112-1116.e1112. https://doi.org/10.1016/j.cub.2023.01.024

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci*, *7*(6), 464-476. https://doi.org/10.1038/nrn1919

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168-179. https://doi.org/10.1016/j.neuron.2012.05.010

Zentall, T. R., Wasserman, E. A., & Urcuioli, P. J. (2014). Associative concept learning in animals. *J Exp Anal Behav*, *101*(1), 130-151. https://doi.org/10.1002/jeab.55

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition*, *196*, 104135. https://doi.org/10.1016/j.cognition.2019.104135

Öze, A., Nagy, A., Benedek, G., Bodosi, B., Kéri, S., Pálinkás, É., . . . Braunitzer, G. (2017). Acquired equivalence and related memory processes in migraine without aura. *Cephalalgia*, *37*(6), 532-540. https://doi.org/10.1177/0333102416651286

Co-author certification

I, myself as a corresponding author of the following publication(s) declare that the authors have no conflict of interest, and **Tót Kálmán** Ph.D. candidate had significant contribution to the jointly published research(es). The results discussed in his thesis were not used and not intended to be used in any other qualification process for obtaining a PhD degree.

date: Szeged, 2023. 08. 09.

J. Nagy A.

Dr Attila Nagy, corresponding author

The publication(s) relevant to the applicant's thesis:

I. Eördegh, G., Tót, K., Kelemen, A., Kiss, Á., Bodosi, B., Hegedűs, A., Lazsádi, A., Hertelendy, Á., Kéri, S., & Nagy, A. (2022). The Influence of Stimulus Complexity on the Effectiveness of Visual Associative Learning. *Neuroscience*, *487*, 26–34. https://doi.org/10.1016/j.neuroscience.2022.01.022

II. Tót, K., Eördegh, G., Kiss, Á., Kelemen, A., Braunitzer, G., Kéri, S., Bodosi, B., & Nagy, A. (2022). Visual consequent stimulus complexity affects performance in audiovisual associative learning. *Scientific reports*, *12*(1), 17793. https://doi.org/10.1038/s41598-022-22880-z

NEUROSCIENCE
RESEARCH ARTICLE
G. Eördegh et al. / Neuroscience 487 (2022) 26–34

The Influence of Stimulus Complexity on the Effectiveness of Visual Associative Learning

Gabriella Eördegh,[a] Kálmán Tót,[b] András Kelemen,[c] Ádám Kiss,[b] Balázs Bodosi,[b] András Hegedűs,[b] Anna Lazsádi,[b] Ábel Hertelendy,[b] Szabolcs Kéri[b] and Attila Nagy[b]*

[a] Faculty of Health Sciences and Social Studies, University of Szeged, Szeged, Hungary

[b] Department of Physiology, Faculty of Medicine, University of Szeged, Szeged, Hungary

[c] Department of Applied Informatics, University of Szeged, Szeged, Hungary

Abstract—Visually guided equivalence learning is a special type of associative learning, which can be evaluated using the Rutgers Acquired Equivalence Test (RAET) among other tests. RAET applies complex stimuli (faces and colored fish) between which the test subjects build associations. The complexity of these stimuli offers the test subject several clues that might ease association learning. To reduce the number of such clues, we developed an equivalence learning test (Polygon), which is structured as RAET but uses simple grayscale geometric shapes instead of faces and colored fish. In this study, we compared the psychophysical performances of the same healthy volunteers in both RAET and Polygon test. Equivalence learning, which is a basal ganglia-associated form of learning, appears to be strongly influenced by the complexity of the visual stimuli. The simple geometric shapes were associated with poor performance as compared to faces and fish. However, the difference in stimulus complexity did not affect performance in the retrieval and transfer parts of the test phase, which are assumed to be mediated by the hippocampi.

Key words: visual, equivalence learning, basal ganglia, hippocampi, psychophysics, healthy human.

INTRODUCTION

Associative learning is an evolutionarily ancient basic cognitive function in which discrete and even different ideas and perceptions are linked together and thus can elicit similar behavioral responses. Typical forms of this learning include classical conditioning (Ito et al., 2008), probabilistic learning (Shohamy et al., 2009), weather prediction (Gluck et al., 2002), latent inhibition (Weiss and Brown, 1974), and sensory preconditioning (Rescorla, 1980). Visually guided equivalence learning is a special type of associative learning, which can be evaluated using the Rutgers Acquired Equivalence Test (RAET) among other tests. In this simple test with well-established neural background (Shohamy and Wagner, 2008), the subject learns that two or more stimuli are equivalent in terms of being mapped onto the same outcomes or responses (Myers et al., 2003). Basically, the paradigm consists of two primary components, acquisition and test phases. In the acquisition phase, the subjects learn to associate pairs of visual stimuli (cartoon faces and colored fish) on the basis of computer feedback on the correctness of the associations (trial-and-error learning). The acquisition phase consists of parts characterized by low working memory load (shaping and the equivalence training) and high working memory load (introduction of new consequents, see Table 1, Puszta et al., 2020).

In the test phase, where no feedback is given on the correctness of the responses, the previously learned (retrieval part of the test phase) or hitherto not shown, but predictable associations are tested (generalization or transfer parts of the test phase). The original paper by Myers et al. (2003) reported that performance in the acquisition phase was affected in Parkinson's disease, and the generalization part of the test phase was affected in hippocampal atrophy. Subsequent psychophysical and neuroimaging studies (Cohen et al., 1999; Gogtay et al., 2006; Persson et al., 2014; Larsen and Luna, 2015; Porter et al., 2015) demonstrated that the equivalence learning phase is linked primarily to the basal ganglia-frontal cortex loops and the test phase is linked to the hippocampi. These observations allow the conclusion that the striatum and hippocampi are structures of key importance for association and generalization, respectively, which is in line with our knowledge on the memory functions of these structures (Cohen et al., 1999; Packard and Knowlton, 2002). Not surprisingly, in several other neurological and psychiatric disorders characterized by the dysfunction of the basal ganglia and hippocampi, per-

*Corresponding author. Address: H-6720 Szeged, Dóm tér 10., Hungary.
E-mail address: nagy.attila.1@med.u-szeged.hu (A. Nagy).

https://doi.org/10.1016/j.neuroscience.2022.01.022
0306-4522/

Table 1. A summary of the visual associative learning paradigms. A,B: antecedents (faces in RAET and circles in Polygon), X,Y: consequents (fish in RAET and simple geometric forms in Polygon)

ACQUISITION			TEST	
Shaping	Equivalence training	New consequents	Retrieval	Generalization
A1 → X1	A1 → X1	A1 → X1	A1 → X1	
	A2 → X1	A2 → X1	A2 → X1	
		A1 → X2	A1 → X2	
				A2 → X2
B1 → Y1	B1 → Y1	B1 → Y1	B1 → Y1	
	B2 → Y1	B2 → Y1	B2 → Y1	
		B1 → Y2	B1 → Y2	
				B2 → Y2

formance deficit was found. Such disorders include Parkinson's disease (Myers et al., 2003), adult migraine (Oze et al., 2017), the Tourette syndrome (Eördegh et al., 2020), hippocampal atrophy (Moustafa et al., 2010), and schizophrenia (Keri et al., 2005). In contrast, no deficit was found in children with obsessive–compulsive disorder (Pertich et al., 2020) and migraine (Giricz et al., 2021). However, it is possible that in the latter two cases, it is only that RAET is not sensitive enough to detect the difference. We argue that in RAET, which applies potentially meaningful and colored stimuli, the stimuli might also elicit emotional responses, which could serve as clues to help association learning. To address this issue, we developed a visual associative learning test (Polygon), which is based on the same principles as RAET, but it uses simple grayscale geometric shapes instead of faces and fish. In the present study, we examined the performance of healthy volunteers in both tests.

Previous studies demonstrated that stimulus complexity could affect auditory-guided associative learning, and more complex stimuli could elicit better responses with more prominent cortical activation (Gucluturk et al., 2018; Staib and Bach, 2018; Maor et al., 2020). Based on these findings, we explored whether the complexity of visual stimuli could influence performance in visually guided equivalence learning. Our hypothesis was that our subjects' performance would be inferior in Polygon as compared to RAET because the stimuli in Polygon contain fewer clues. Special attention was focused on whether these possible differences are similar or different in the acquisition and test phases of the learning paradigms.

EXPERIMENTAL PROCEDURES

Subjects

Fifty-five healthy adults participated in this study (26 women and 29 men, mean age: 35.11 ± 13.925 years, range: 18–65 years). The participants were recruited on a voluntary basis, received no compensation for their participation, and they were free to quit at any time without any consequence (one subject did so). The volunteers were informed about the aims and procedures of the study, and their medical history was taken with emphasis on any neurological or psychiatric disorders. Volunteers with such disorders in their history were not eligible for the study. The volunteers were also tested with the Ishihara plates to exclude color blindness. Those who decided to participate, signed an informed consent form. The study protocol followed the tenets of the Declaration of Helsinki in all respects, and it was approved by the Regional Research Ethics Committee for Medical Research at the University of Szeged, Hungary (27/2020-SZTE).

Visually guided associative learning paradigms

Tests were run on two laptops (Lenovo Think Book 15-IIL). The subjects were tested in a quiet room sitting at a standard distance (57 cm) from the laptop screen (stimuli were equal in size, of a maximum diameter of 5 cm, which corresponds to a 5° angle of view). The subjects were tested separately, one subject at a time. No time limit was set, and no forced quick responses were expected. The keys X and M were labeled as "left" and "right" on the laptop's keyboard. The subjects used these keys to indicate their choices in both test paradigms.

In this study, we applied two visually guided associative learning paradigms: RAET and Polygon.

RAET was carried out according to Myers and co-workers (Myers et al., 2003). The testing software (originally written for iOS) was used and rewritten in Assembly (for Windows) with the written permission of Myers and colleagues at Rutgers University, NJ (Oze et al., 2017).

The antecedent stimuli were cartoon faces of a woman (A1), a girl (A2), a man (B1) and a boy (B2) with black or brown hair. The consequent stimuli were yellow (X1), red (X2), green (Y1) and blue (Y2) fish. The shape of the fishes were the same, they differed only in their color.

During a trial, the participant was presented with an antecedent (a face) and two consequents (a pair of fish of different color) and asked to choose one of the latter by pressing either the "left" or "right" button on the keyboard (Fig. 1).

The trials were organized into two main phases: acquisition and test. The test phase was further broken down to retrieval and generalization (see below). In the acquisition phase, the choice was followed by feedback on the correctness of the choice (trial and error learning) and there was no feedback in the test phase.

