Neuroscience and Biobehavioral Reviews 127 (2021) 607–618

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review article

Prototype-based category learning in autism: A review

Wolf Vanpaemel [a], Janine Bayer [b,*]

[a] Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium
[b] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

ARTICLE INFO

Keywords:
Category learning
Learning styles
Basic cognition
Classification
Family resemblance
Memory systems
Neurodiversity
Developmental disorder
Prototype & exemplar model
Computational modeling

ABSTRACT

Similarity-based categorization, as an important cognitive skill, can be performed by abstracting a categories' central tendency, the so-called prototype, or by memorizing individual exemplars of a category. The flexible selection of an appropriate strategy is crucial for effective cognitive functioning. The detail-focused cognitive style in individuals with autism spectrum disorders (ASD) has been hypothesized to specifically impair prototype-based categorization but to leave exemplar-based categorization unimpaired. We first give an overview of approaches to investigate prototype-based abstraction in the prototype-distortion task, with an emphasis on model-based approaches suitable to discern the two strategies on the individual level. The second part summarizes literature speaking to prototype-based categorization in ASD using that task. Despite considerable inconsistencies, most studies appear to confirm that autistic individuals have more difficulties to perform prototype-distortion tasks than non-autistic individuals. We highlight how inconsistencies in literature can be resolved by taking the differences in task designs into account. The current review illustrates the need for sensitive computational approaches, suitable to detect hidden individual differences and potential compensatory strategies.

1. Introduction

Autism spectrum disorders (ASD) are defined by 'persistent deficits in social communication and social interaction' and 'restricted, repetitive patterns of behavior, interests, or activities' (American Psychiatric Association, 2013). Aside from these most prominent characteristics of ASD, individuals with ASD have been shown to pay an exceedingly high focus to details and to attend less to global, contextual information (Happé and Frith, 2006). While this focus on detail can be of advantage when the processing of details is required, as, for example, in, visual search (Mottron et al., 2006), it can be hindering in situations which require the abstraction of contextual information and the generalization of acquired knowledge to new instances, as, for example, in categorization (Church et al., 2010).

Through the process of categorization, organisms assign stimuli to separate groups or 'categories'. Categorization enables quick responses to environmental demands (Koriat and Sorka, 2015; Seger and Miller, 2010), understanding social situations (e.g., emotion recognition; Etcoff and Magee, 1992; Mercado et al., 2020; Plate et al., 2019), and is involved in speech acquisition and processing (Liu and Holt, 2009). More generally, it allows to apply acquired knowledge to new situations. Without categorization, new situations would require a tremendous effort of processing.

Some category structures allow to abstract verbalizable rules, so that the presence of certain features (e.g. four equal straight sides and four right angles) can be used as indicators for category membership (e.g., square). In contrast, numerous other everyday structures are easier defined by similarities between the stimuli, so that category membership can be quickly determined by comparing the overall similarity of an item to our perceptual representation of the category. Similarity-based categorization can be highly efficient, because simple explicit rules are often not feasible, and complex rules can become very demanding. Two of the most prominent strategies for similarity-based categorization are the prototype and exemplar strategies. The former involves representing the category structure by a summary of all members, the so-called 'prototype', whereas in the latter, a category is represented by the collection of individual members, separately stored in memory as 'exemplars'.

The *prototype strategy* requires learners to abstract a summary (often the average) of a category by focusing on similarities between category members and ignoring category-unrelated differences among them (Minda and Smith, 2001; Posner and Keele, 1968). Category decisions of new items are then based on their similarity to the stored prototype. The

[*] Corresponding author.
E-mail address: j.bayer@uke.de (J. Bayer).

https://doi.org/10.1016/j.neubiorev.2021.05.016
Received 16 December 2020; Received in revised form 12 May 2021; Accepted 17 May 2021
Available online 19 May 2021

prototype strategy has been suggested to be particularly useful for large and coherent categories, in which category members are only gradually distinct from each other (Bowman and Zeithamova, 2020). The prototype strategy is assumed to rely on implicit perceptual processes, which is associated to a 'feeling of familiarity' rather than voluntary memory retrieval (Casale and Ashby, 2008; Schacter, 1990; Zeithamova et al., 2008).

The *exemplar strategy*, in contrast, requires learners to store and retrieve single category members in and from memory (Medin and Schaffer, 1978; Nosofsky, 1988). The decision whether a newly encountered item belongs to a certain category is then based on its similarity to all the stored individual exemplars. The exemplar strategy has been suggested to be most advantageous for small categories in which category members have little coherence (Bowman and Zeithamova, 2020; Homa et al., 1981). It is mediated by processes involved in declarative memory encoding and retrieval (Casale and Ashby, 2008; Zeithamova et al., 2008).