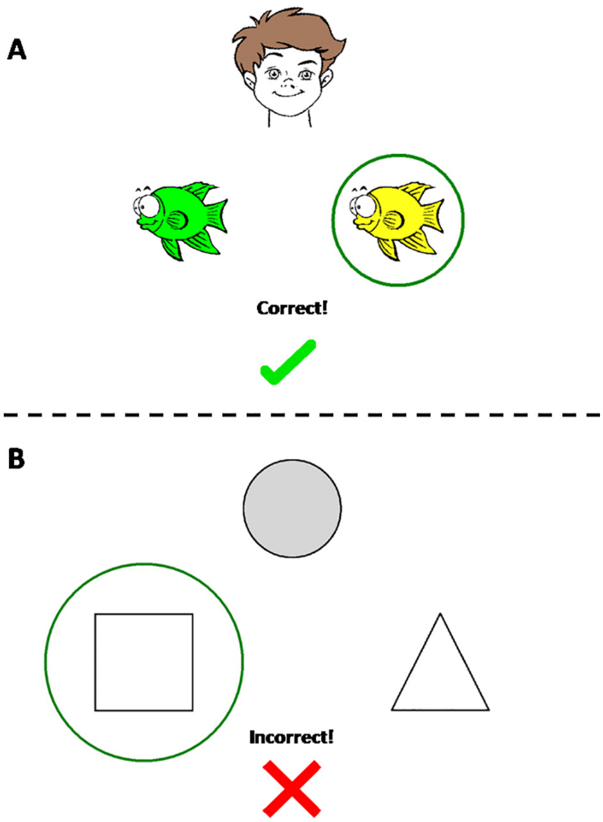

Fig. 1. A trial in the acquisition phase of RAET **(A)** and Polygon **(B).** Above is the antecedent and below are the possible consequents. By pressing the "left" or "right" button, the subject guesses which consequent belongs to the given antecedent. Immediate visual feedback is given. If the guess is right, a green checkmark appears. If the guess is wrong, it is indicated by a red X mark. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

During the acquisition phase, the participants learned a series of antecedent-consequent pairs via trial and error. When face A1 or face A2 were shown, the correct choice was fish X1 over fish Y1; however, when face B1 or face B2 appeared on the screen, the correct answer was fish Y1, instead of fish X1 (Table 1, Fig. 2). Visual feedback on the correctness of the subject's choice was always given immediately in the form of a checkmark (in green) or an X mark (in red) displayed on the screen under the actual antecedent-consequent pair. This way, besides the face-fish associations, the participants also learned that the face A1 is equivalent to face A2 in terms of their relation to the consequents (fish). New associations were introduced gradually, and they were presented mixed with trials of previously learned associations. The subjects had to achieve a certain number of consecutive correct answers after the presentation of each new association to be allowed to proceed. This number was 2 when the first association was presented, and it was increased by 2 upon the presentation of each association that followed (up to a maximum of 12). Thus, the length of the acquisition phase varied among the participants, depending upon how efficiently they learned. Altogether six of the eight possible associations were presented in the acquisition phase (each of the 4 faces associated with 2 fish). The rule of association, i.e. which stimulus feature (age, sex or hair color) is used to link the antecedent pairs, was generated randomly by the software for each subject, and it remained the same until the end of the test. The subjects were not aware of the rule of association at the beginning of testing, they had to figure it out for themselves through trial and error.

If the acquisition was successful, subjects continued with the test phase. In the test phase, the task remained the same, but visual feedback was no longer provided. In this phase, the subjects had to recall the previously learned six associations (retrieval part) and they had to identify the two new, hitherto not presented but predictable associations (generalization part). In contrast to the acquisition phase, the test phase always involved 48 trials (12 new and 36 previously learned associations). Subjects were not informed that new associations would have to be formed, only that their task remained the same, but without feedback.

Polygon is a modified version of RAET with simple geometric shapes as stimuli (Table 1, Fig. 1B, and Fig. 3).

We applied simple geometric shapes to reduce the chance that the stimuli evoke emotional responses or cognitive associations that could serve as clues for associative learning. Instead of faces, we applied circles with different contrasts (white, light gray, dark gray, and black) as antecedents. Instead of fish with different colors, we applied simple geometric shapes (triangle, square, rhombus, and concave deltoid) with no coloring as consequents (Fig. 3).

The subjects completed both equivalence learning tests one after the other, and in random order to avoid carryover.

Data analysis

We analyzed the number of trials required for completing the acquisition phase (NAT), response accuracy for the various stages of the paradigms (error ratios), and reaction times (RTs, the time between the appearance of the stimuli and the decision of the participant as indicated by pressing one of the designated keys on the keyboard). Error ratios were calculated in Microsoft Excel (2016) with a custom-made script by dividing the number of incorrect answers by the total number of trials in the acquisition phase (acquisition learning error ratio = ALER), the retrieval (retrieval error ratio = RER), and generalization parts (generalization error ratio = GER) of the test phase. RT was measured with millisecond accuracy. RT values of >3 SD were excluded from further analysis.

Statistical analysis was conducted in Statistica 13.4.0.14 (TIBCO Software Inc., USA). Data distributions were evaluated using the Shapiro–Wilk normality test. As the data sets were non-normally distributed, the Wilcoxon matched-pairs test was used for the comparisons between the paradigms. We also performed effect size and power calculations for the significant differences in G*Power 3.1.9.2 (Düsseldorf, Germany).

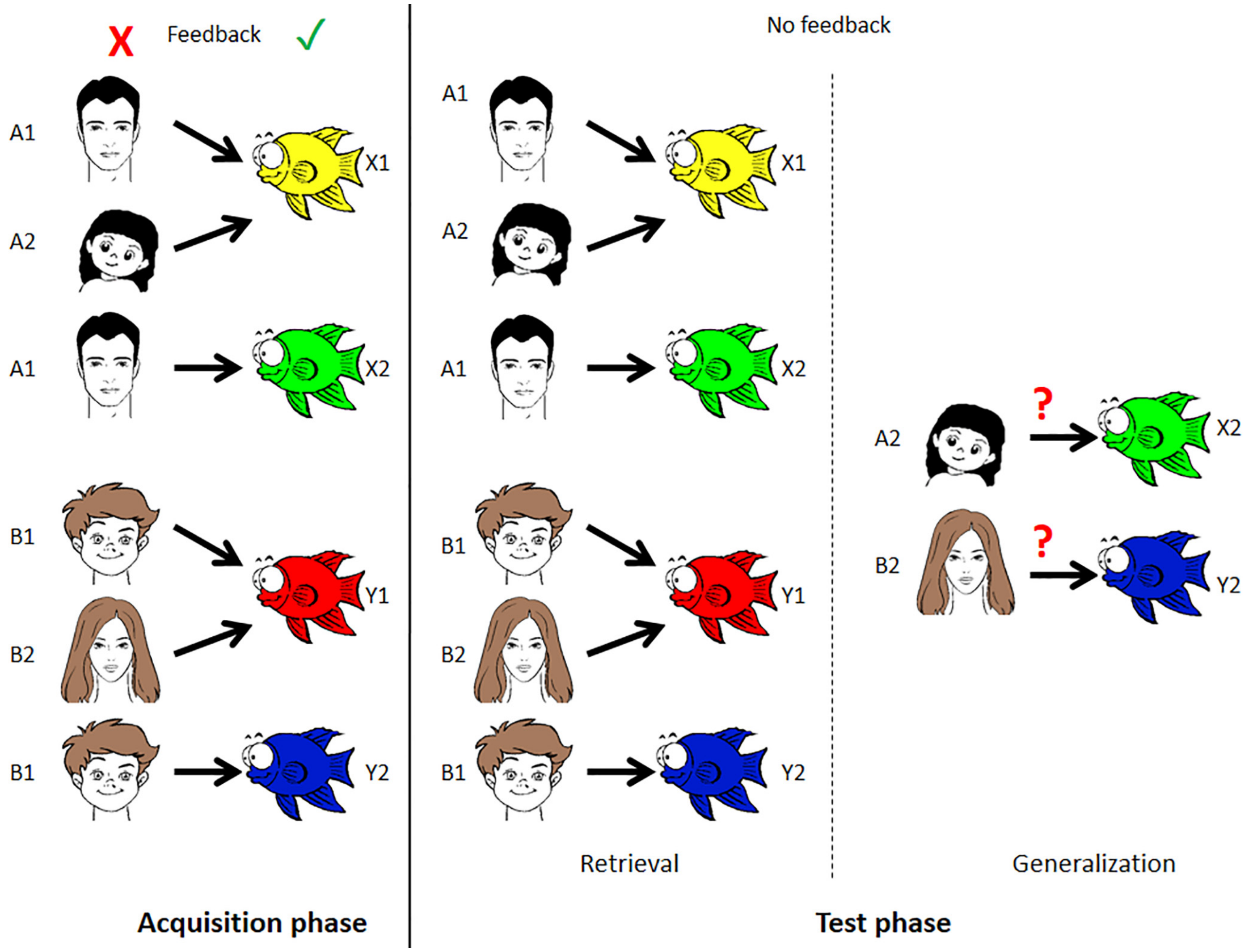

Fig. 2. Overview structure of the original Rutgers Acquired Equivalence Test. The antecedent–consequent pairs of the test. The antecedents were cartoon faces of a man (A1) a girl (A2), a boy (B1) and woman (1). The consequents (responses) were drawings of fish of yellow (X1), green (X2), red (Y1) and blue (Y2) colors. In this example, the basis of equivalence is hair color.

Data availability

The datasets are available from the corresponding author on reasonable request.

RESULTS

Fifty-four of the 55 subjects completed both paradigms. One subject could not complete Polygon. The data of this subject were not included in the analyses.

As a preliminary analysis, we assessed if the order of the administration of the paradigms (RAET-Polygon $n = 26$ or Polygon-RAET $n = 28$) had any effect on the subjects' performance. The subjects were divided into two groups based on the order in which they completed the paradigms, and we compared their performance according to the already described parameters (NAT, ALER RER, GER, RT, see before in the Experimental Procedures) for both paradigms. For this analysis, the Mann–Whitney *U* test was used. The analysis found no significant difference in any of the parameters in either paradigm (Mann–Whitney *U* test, $p > 0.05$). The temporal evolution of the psychophysical performances (NAT, ALER RER, GER, RT) were not affected to a considerable degree by the order of administration either (Supplementary Figs. 1–4). It was also tested for both paradigms if completing a paradigm as first or second in the testing sequence influenced the subjects' performance. No significant difference was detected (Mann–Whitney *U* test, $p > 0.05$), that is, whether the same paradigm was administered as first or second had no or negligible effect.

Acquisition phase

The median of the number of trials required for completing the acquisition phase (NAT) in RAET was 54.0 (range: 43.0–136.0), but it was 68.5 (range: 42.0–213.0) in Polygon. In Polygon, significantly more trials were required for the learning of the associations than in RAET (Wilcoxon matched-pairs test $Z = 3.731$, $p = 0.0002$, effect size 0.6268, power 0.9538, Fig. 4). In RAET, the median of acquisition learning error ratio

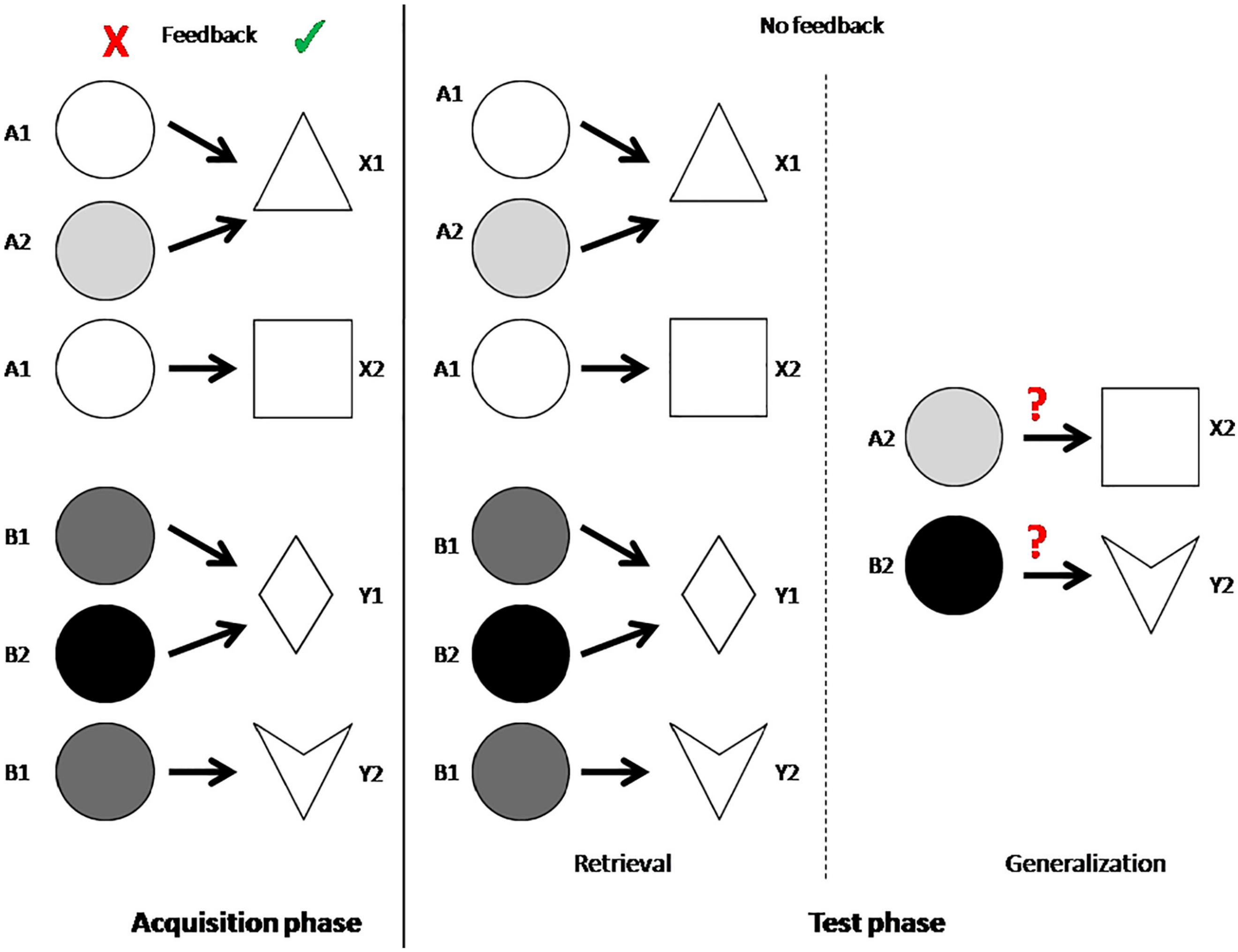

Fig. 3. Overview structure of the new visual acquired equivalence paradigm (Polygon). The antecedent–consequent pairs of the test. The antecedents were circles filled with different shades of the grayscale: white (A1), light gray (A2), dark gray (B1) and black (1). The consequents (responses) were different polygons: a triangle (X1), square (X2), rhombus (Y1) and concave deltoid (Y2). In this example, the basis of equivalence is low contrast difference.

(ALER) was 0.052 (range: 0.0–0.240), and in the Polygon test, it was 0.096 (range: 0.0–0.340). This difference was also highly significant (Wilcoxon matched-pairs test $Z = 3.939$, $p = 0.00008$, effect size 0.6984, power 0.9520, Fig. 4).

The comparison of the low and high working memory load parts of the acquisition phase (see Table 1) revealed that the performances in both the low- and high-load parts were significantly superior in RAET. The median value of NAT in the equivalence training part (low working memory load) was 27.0 trials (range: 22.0–76.0) in RAET, and 33.0 trials (range: 22.0–72.0) in Polygon (Wilcoxon matched-pairs test: $p = 0.0019$). The median value of NAT in the high working memory load part (introduction of new consequents) was 26.5 (range: 22.0–91.0) in RAET and 32.5 (range: 22.0–157.0) in Polygon (Wilcoxon matched-pairs test: $p = 0.0091$).

ALERs were also lower in both the low working memory load and the high working memory load parts of the acquisition phase in RAET. The median value of the ALER the equivalence training part (low working memory load) was 0.07275 (range: 0.0–0.2933) in RAET and 0.12702 (range: 0.0–0.4688) Polygon (Wilcoxon matched-pairs test: $p = 0.0013$). The median value of the ALER in the high working memory load part (introduction of new consequents) was 0.04348 (range: 0.0–0.1778) in RAET and 0.06155 (range: 0.0–0.3013) in Polygon (Wilcoxon matched-pairs test: $p = 0.0648$).

Reaction times (RTs) were also significantly longer in Polygon (Wilcoxon matched-pairs test $Z = 2.983$, $p = 0.003$, effect size 0.4862, power 0.9521). The median RT in RAET was 1606.22 ms (range: 1004.74–3 052.88 ms), and 1802.06 ms in Polygon (range: 888.4 9–4618.79 ms).

Test phase

In contrast to the acquisition phase, no significant performance differences were found between the paradigms either in the retrieval (Wilcoxon matched-pairs test $Z = 0.739$, $p = 0.460$) or the generalization parts of the test phase (Wilcoxon matched-pairs test

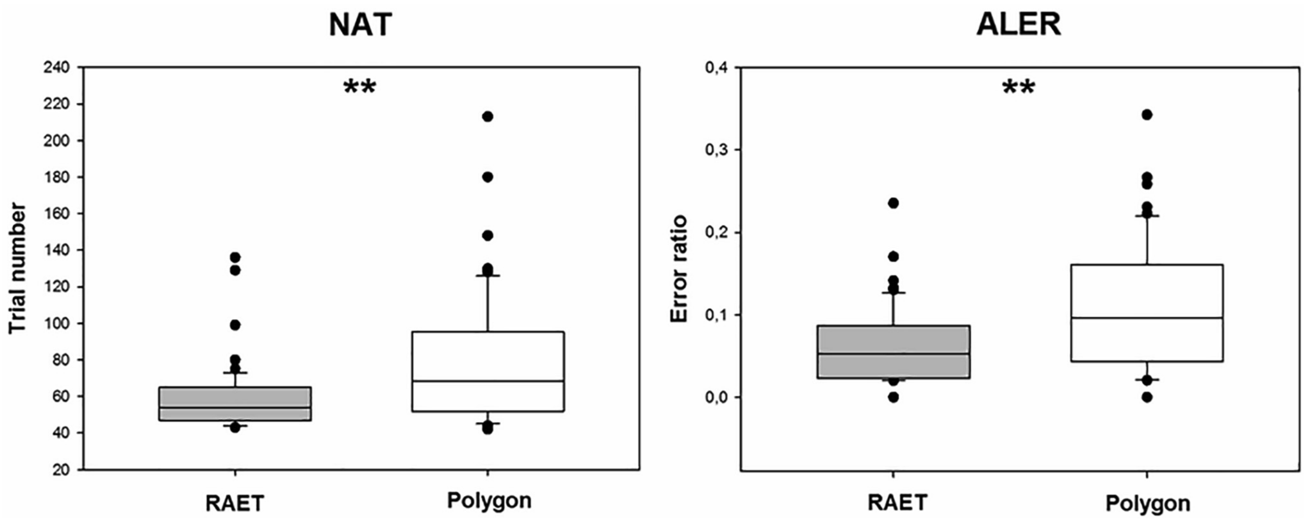

Fig. 4. Comparison of performances in the acquisition phase of the two associative learning paradigms. NAT: number of acquisition trials. ALER: acquisition learning error ratio. The lower margin of the boxes marks the 25th percentile; the line within the boxes indicates the median; and the upper margin indicates the 75th percentile. The whiskers above and below the boxes indicate the 90th and 10th percentiles, respectively. The dots above and below the whiskers represent extreme outliers. **Indicates a highly significant difference ($p < 0.01$).

$Z = 1.624$, $p = 0.104$) (Fig. 5). In RAET, the median retrieval error ratio (RER) was 0.028 (range: 0.00–0.31), and in Polygon, it was 0.00 (range: 0.00–0.42). The median generalization error ratio (GER) in RAET was 0.083 (range: 0.00–1.00), and it was 0.00 (range: 0.00–1.00) in Polygon.

RTs did not differ significantly between the two paradigms either in the retrieval (Wilcoxon matched-pairs test $Z = 0.667$, $p = 0.505$) or the generalization (Wilcoxon matched-pairs test $Z = 0.595$, $p = 0.552$) parts of the test phase. The median RT in the retrieval part in RAET was 1840.71 ms (range: 1156.26–4046.87 ms), and 1763.73 ms in Polygon (range: 934.56–4036.71 ms). The median RT in the generalization part of RAET was 2127.58 ms (range: 1300.67–13075.50 ms), and it was 2450.20 ms in Polygon (range: 1048.25–8230.50 ms).

DISCUSSION

The Rutgers Acquired Equivalence Test (RAET) was developed originally to dissociate the different phases of the visually guided associative learning of neurological patients with hippocampal and basal ganglia dysfunctions (Myers et al., 2003). Performance in the equivalence learning phase of the test provides information about the function of the frontostriatal loops, while performance in the test phase is assumed to rely on the hippocampi (Myers et al., 2003; Moustafa et al., 2009, 2010).

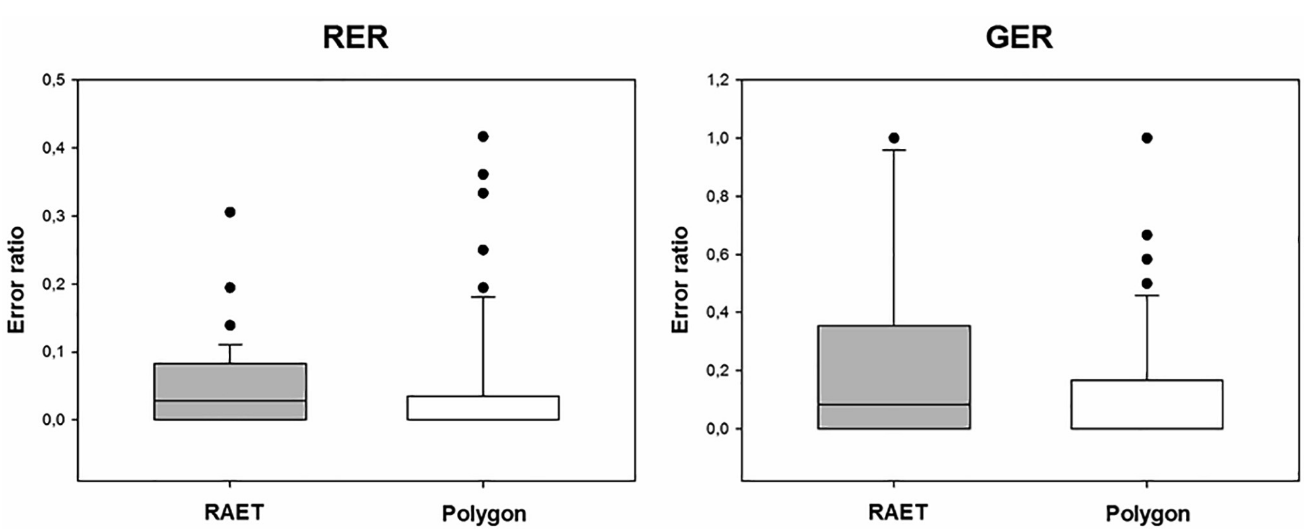

Fig. 5. Comparison of performances in the test phase of the two associative learning paradigms. RER: retrieval error ratio. GER: generalization error ratio. Otherwise, the conventions are the same as in Fig. 3.

The stimuli in the original version of RAET are human cartoon faces and colored fish. These are complex stimuli that contain various cues that can enhance association learning performance. In fact, having completed the tests, several participants were able even to verbalize the rule of association. During our long experience with RAET, we have never encountered a healthy volunteer who could not complete the test (Braunitzer et al., 2017; Eordegh et al., 2019). As for non-healthy populations, altered performance has been reported in certain conditions, but not in others (Myers et al., 2003; Keri et al., 2005; Bodi et al., 2009; Eordegh et al., 2020; Pertich et al., 2020; Giricz et al., 2021). Therefore, the sensitivity of this test is questionable in detecting mild changes in psychophysical performance. Having recognized this, we aimed to improve the sensitivity of this test, and developed a new visually guided equivalence learning test (Polygon). The new test is based on the principles of RAET, but it applies simple geometric shapes as stimuli instead of cartoon faces and fish. By this change, we sought to reduce the number of stimulus features to a minimum to avoid emotional responses or cognitive associations that could serve as extra cues for equivalence learning. To our knowledge, this is the first study to compare the visually guided equivalence learning performance of healthy volunteers across paradigms with stimuli of different complexity.

The basic structure of the two tests is very similar but there are some remarkable differences between them. In RAET, the antecedent stimuli are faces, each of which has three features (sex, age, and hair color). In contrast, antecedent stimuli in Polygon have only one feature, their shading (grayscale). Furthermore, the consequent stimuli in the RAET have only one distinctive feature, their color. In Polygon, the consequents have more than one distinctive features (such as the number of angles, sides or whether the stimulus points upward or downward). The primary aim during the construction of the stimuli in Polygon was to reduce the supporting information (emotional content, sematic content, and color information) to reduce the cortical contribution (Puszta et al., 2018; 2019) to the tasks. Additionally the more difficult task could be suitable to detect such weak differences, which were not detectable with the original RAET (Pertich et al., 2020; Giricz et al., 2021). It is true that in this respect, RAET and Polygon are not fully identical, but the modification of stimulus features allowed us to make the test more difficult (and sensitive) while keeping the original structure and logic. This way, the same testing paradigm is applied, but with stimuli that enable the detection of finer performance differences. Naturally, it would be possible to make the two tests even more similar, for instance, by using another set of consequent stimuli that would differ only in one feature.

As a preliminary analysis, we examined whether the order of the administration of the two tests was a significant factor of performance on the tests. We found no significant effect, which means that, at least in this study, the effects of learning, practice and fatigue can be excluded as confounding factors.

Our results showed that equivalence learning, which is linked primarily to the basal ganglia, is strongly influenced by the applied visual stimuli. The performance of the subjects was significantly weaker in the acquisition phase of Polygon than in the same phase of RAET, as indicated by the significantly higher error ratios and number of required acquisition trials. Based on working memory load, the acquisition phase can be divided into a low-load and a high-load part (Puszta et al., 2020). Since working memory load could influence the effectiveness of implicit learning (Collins and Frank, 2012), we also compared the subjects' performance in these two parts. The comparisons revealed that the performances in both the low- and high-load load parts of the acquisition phase were significantly better in RAET than in Polygon. In a recently published study (Eordegh et al., 2019), we administered a unimodal visual, a unimodal auditory, and an audiovisual version of the RAET paradigm to healthy subjects. We found no significant difference between the acquisition learning error ratios across the paradigms, which suggest that stimulus modality has no significant influence on performance in the acquisition phase. A possible explanation is that this type of feedback-based pair learning is a very old and conserved function, which is so simple that the different modalities contribute to the associative learning equally, and thus, the multisensory information has no priority in these learning processes. This is consistent with earlier findings that the basal ganglia, which predominate the acquisition phase of the associative learning test, are more active when rare stimulus associations appear, and this is not affected by stimulus modality (Amso et al., 2005). However, stimulus complexity and salience appear to be important determinants of learning effectiveness in sensory-guided equivalence learning. This is in agreement with the findings of previous studies, which demonstrated that stimulus complexity could influence auditory-guided associative learning, and more complex stimuli could elicit more accurate responses, better performance, and more prominent cortical activation (Brown and Proulx, 2013).

Reaction times in the acquisition phase were also significantly longer in Polygon, which shows that stimulus complexity affected this parameter as well. However, the longer reaction times did not result in performance improvement.

As for the error ratios in the retrieval and transfer (generalization) parts of the test phase, these did not differ between RAET and Polygon and neither did reaction times. In other words, stimulus complexity had no effect whatsoever on the retrieval of the previously learned associations and the transfer of the acquired rule of association to previously unseen stimulus pairs.

In summary, our results suggest that stimulus complexity can have a considerable influence on equivalence learning in healthy humans, but it has no or only a very weak effect on the connected memory processes (retrieval and transfer). The fact that the subjects made more mistakes in the acquisition phase of the Polygon can indicate that Polygon is a more sensitive test in healthy adults than the original RAET. If this is true in basal ganglia-related neurological/ psychiatric disorders as well, then Polygon could be a tool to detect fine learning alterations in such conditions that RAET is not capable of detecting. Upcoming studies should test this hypothesis.

CONTRIBUTIONS

A.N., Sz.K., B.B. and G.E. conceived the study conception and design. Data collection was made by K. T., A.H., A.L., Á.K., Á.H., and G.E.. Data analysis were performed by K.T., G.E., Á.K. and A.K. The manuscript was written by A.N., K.T. and G.E.. All authors discussed data analysis and interpretation. Funding acquisition was made by A.N. All authors reviewed/ edited the manuscript and approved the final version.

ACKNOWLEDGMENTS

The authors thank Réka Demeter, Mercédesz Kovács, and Márton Gaszner for their help in conducting the investigation and data collection. This work was supported by a grant from SZTE ÁOK-KKA Grant No. 2019/270-62-2. KT was supported by EFOP 3.6.3-VEKOP-16-2017-00009 grant. KT and ÁK were supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the Source of the National Research, Development and Innovation Fund. GE was supported by EFOP-3.6.1-16-2016-00008 grant.

REFERENCES

Amso D, Davidson MC, Johnson SP, Glover G, Casey BJ (2005) Contributions of the hippocampus and the striatum to simple association and frequency-based learning. Neuroimage 27:291–298.

Bodi N, Csibri E, Myers CE, Gluck MA, Keri S (2009) Associative learning, acquired equivalence, and flexible generalization of knowledge in mild Alzheimer disease. Cogn Behav Neurol 22:89–94.

Braunitzer G, Oze A, Eördegh G, Pihokker A, Rózsa P, Kasik L, Kéri S, Nagy A, Ito E (2017) The development of acquired equivalence from childhood to adulthood-A cross-sectional study of 265 subjects. PLoS One 12 e0179525.

Brown DJ, Proulx MJ (2013) Increased Signal Complexity Improves the Breadth of Generalization in Auditory Perceptual Learning. Neural Plast 2013 879047.

Cohen NJ, Ryan J, Hunt C, Romine L, Wszalek T, Nash C (1999) Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. Hippocampus 9:83–98.

Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 35:1024–1035.

Eordegh G, Oze A, Bodosi B, Puszta A, Pertich A, Rosu A, Godó G, Nagy A (2019) Multisensory guided associative learning in healthy humans. PLoS One 14 e0213094.

Eordegh G, Pertich A, Tárnok Z, Nagy P, Bodosi B, Giricz Z, Hegedűs O, Merkl D, Nyujtó D, Oláh S, Oze A, Vidomusz R, Nagy A, Kavushansky A (2020) Impairment of visually guided associative learning in children with Tourette syndrome. PLoS One 15 e0234724.

Giricz Z, Pertich Á, Oze A, Puszta A, Fehér Á, Eördegh G, Kóbor J, Bihari K, Pálinkás É, Braunitzer G, Nagy A (2021) Visually guided associative learning in pediatric and adult migraine without aura. Cephalalgia 41:176–184.

Gluck MA, Shohamy D, Myers C (2002) How do people solve the ‘‘weather prediction” task?: Individual variability in strategies for probabilistic category learning. Learn Memory 9:408–418.

Gogtay N, Nugent TF, Herman DH, Ordonez A, Greenstein D, Hayashi KM, Clasen L, Toga AW, Giedd JN, Rapoport JL, Thompson PM (2006) Dynamic mapping of normal human hippocampal development. Hippocampus 16:664–672.

Gucluturk Y, Guclu U, van Gerven M, van Lier R (2018) Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. Sci Rep 8:3439.

Ito R, Robbins TW, Pennartz CM, Everitt BJ (2008) Functional interaction between the hippocampus and nucleus accumbens shell is necessary for the acquisition of appetitive spatial context conditioning. J Neurosci 28:6950–6959.

Keri S, Nagy O, Kelemen O, Myers CE, Gluck MA (2005) Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. Schizophr Res 77:321–328.

Larsen B, Luna B (2015) In vivo evidence of neurophysiological maturation of the human adolescent striatum. Dev Cogn Neurosci 12:74–85.

Maor I, Shwartz-Ziv R, Feigin L, Elyada Y, Sompolinsky H, Mizrahi A (2020) Neural correlates of learning pure tones or natural sounds in the auditory cortex. Front Neural Circuit 13:82.

Moustafa AA, Keri S, Herzallah MM, Myers CE, Gluck MA (2010) A neural model of hippocampal-striatal interactions in associative learning and transfer generalization in various neurological and psychiatric patients. Brain Cogn 74:132–144.

Moustafa AA, Myers CE, Gluck MA (2009) A neurocomputational model of classical conditioning phenomena: a putative role for the hippocampal region in associative learning. Brain Res 1276:180–195.

Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, et al. (2003), Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J Cogn Neurosci 15:185-193.

Oze A, Nagy A, Benedek G, Bodosi B, Kéri S, Pálinkás E, Bihari K, Braunitzer G (2017) Acquired equivalence and related memory processes in migraine without aura. Cephalalgia 37:532–540.

Packard MG, Knowlton BJ (2002) Learning and memory functions of the Basal Ganglia. Annu Rev Neurosci 25:563–593.

Persson J, Spreng RN, Turner G, Herlitz A, Morell A, Stening E, Wahlund L-O, Wikström J, Söderlund H (2014) Sex differences in volume and structural covariance of the anterior and posterior hippocampus. NeuroImage 99:215–225.

Pertich Á, Eördegh G, Németh L, Hegedüs O, Öri D, Puszta A, Nagy P, Kéri S, Nagy A (2020) Maintained visual-, auditory-, and multisensory-guided associative learning functions in children with obsessive-compulsive disorder. Front Psychiatry 11 571053.

Porter JN, Roy AK, Benson B, Carlisi C, Collins PF, Leibenluft E, Pine DS, Luciana M, Ernst M (2015) Age-related changes in the intrinsic functional connectivity of the human ventral vs. dorsal striatum from childhood to middle age. Dev Cogn Neurosci 11:83–95.

Puszta A, Katona X, Bodosi B, Pertich Á, Nyujtó D, Braunitzer G, Nagy A (2018) Cortical power-density changes of different frequency bands in visually guided associative learning: a human EEG-study. Front Hum Neurosci 12:188.

Puszta A, Pertich Á, Katona X, Bodosi B, Nyujtó D, Giricz Z, Eördegh G, Nagy A (2019) Power-spectra and cross-frequency coupling changes in visual and Audio-visual acquired equivalence learning. Sci Rep 9:9444.

Puszta A, Pertich Á, Giricz Z, Nyujtó D, Bodosi B, Eördegh G, Nagy A (2020) Predicting stimulus modality and working memory load during visual- and audiovisual-acquired equivalence learning. Front Hum Neurosci 14 569142.

Rescorla RA (1980) Simultaneous and successive associations in sensory preconditioning. J Exp Psychol Anim Behav Process 6:207–216.

Shohamy D, Myers CE, Hopkins RO, Sage J, Gluck MA (2009) Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. J Cogn Neurosci 21:1820-1832.

Shohamy D, Wagner AD (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60:378–389.

Staib M, Bach DR (2018) Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. Neuroimage 166:276–284.

Weiss KR, Brown BL (1974) Latent inhibition: a review and a new hypothesis. Acta Neurobiol Exp 34:301–316.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroscience.2022.01.022.

(Received 10 August 2021, Accepted 27 January 2022)
(Available online 3 February 2022)

scientific reports

Check for updates

OPEN

Visual consequent stimulus complexity affects performance in audiovisual associative learning

Kálmán Tót[1], Gabriella Eördegh[2], Ádám Kiss[1], András Kelemen[3], Gábor Braunitzer[4], Szabolcs Kéri[1,4], Balázs Bodosi[1] & Attila Nagy[1]✉

In associative learning (AL), cues and/or outcome events are coupled together. AL is typically tested in visual learning paradigms. Recently, our group developed various AL tests based on the Rutgers Acquired Equivalence Test (RAET), both visual and audiovisual, keeping the structure and logic of RAET but with different stimuli. In this study, 55 volunteers were tested in two of our audiovisual tests, SoundFace (SF) and SoundPolygon (SP). The antecedent stimuli in both tests are sounds, and the consequent stimuli are images. The consequents in SF are cartoon faces, while in SP, they are simple geometric shapes. The aim was to test how the complexity of the applied consequent stimuli influences performance regarding the various aspects of learning the tests assess (stimulus pair learning, retrieval, and generalization of the previously learned associations to new but predictable stimulus pairs). In SP, behavioral performance was significantly poorer than in SF, and the reaction times were significantly longer, for all phases of the test. The results suggest that audiovisual associative learning is significantly influenced by the complexity of the consequent stimuli.

Associative learning is a basic cognitive function, in which different stimuli, cues, and/or outcome events are coupled. This learning type includes cognitive tasks like probabilistic learning[1,2], latent inhibition[3] and sensory preconditioning[4], and equivalence learning[5,6]. Equivalence learning is a specific kind of associative learning in which two discrete and often different percepts (antecedents) are linked together based on a shared outcome (consequent). A visually guided paradigm, the Rutgers Acquired Equivalence Test (RAET), was developed by Myers et al.[7] to investigate equivalence learning in humans. The test is computer-based and divided into two main phases: the acquisition and the test phases. In the acquisition phase, the subject's task is to associate two different visual stimuli based on feedback information about the correctness of the choices. This way, the rule of pairing is acquired. In the subsequent test phase, the subject must recall the already learned associations (retrieval) and build new, hitherto not seen but predictable associations (generalization or transfer). Regarding the neural correlates, both the original study of the Myers group[7] and subsequent investigations[8–13] demonstrated that the acquisition phase is linked to the fronto-striatal loops, while the test phase is linked primarily to the hippocampi and the medial temporal lobe[7,14–19]. The basal ganglia and the hippocampi are structures of key importance in equivalence learning, and they are also involved in multisensory processing[20–23].

Several studies reported that stimulus complexity influences auditory guided associative learning and more complex stimuli cause better responses and greater cortical activation[24–26]. It has also been demonstrated in studies from the cellular to the behavioral level that responses are quicker and more precise in the case of multimodal stimuli[22,27–30]. A recent study by our research group[31], applying a modified but structurally identical version of RAET, showed that the complexity of the applied visual stimuli could also strongly influence the efficiency of associative learning: simple visual stimuli (antecedents: white, light gray, dark gray and black circles; consequents: colorless triangle, square, rhombus, and concave deltoid) with restricted semantic and color information allowed significantly poorer equivalence acquisition than more complex stimuli (antecedents: cartoon faces of a woman, a man, a boy and a girl; consequents: green, yellow, red and blue fish) without such feature restrictions. However, stimulus complexity did not affect retrieval and generalization (transfer).

Given that the key neural structures associated with RAET also play a role in multisensory processing, the question arises whether the complexity of the applied visual stimuli can also influence the effectiveness of multisensory (audiovisually guided) associative learning. In this study, we sought to answer this question. For this,

[1]Department of Physiology, Faculty of Medicine, University of Szeged, Dóm Tér 10, Szeged 6720, Hungary. [2]Faculty of Health Sciences and Social Studies, University of Szeged, Szeged, Hungary. [3]Department of Applied Informatics, University of Szeged, Szeged, Hungary. [4]Nyírő Gyula National Institute of Psychiatry and Addictions, Budapest, Hungary. ✉email: nagy.attila.1@med.u-szeged.hu

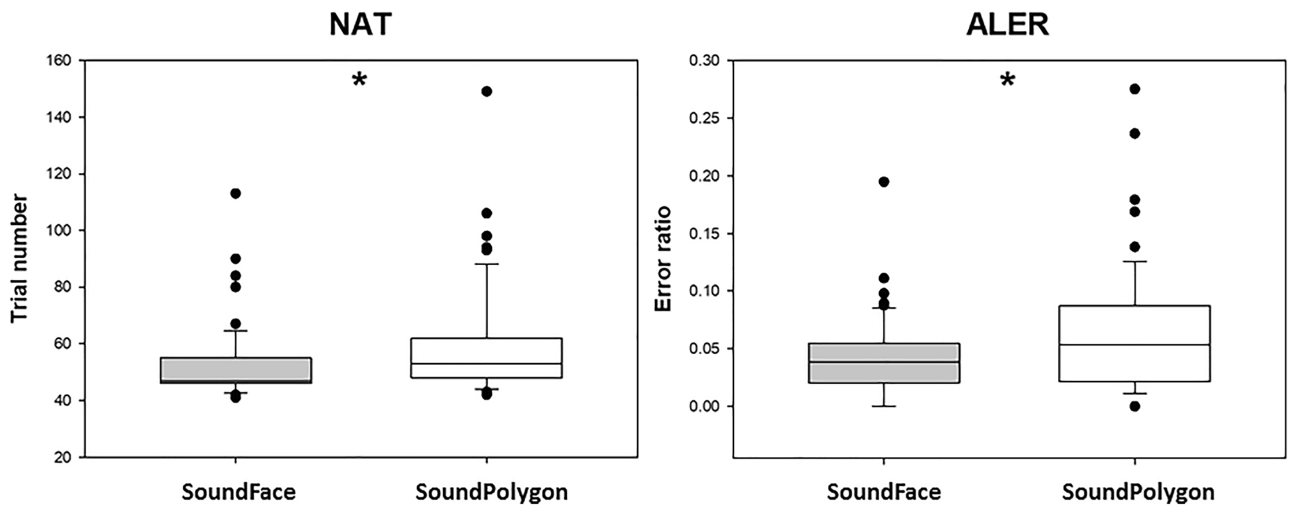

Figure 1. Performance in the acquisition phase in the two tests. NAT: the number of trials needed to complete the acquisition phase. ALER: error ratios in the acquisition phase. The lower margin of the boxes indicates the 25th percentile, the upper margin the 75th percentile, while the line within the boxes marks the median. The error bars (whiskers) above and below the boxes are the 90th and 10th percentiles, respectively. The dots over and under the whiskers represent the extreme outliers. Asterisk (*) indicates a significant difference at the level $p < 0.05$.

we used two audiovisual tests, both of which follow the RAET paradigm: SoundFace (SF) and SoundPolygon (SP). These tests have been developed in our laboratory. Both tests use sounds as antecedent stimuli, but SF uses cartoon faces and SP uses simple geometric shapes as consequents. That is, the consequent stimuli in SF (colored cartoon faces of a boy, a girl, a man and a woman) are relatively complex in the sense that they have well-defined, readily detectable and readily verbalizable distinctive features (e.g., colors, gender, age), while the consequents in SP lack such features. By comparing our volunteers' performance on these tests, we sought to test if consequent stimulus complexity influences audiovisual associative learning at all. It is important to note that the goal of this study was not to analyze how the gradual extraction of the different features of the consequent stimuli influences audiovisual associative learning (and to learn this way which features are more important, and which are less important for audiovisual associative learning). Instead, we chose to use the cartoon faces of RAET and a set of completely different, simple, non-face stimuli that lack all the specific features associated with the cartoon faces merely to establish the lack or existence of an effect.

Results

All 55 volunteers completed both tests. The analysis of their performance data is presented according to the two main phases of the test paradigm (acquisition and test).

Acquisition phase. The median number of trials in the acquisition phase (NAT) in SF was 47 (range: 41–113), and in SP, it was 53 (range: 42–149). The participants needed significantly more trials to learn the associations in SP ($Z = 2.417$, $p = 0.016$) (Fig. 1).

In SF, the median of error ratios in the acquisition (ALER) was 0.038 (range: 0.00–0.19), and in SP, it was 0.058 (range: 0.00–0.28). The difference between the two tests was significant ($Z = 2.213$, $p = 0.027$) (Fig. 1). The median reaction time (RT) for the acquisition trials in SF was 1611.619 ms (range: 1095.022–4016.42), and in SP, it was 1834.810 ms (range: 1204.867–4762.33). The difference was significant ($Z = 3.703$, $p = 0.0002$).

Test phase. The median of retrieval error ratio (RER) in SF was 0.00 (range: 0.00–0.25), and in SP, it was 0.028 (range: 0.00–0.39). The difference was significant ($Z = 2.727$, $p = 0.0064$) (Fig. 2). As for the reaction times, the median RT for the retrieval trials in SF was 1586.39 ms (range: 980.056–3802.11), and in SP, it was 2000.42 ms (range: 1222.33–4790.44). The difference was significant ($Z = 4.994$, $p = 0.000001$).

The median of generalization error ratio (GER) in SF was 0.00 (range: 0.00–0.58), and in SP, it was 0.083 (range: 0.00–1.00). The difference was significant ($Z = 3.085$, $p = 0.002$) (Fig. 2). The median RT for the generalization trials in SF was 1769.17 ms (range: 1145.75–5722.83), and in SP, it was 2544.25 ms (range: 1282.58–18,381.00). The difference was significant ($Z = 3.938$, $p = 0.00008$).

Discussion

To our knowledge, this is the first study to investigate the effect of visual stimulus complexity on multisensory (audiovisual) guided associative learning. The same set of auditory stimuli were applied with two different series of visual stimuli in two tests based on the same paradigm. The two sets of visual stimuli differed in their feature richness and complexity. Multisensory guided equivalence learning and subsequent retrieval and generalization

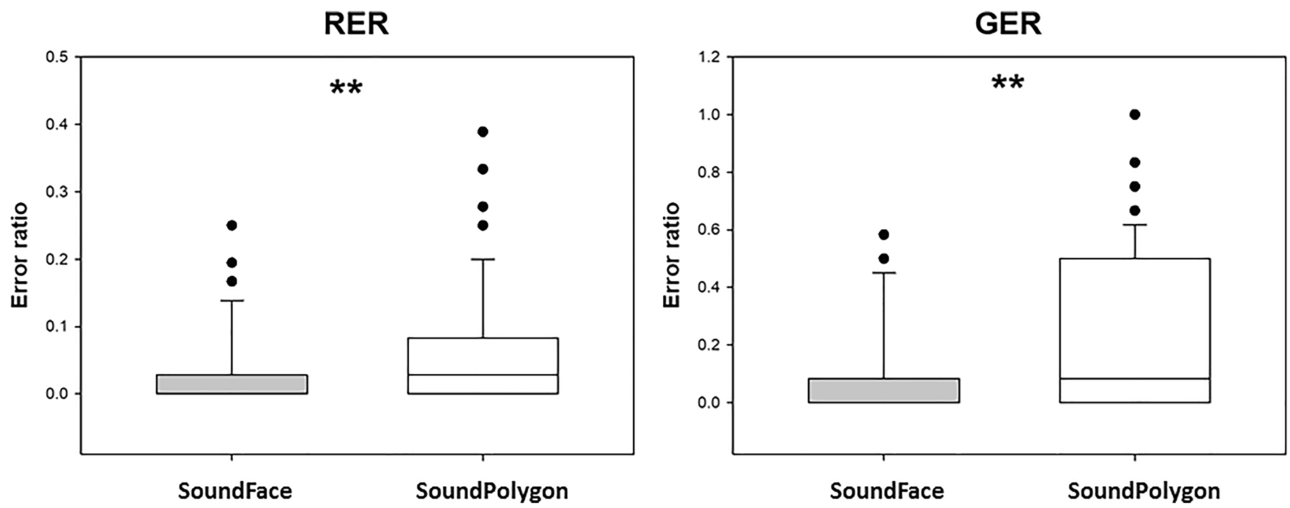

Figure 2. Performance in the test phase in the two tests. RER: retrieval error ratio, GER: generalization error ratio. Asterisk (*) indicates significant difference at the level $p < 0.01$. Otherwise, the conventions are the same as in Fig. 1.

were all influenced markedly by the complexity of the visual stimuli. The difference also showed in significantly shorter reaction times when the more complex, feature-rich visual stimuli were used.

In this study, we used two audiovisual tests that were developed in our laboratory, based on RAET, a visually guided equivalence learning test designed by Catherine E. Myers and colleagues at Rutgers University[7]. The original paradigm tests visually guided pair learning, the retrieval of the already learned stimulus pairs and the ability to apply the previously learned associations to build new stimulus pairs. The key brain structures associated with this task (the hippocampi and the basal ganglia) are also known for their role in multisensory processing[22,23,32,33]. Therefore, we developed SF that uses cartoon faces as consequents with auditory antecedents[34]. SF was administered to healthy adult subjects and in psychiatric patient populations and the results were compared to those of the original, visually guided RAET test[13,34]. The comparison indicated that the fact alone that the task had become multisensory did not influence the volunteers' performance to a significant degree, which led us to the conclusion that multimodality itself does not interfere with the efficiency of associative learning, retrieval, or generalization.

The visual stimuli both in the visual RAET and the audiovisual SoundFace are complex, colored stimuli with the potential to evoke associations and emotional responses, which, in turn, can serve as extra clues that recruit various cortical areas to enhance performance[35–37]. Such clues can thus mask the contribution of subcortical structures. We developed a new visually guided test, Polygon[31] to reduce this effect. Polygon uses simple geometric shapes both as antecedents and consequents. Such simple shapes are relatively meaningless in themselves, and they can hardly evoke emotions. Therefore, we hypothesized that the use of geometric shapes would allow us to minimize cortical contributions to task performance and thus allow a better assessment of subcortical contributions. Indeed, the first study with Polygon[31] revealed a specific pattern: it took significantly more trials for the volunteers to learn the stimulus pairs (and they made significantly more mistakes), but the reduced complexity of the stimuli had no significant effect on either the retrieval or the generalization part of the test phase.

The next logical step was to investigate what effect visual stimulus complexity might have on multisensory guided (audiovisual) equivalence learning. For this purpose, we combined the antecedent sounds of SF[31] and the geometric shapes of Polygon[31] into a new test (SP) and compared volunteers' performance on this test to their performance on SF. This comparison is presented in this study. In contrast to what was found when tests using visual stimuli only were compared (RAET vs. Polygon)[31], in the case of these multisensory tests, decreased stimulus complexity affected not only the acquisition phase, but the entire test phase, including retrieval and generalization. That is, it seems that when only visual stimuli are used, decreased stimulus complexity makes learning difficult, but if learning has been successful, retrieval and generalization are spared. Such a sparing does not seem to occur when visual stimulus complexity is decreased in an audiovisual (multimodal or multisensory) learning environment. While it comes as no surprise (in fact, it is somewhat intuitive) that stimulus complexity influences the efficiency of associative learning[24–26], it is difficult to tell why decreased stimulus complexity affects performance in all phases of the audiovisual version of the test paradigm, while in the visual version, only acquisition is affected.

In an earlier study[38], based on developmental data, we argued that learning and memory in this specific paradigm might be best described by the integrative encoding account of associative learning[39,40]. This account concentrates on two specific neural loops, the substantia nigra (SN)- striatum loop and the ventral tegmentum (VT)-hippocampus loop, which can be activated in parallel. While the SN- striatum loop supports primarily the voluntary learning of stimulus pairs with the help of feedback, the VT- hippocampus loop transfers information to the hippocampi, where a network of all encountered stimuli is constructed, with their connections and overlaps[41–46]. Then, this hippocampal network is activated in the test phase, which makes both retrieval and generalization possible. Based on this account, it is possible that the hippocampi, even if they are typically discussed as structures of key importance in explicit memory[39,47], can support implicit functions as well. That is why, as we earlier argued[38,46], children can generalize at a high level with poor acquisition and retrieval. In other words, they have the information, and they can use it as long as no conscious effort is involved. It must be noted that the integrative encoding account was developed based on visual (non-multisensory) learning paradigms, and we have no information whatsoever if it can be applied to multisensory learning as well. Based on the information available at this point, it might be hypothesized that decreased stimulus complexity affects all phases of the multisensory test (but not the visual test, as demonstrated earlier) because hippocampal compensation is either specific to the visual stimulus modality or it works only if stimuli of the same modality are used. This, however, is only a crude hypothesis, which is made by inference from the literature.

At the same time, it must be also noted that a direct comparison between SF/SP and the purely visual version of RAET is not possible as the latter uses colored cartoon fish as consequent stimuli. While it is not entirely obvious how this could contribute to the observed difference, the confounding effect of this methodological factor cannot be ruled out. Another possible explanation is that in SF, equivalence might be established between the face consequents too, based on their various features, which makes learning easier in SF than in SP, where consequents do not share such readily detectable features. However, assuming that the integrative encoding account[39,40] (see above) can be applied in an audiovisual context too, this should affect only the acquisition phase. It may be that the lack of equivalence between the consequents in SP can explain poorer acquisition, but it does not seem to be a good explanation for poorer transfer. In the sense of the integrative encoding account, the fact that the stimulus pairs are more difficult to learn (which shows as more errors and a longer acquisition phase in RAET and its various versions) does not necessarily imply poor transfer. This is exactly what we saw when we administered the original version of RAET to small children.[46].

Beside the poorer performance in all phases, reaction times were also significantly longer in SP. This is consistent with earlier findings, where, in a face-name association task, subjects' reaction times were significantly shorter in response to more complex, high-salience colored faces expressing emotions than to less distinctive, grayscale face stimuli.[48] This shows that complexity facilitates decision making, while in the (relative) lack of distinctive features, the facilitating effect is absent.

As for the limitations, we would like to point out the following.

First, this study is best understood as an exploratory study that sought to establish if the complexity of visual consequent stimuli has any effect on subjects' performance in an audiovisual associative learning paradigm. By complexity, we simply meant how rich the applied stimuli were in well-described, readily identifiable (and possibly verbalizable) features that can be used as cues for learning. The cartoon faces are relatively rich in these: age, gender, hair color and facial expression are all such cues. In contrast, the polygons are colorless and they definitely do not have age, gender or any feature that is even close to a facial expression. This is a crude comparison, and it does not allow a finer analysis of the difference; all it allows is the conclusion that the presence or lack of such easily identifiable and obvious features does make a difference. Whether this is because these cues are easy to verbalize or because they are characteristically human features (which activate additional neural circuits) or for some other reason should be addressed in studies designed for that purpose. A logical next step would be to generate several sets of the cartoon face consequents with gradually decreasing complexity (cue content) and repeat the measurements with all the sets.

Second, it is a limitation of SP sepcifically that it is possible that the female voice and the female face are matched, which could provide an extra cue for learning. While it is not always the case (the stimulus pairings are randomly generated for each session), this could interfere with the results in some cases, even if not to a major extent.

In summary, in this study we have demonstrated that in a multisensory associative learning paradigm, where the antecedent stimuli are sounds, the complexity of the consequent visual stimuli has a significant effect on both learning and generalization.

Methods

Subjects. Fifty-five healthy adult volunteers participated in the presented study (27 females and 28 males, age mean: 31.36 ± 14.56 years, range: 18–69; five participants were over the age of 60). The estimated minimum sample size was 47, assuming $p < 0.05$, $1 - \beta = 0.95$ and an effect size of 0.5. The sample size estimation was performed in G*Power 3.1.9.2 (Düsseldorf, Germany). The volunteers received no compensation and were free to quit without any negative consequence. The volunteers were informed about the study's background, goals, and procedures and gave their written informed consent. Any psychiatric, neurological, otological or ophthalmological condition that could interfere with the participant's performance was an exclusion criterion. Before each testing session, the participants were shown the stimuli of the tests one by one (each stimulus once) to make sure that they could see and hear them correctly. The study protocol conformed to the Declaration of Helsinki in all respects and was approved by the Regional Research Ethics Committee for Medical Research at the University of Szeged, Hungary (27/2020-SZTE).

The applied multisensory tests. Two audiovisual tests of our own development were administered (SoundFace and SoundPolygon, see below). Both tests were run on laptops (Lenovo ThinkBook 15-IIL, Lenovo, China), and the auditory stimuli were administered through over-ear headphones (Sennheiser HD439, Sennheiser, Germany). The volunteers were tested one-by-one in a quiet room, sitting at a comfortable distance (57 cm) from the screen. No forced quick responses were expected to avoid performance anxiety, but the participants were instructed to respond as quickly as possible. This way, explicit time pressure could be avoided, yet, the participants were aware that it was desirable that they spent a limited amount of time with each trial. The

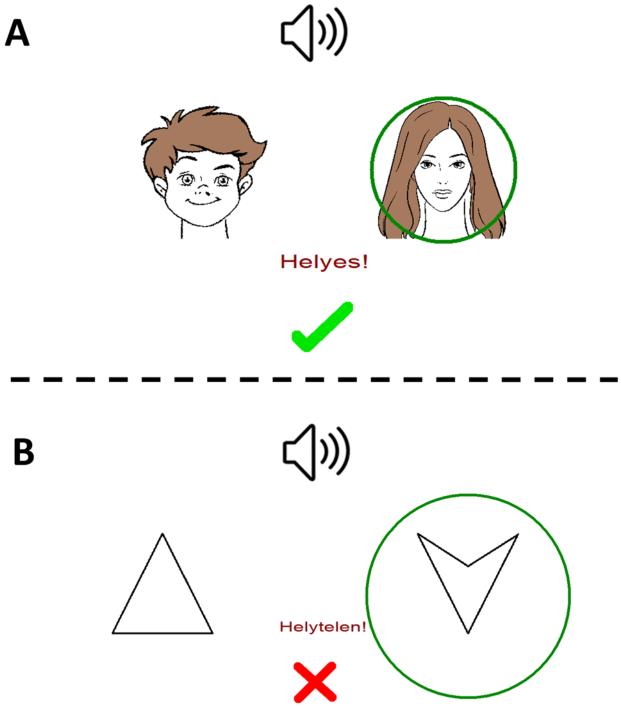

Figure 3. A trial in the acquisition phase of SoundFace (**A**) and SoundPolygon (**B**). In each trial, the subject simultaneously hears a sound (antecedent) and sees two faces (consequents) on the left and right side of the screen. Then the subject guesses which consequent belongs to the given antecedent sound by pressing the "left" or "right" button. The choice is indicated with the green circle. Immediate visual feedback is given. A green checkmark with the word Helyes! (Correct!) indicates a correct guess and a red X with the word Helytelen! (Incorrect!) indicates an incorrect guess.

volunteers completed both tests immediately one after another, in a pseudorandom order to avoid the carry-over effect.

SoundFace is based on RAET as described in Myers et al.[7,34]. The structure of RAET was kept, but it was translated into the Hungarian language and transformed into an audiovisual paradigm in Assembly for Windows. These modifications were performed with the written permission of Catherine E. Myers. In SoundFace, the subject is first asked to learn associations through trial and error. There are four sounds as antecedents and four possible faces as consequents. The antecedent stimuli are different and clearly distinguishable sounds: a cat (A1), a guitar note (A2), the sound of a vehicle (B1) and a woman's voice (B2). The consequents are different cartoon faces: an adult male (X1), a girl (X2), an adult woman (Y1), and a boy (Y2). The auditory and visual stimuli were semantically incongruent (except for the case when a woman's voice is matched with a woman's face, but this is not always the case). In each trial, the subject simultaneously hears a sound and sees two faces on the left and right sides of the screen. The subject is instructed to guess which face belongs to the given sound and indicate his or her guess by pressing either the „left" or the „right" button. The duration of the auditory stimulus was consistently 1.5 s. The visual stimuli lasted until the participant made the decision with the pressing of the "left" or "right" button. In this respect, SoundFace and SoundPolygon are identical (Fig. 3). The pairs are randomly generated by the software for each subject.

The paradigm is divided into two main phases: the acquisition and test phases. In the acquisition phase, visual feedback was given about the correctness of the choice. In the initial part of the acquisition phase, the subject learns through trial-and-error that if sounds A1 or A2 are presented, the correct response is to choose face X1 over Y1. Similarly, if sounds B1 or B2 are presented, the correct response is to choose face Y1 over X1. This way, it is learned that in terms of their consequents, A1 = A2 and B1 = B2. Once this has been established, new stimulus pairs are added. This time, the subject learns that if sound A1 is presented, the correct response is to choose X2 over Y2, and if sound B1 is presented, the correct response is to choose Y2 over X2. This way, antecedents A1 and B1 gain additional consequents. At this point, the subject knows that A1: X1, X2 and B1:Y1, Y2. Six items are presented in the acquisition phase from the eight possible stimulus pairs. A2:X2 and B2:Y2 are not presented, but it is implied by the connection A1 = A2 and B1 = B2. After each newly introduced stimulus pair, the participant must give a certain number of subsequent correct answers (4, 6, 8, 10, 12 after each new association, respectively) to accomplish the acquisition phase. Because of this, the number of trials in this phase is not constant, and it depends on how efficiently the given individual learns.

Once having completed the acquisition phase, the participant continues with the test phase, where feedback is no longer given about the correctness of the responses. In this phase, the retrieval and generalization are tested. Retrieval refers to the recall of the already known (learned) stimulus pairs, while generalization refers to

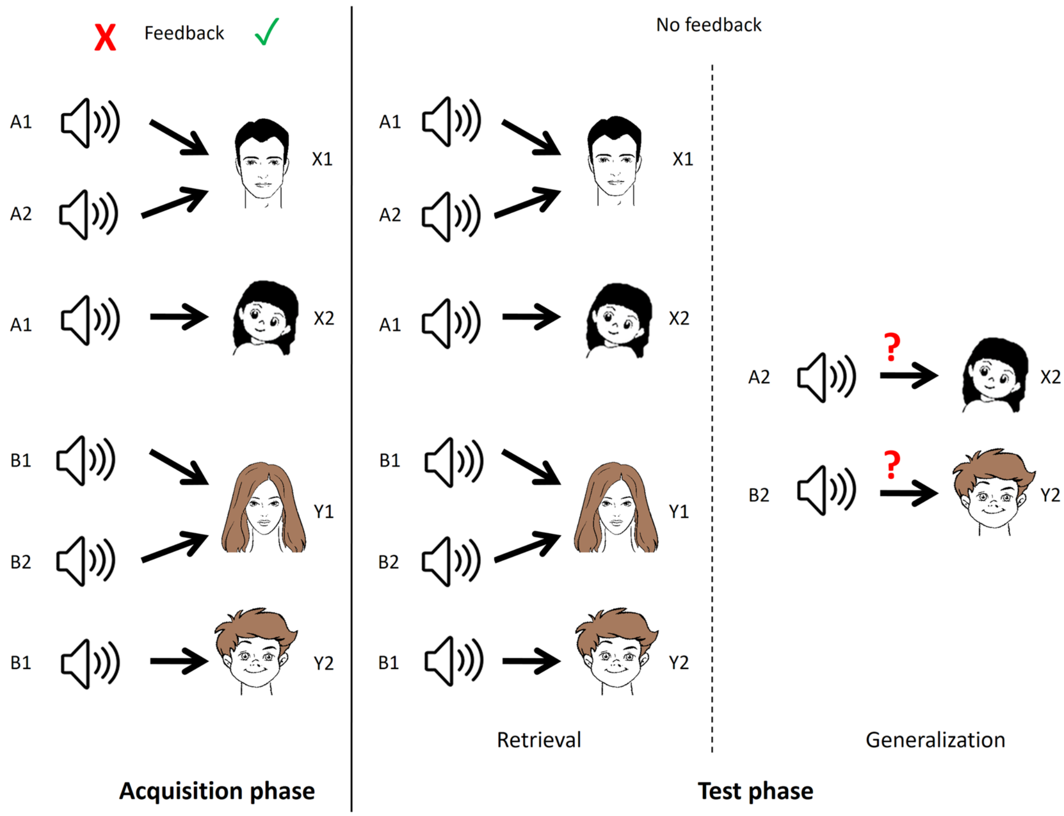

Figure 4. Overview of the structure of the SoundFace test. The antecedent stimuli are sounds of a cat (A1), a guitar note (A2), the sound of a vehicle (B1), and a woman's voice (B2). The consequents are cartoon faces of a man (X1), a girl (X2), a woman (Y1), and a boy (Y2).

Acquisition			Test	
Shaping	Equivalence training	New consequents	Retrieval	Generalization
A1—>X1	A1—>X1	A1—>X1	A1—>X1	
	A2—>X1	A2—>X1	A2—>X1	
		A1—>X2	A1—>X2	
				A2—>X2
B1—>Y1	B1—>Y1	B1—>Y1	B1—>Y1	
	B2—>Y1	B2—>Y1	B2—>Y1	
		B1—>Y2	B1—>Y2	
				B2—>Y2

Table 1. A summary of the audiovisual associative learning paradigms. A,B: antecedents (the same sounds in both tests), X,Y: consequents (faces in SoundFace and simple geometric forms in SoundPolygon tests).

making the A2:X2 and B2:Y2 stimulus pairs not presented in the acquisition phase but implied by the connection A1 = A2 and B1 = B2. If the subject has successfully acquired the said associations, he or she will choose X2 when A2 is presented and Y2 when B2 is presented, even if he or she has not seen these pairs before. The subject is not informed that new stimulus pairs are to be expected in the test phase. The number of trials in the test phase is constant. There are altogether 48 trials, of which 36 are retrieval trials, and 12 are generalization trials. These are mixed in random order. The overview of the paradigm is given in Fig. 4. See also Table 1 for clarification.

SoundPolygon has the same structure as SoundFace, but with simplified visual stimuli. Instead of cartoon faces, simple geometric shapes are used as consequents: a triangle (X1), a square (X2), a rhombus (Y1), and a concave deltoid (Y2). The auditory stimuli are the same as in SoundFace. Figure 5 summarizes SoundPolygon.

Data analysis. The performance of the participants was characterized by four main parameters: the number of trials necessary for the completion of the acquisition phase (NAT), association learning error ratio (the ratio of incorrect choices during the acquisition trials, ALER), retrieval error ratio (RER), and generalization error

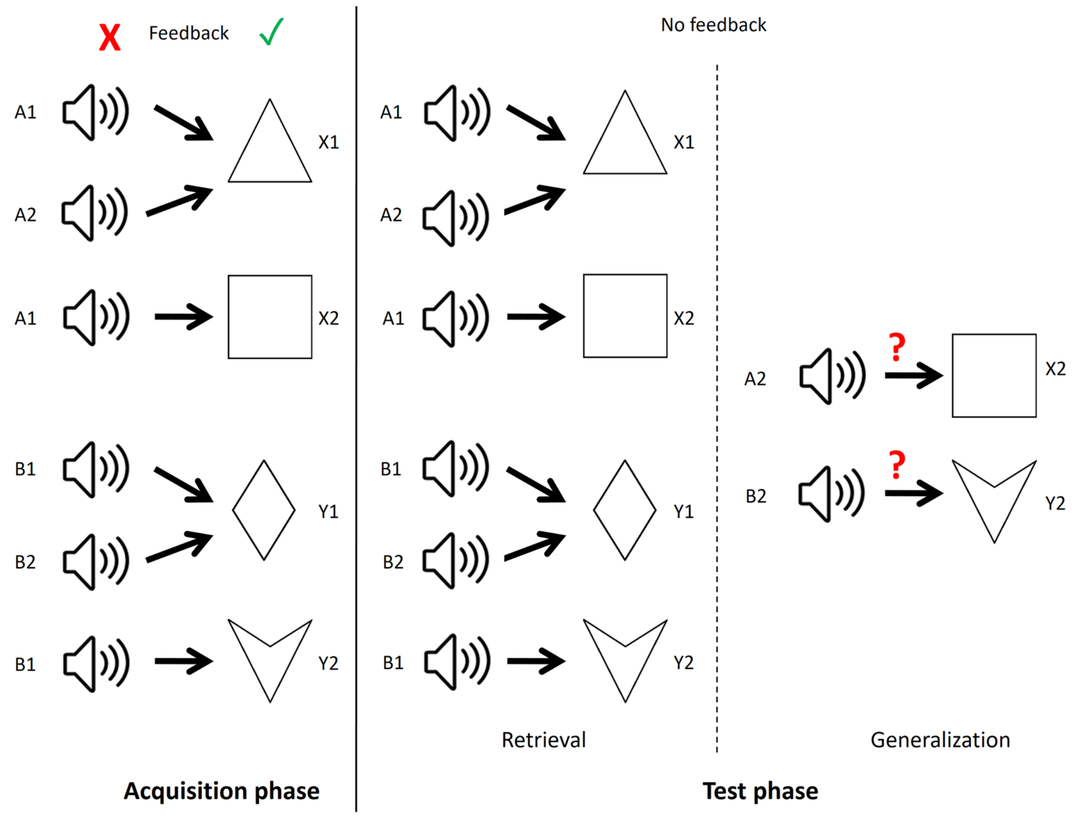

Figure 5. Overview of the structure of the SoundPolygon test. The consequents are simple geometric shapes: a triangle (X1), a square (X2), a rhombus (Y1), and a concave deltoid (Y2).

ratio (GER). NAT and ALER are performance parameters of the acquisition phase. RER and GER are performance parameters of the test phase. Error ratios were calculated by dividing the number of incorrect responses by the total number of trials. Reaction times were recorded for each trial, and they were analyzed for the acquisition, retrieval, and generalization trials separately. Reaction time was defined as the time elapsed between the appearance of the stimuli and the subject's response. Only RTs of the correct choices were included, and values over 3SD were excluded.

Statistical analysis was performed in Statistica 13.4.0.14 (TIBCO Software Inc., USA). NAT, ALER, RER, and GER were compared between the two paradigms. As the data were non-normally distributed (Shapiro–Wilk $p < 0.05$), the Wilcoxon matched-pairs test was used for the hypothesis tests.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 3 June 2022; Accepted: 20 October 2022
Published online: 22 October 2022

References

1. Shohamy, D., Myers, C. E., Hopkins, R. O., Sage, J. & Gluck, M. A. Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *J. Cogn. Neurosci.* **21**, 1821–1833. https://doi.org/10.1162/jocn.2009.21138 (2009).
2. Sáringer, S., Fehér, Á., Sáry, G. & Kaposvári, P. Online measurement of learning temporal statistical structure in categorization tasks. *Mem. Cognit.* https://doi.org/10.3758/s13421-022-01302-5 (2022).
3. Weiss, K. R. & Brown, B. L. Latent inhibition: A review and a new hypothesis. *Acta Neurobiol. Exp. (Wars)* **34**, 301–316 (1974).
4. Rescorla, R. A. Simultaneous and successive associations in sensory preconditioning. *J. Exp. Psychol. Anim. Behav. Process* **6**, 207–216. https://doi.org/10.1037//0097-7403.6.3.207 (1980).
5. Ward-Robinson, J. & Hall, G. The role of mediated conditioning in acquired equivalence. *Q. J. Exp. Psychol. B* **52**, 335–350. https://doi.org/10.1080/027249999393031 (1999).
6. Molet, M., Miller, H. & Zentall, T. R. Acquired equivalence of cues by presentation in a common context in rats. *Anim. Cogn.* **15**, 143–147. https://doi.org/10.1007/s10071-011-0431-4 (2012).
7. Myers, C. E. *et al.* Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J. Cogn. Neurosci.* **15**, 185–193. https://doi.org/10.1162/089892903321208123 (2003).
8. Coutureau, E. *et al.* Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *J. Exp. Psychol. Anim. Behav. Process* **28**, 388–396 (2002).
9. Bódi, N., Csibri, E., Myers, C. E., Gluck, M. A. & Kéri, S. Associative learning, acquired equivalence, and flexible generalization of knowledge in mild Alzheimer disease. *Cogn. Behav. Neurol.* **22**, 89–94. https://doi.org/10.1097/WNN.0b013e318192ccf0 (2009).

10. Myers, C. E. *et al.* Learning and generalization deficits in patients with memory impairments due to anterior communicating artery aneurysm rupture or hypoxic brain injury. *Neuropsychology* **22**, 681–686. https://doi.org/10.1037/0894-4105.22.5.681 (2008).
11. Kéri, S., Nagy, O., Kelemen, O., Myers, C. E. & Gluck, M. A. Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. *Schizophr. Res.* **77**, 321–328. https://doi.org/10.1016/j.schres.2005.03.024 (2005).
12. Eördegh, G. *et al.* Impairment of visually guided associative learning in children with Tourette syndrome. *PLoS ONE* **15**, e0234724. https://doi.org/10.1371/journal.pone.0234724 (2020).
13. Pertich, Á. *et al.* Maintained visual-, auditory-, and multisensory-guided associative learning functions in children with obsessive-compulsive disorder. *Front. Psychiatry* **11**, 571053. https://doi.org/10.3389/fpsyt.2020.571053 (2020).
14. Cohen, N. J. *et al.* Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus* **9**, 83–98. https://doi.org/10.1002/(SICI)1098-1063(1999)9:1%3c83::AID-HIPO9%3e3.0.CO;2-7 (1999).
15. Gogtay, N. *et al.* Dynamic mapping of normal human hippocampal development. *Hippocampus* **16**, 664–672. https://doi.org/10.1002/hipo.20193 (2006).
16. Larsen, B. & Luna, B. In vivo evidence of neurophysiological maturation of the human adolescent striatum. *Dev. Cogn. Neurosci.* **12**, 74–85. https://doi.org/10.1016/j.dcn.2014.12.003 (2015).
17. Moustafa, A. A., Myers, C. E. & Gluck, M. A. A neurocomputational model of classical conditioning phenomena: A putative role for the hippocampal region in associative learning. *Brain Res.* **1276**, 180–195. https://doi.org/10.1016/j.brainres.2009.04.020 (2009).
18. Persson, J. *et al.* Sex differences in volume and structural covariance of the anterior and posterior hippocampus. *Neuroimage* **99**, 215–225. https://doi.org/10.1016/j.neuroimage.2014.05.038 (2014).
19 Porter, J. N. *et al.* Age-related changes in the intrinsic functional connectivity of the human ventral vs. dorsal striatum from childhood to middle age. *Dev. Cogn. Neurosci.* **11**, 83–95. https://doi.org/10.1016/j.dcn.2014.08.011 (2015).
20. Chudler, E. H., Sugiyama, K. & Dong, W. K. Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Res.* **674**, 33–45. https://doi.org/10.1016/0006-8993(94)01427-j (1995).
21. Schwarz, M., Sontag, K. H. & Wand, P. Sensory-motor processing in substantia nigra pars reticulata in conscious cats. *J. Physiol.* **347**, 129–147. https://doi.org/10.1113/jphysiol.1984.sp015057 (1984).
22. Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z. & Benedek, G. Multisensory integration in the basal ganglia. *Eur. J. Neurosci.* **24**, 917–924. https://doi.org/10.1111/j.1460-9568.2006.04942.x (2006).
23. Bates, S. L. & Wolbers, T. How cognitive aging affects multisensory integration of navigational cues. *Neurobiol. Aging* **35**, 2761–2769. https://doi.org/10.1016/j.neurobiolaging.2014.04.003 (2014).
24. Güçlütürk, Y., Güçlü, U., van Gerven, M. & van Lier, R. Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Sci. Rep.* **8**, 3439. https://doi.org/10.1038/s41598-018-21636-y (2018).
25. Staib, M. & Bach, D. R. Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. *Neuroimage* **166**, 276–284. https://doi.org/10.1016/j.neuroimage.2017.11.009 (2018).
26. Maor, I. *et al.* Neural correlates of learning pure tones or natural sounds in the auditory cortex. *Front. Neural Circuits* **13**, 82. https://doi.org/10.3389/fncir.2019.00082 (2019).
27. Frens, M. A. & Van Opstal, A. J. Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res. Bull.* **46**, 211–224. https://doi.org/10.1016/s0361-9230(98)00007-0 (1998).
28. Harrington, L. K. & Peck, C. K. Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp. Brain Res.* **122**, 247–252. https://doi.org/10.1007/s002210050512 (1998).
29. Giard, M. H. & Peronnet, F. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* **11**, 473–490. https://doi.org/10.1162/089892999563544 (1999).
30. Patching, G. R. & Quinlan, P. T. Cross-modal integration of simple auditory and visual events. *Percept. Psychophys.* **66**, 131–140. https://doi.org/10.3758/bf03194867 (2004).
31. Eördegh, G. *et al.* The influence of stimulus complexity on the effectiveness of visual associative learning. *Neuroscience* **487**, 26–34. https://doi.org/10.1016/j.neuroscience.2022.01.022 (2022).
32. Nagy, A., Paróczy, Z., Norita, M. & Benedek, G. Multisensory responses and receptive field properties of neurons in the substantia nigra and in the caudate nucleus. *Eur. J. Neurosci.* **22**, 419–424. https://doi.org/10.1111/j.1460-9568.2005.04211.x (2005).
33. Ravassard, P. *et al.* Multisensory control of hippocampal spatiotemporal selectivity. *Science* **340**, 1342–1346. https://doi.org/10.1126/science.1232655 (2013).
34. Eördegh, G. *et al.* Multisensory guided associative learning in healthy humans. *PLoS ONE* **14**, e0213094. https://doi.org/10.1371/journal.pone.0213094 (2019).
35. Puszta, A. *et al.* Cortical power-density changes of different frequency bands in visually guided associative learning: A human EEG-study. *Front. Hum. Neurosci.* **12**, 188. https://doi.org/10.3389/fnhum.2018.00188 (2018).
36. Puszta, A. *et al.* Power-spectra and cross-frequency coupling changes in visual and audio-visual acquired equivalence learning. *Sci. Rep.* **9**, 9444. https://doi.org/10.1038/s41598-019-45978-3 (2019).
37. Puszta, A. *et al.* Predicting stimulus modality and working memory load during visual- and audiovisual-acquired equivalence learning. *Front. Hum. Neurosci.* **14**, 569142. https://doi.org/10.3389/fnhum.2020.569142 (2020).
38. Öze, A. *et al.* Acquired equivalence and related memory processes in migraine without aura. *Cephalalgia* **37**, 532–540. https://doi.org/10.1177/0333102416651286 (2017).
39. Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41–50. https://doi.org/10.1038/35036213 (2000).
40. Shohamy, D. & Wagner, A. D. Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389. https://doi.org/10.1016/j.neuron.2008.09.023 (2008).
41. Daniel, R. & Pollmann, S. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiol. Learn. Mem.* **114**, 90–100. https://doi.org/10.1016/j.nlm.2014.05.002 (2014).
42. Hiebert, N. M. *et al.* Striatum in stimulus-response learning via feedback and in decision making. *Neuroimage* **101**, 448–457. https://doi.org/10.1016/j.neuroimage.2014.07.013 (2014).
43. Balleine, B. W., Delgado, M. R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* **27**, 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007 (2007).
44. Scimeca, J. M. & Badre, D. Striatal contributions to declarative memory retrieval. *Neuron* **75**, 380–392. https://doi.org/10.1016/j.neuron.2012.07.014 (2012).
45. Delgado, M. R., Miller, M. M., Inati, S. & Phelps, E. A. An fMRI study of reward-related probability learning. *Neuroimage* **24**, 862–873. https://doi.org/10.1016/j.neuroimage.2004.10.002 (2005).
46. Braunitzer, G. *et al.* The development of acquired equivalence from childhood to adulthood—A cross-sectional study of 265 subjects. *PLoS ONE* **12**, e0179525. https://doi.org/10.1371/journal.pone.0179525 (2017).
47. Tulving, E. & Markowitsch, H. J. Episodic and declarative memory: Role of the hippocampus. *Hippocampus* **8**, 198–204. https://doi.org/10.1002/(SICI)1098-1063(1998)8:3%3c198::AID-HIPO2%3e3.0.CO;2-G (1998).
48. Bender, A. R., Naveh-Benjamin, M., Amann, K. & Raz, N. The role of stimulus complexity and salience in memory for face-name associations in healthy adults: Friend or foe?. *Psychol. Aging* **32**, 489–505. https://doi.org/10.1037/pag0000185 (2017).

Acknowledgements

The authors thank Anna Lazsádi and Ábel Hertelendy for their help in conducting the investigation and data collection. This work was supported by a grant from SZTE ÁOK-KKA Grant No. 2019/270-62-2. KT and ÁK were supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the Source of the National Research, Development and Innovation Fund. GE was supported by EFOP-3.6.1-16-2016-00008 grant.

Author contributions

A.N., Sz.K., G.E. G.B., conceived the study conception and design. Data collection was made by K.T., G.E., Á.K. and B.B. Data analysis were performed by K.T., A.N., Á.K. and A.K. The manuscript was written by A.N., G.B. and K.T. All authors discussed data analysis and interpretation. All authors reviewed/edited the manuscript and approved the final version. Funding acquisition was made by A.N.

Funding

Open access funding provided by University of Szeged.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22880-z.

Correspondence and requests for materials should be addressed to A.N.