Whereas some studies aimed at providing evidence that only one of the strategies underlies similarity-based categorization (Mack et al., 2013; Nosofsky et al., 2012; Smith, 2002), a more nuanced position is that both are useful, depending on the exact circumstances (Bowman and Zeithamova, 2020; Casale and Ashby, 2008; Medin et al., 1984; Smith et al., 2016; Zeithamova and Bowman, 2020). The reasoning is that as individuals are exposed to different category structures, categorization strategies have to be flexibly adapted to meet situational requirements efficiently. From this perspective, an impairment in one of these strategies would make everyday life much more effortful and strenuous. Such a dissociation has, for example, been reported for patients with memory impairments, who exhibited intact performance tasks in which the prototype strategy was useful (Knowlton and Squire, 1993; Sinha, 1999; though see Nosofsky and Zaki, 1998), but were severely impaired in tasks in which the exemplar strategy is assumed to be most efficient (Zaki et al., 2003).

For individuals with ASD, the opposite pattern has been hypothesized (Happé and Frith, 2006): Their elevated focus on details specifically impairs prototype-based categorization (Church et al., 2010), because detail-oriented processing should make it particularly difficult to ignore category-unrelated differences, which is of crucial importance for prototype abstraction. In contrast, detailed processing and accurate representation of individual category exemplars is not expected to interfere with exemplar-based categorization. Therefore, individuals with ASD are expected to rely more on an exemplar strategy than typically developing individuals, strongly stressing memory processes, even in situations where the prototype strategy would be more helpful.

In the current review, we summarize past research on prototype-based categorization in individuals with ASD, focusing on data collected in the commonly used prototype-distortion task. Although the prototype-distortion task is not the only paradigm suitable to study representational abstraction, all investigations about prototype-based categorization in ASD have exclusively relied on this paradigm. Our review compliments a recent review by Mercado et al. (2020), who focus on perceptual category learning in ASD, its relationship to individual differences in basic cognitive processes and potential consequences on social behavior.

The article is organized as follows: First, we will discuss how prototype abstraction is assessed through variants of the prototype-distortion task. We will outline how performance in these tasks is related to prototype abstraction, highlight how formal modeling can help to uncover different strategies, and summarize what is known about neuronal processes mediating prototype-based categorization. The second section provides an exhaustive review of empirical findings about the performance of individuals with ASD on the prototype-distortion task. We first cover studies speaking to ASD-related difficulties in prototype-distortion tasks. We speculate that apparent inconsistencies in the literature can be resolved by taking task characteristics into account. Then, we summarize two studies which applied formal modeling on a data set from autistic and non-autistic individuals, illustrating the usefulness of these computational approaches. In the final part of this section, we review findings on potential neuronal and cognitive processes underlying ASD-related difficulties in prototype-based categorization.

2. Prototype abstraction in category learning

2.1. Prototype-distortion tasks

Prototype-based categorization is often investigated by the application of prototype-distortion tasks in humans (Bowman et al., 2020; Homa et al., 2008a; Minda and Smith, 2001; Nosofsky et al., 2012; Posner et al., 1967) and animals (Antzoulatos and Miller, 2011; Cook and Smith, 2006). In these tasks, a category is organized around a prototypical stimulus, surrounded by category members differing from the prototypical stimulus in varying degrees ('distorted category members'; Fig. 1 A. & B.). Variants of the task either contain a single category plus unrelated stimuli, which are non-members (A/notA; e.g., Aizenstein et al., 2000) or multiple contrasting categories (e.g. A/B, A/B/C etc. tasks; e.g., Bowman et al., 2020).

Typically, a prototype-distortion task starts with a training phase (Fig. 1 C.), where participants learn to assign items to one or more categories, either actively via trial-and-error (e.g., Little et al., 2006) or passively (e.g., Nosofsky et al., 2012) through the presentation of category labels. Typically, the prototype is not shown during training. During the test or transfer phase, participants are then asked to categorize the previously seen stimuli, the prototype(s) and a set of new stimuli, all of which are unlabeled. In most studies, no feedback is provided during the test phase.

A widely used version of the prototype-distortion paradigm is the dot pattern task (Aizenstein et al., 2000; Homa et al., 2008a; Little et al., 2006; Nosofsky et al., 2012), in which the prototype is constructed as a static pattern of several dots at random positions (Fig. 1 A.). In this paradigm, category members are created as distortions of the prototype by probabilistically moving the coordinates of the dots (Posner and Keele, 1968). Distortion levels vary with respect to predefined probabilities of dots moving to certain areas. This procedure allows to construct stimulus sets with a wide, continuous range of distances to the prototype. The perceptual distance between two different dot pattern stimuli is often calculated by averaging the (Euclidian) distances between the individual dots. Other variants of prototype-distortion tasks involve concrete stimuli, such as imaginary animals varying on multiple dimensions (Fig. 1 C.; e.g., Bowman and Zeithamova, 2020). Here, distorted versions of the prototype are constructed via manipulating discrete (e.g., presence of antennae) or continues features (e.g., length of antennae). Perceptual distances between category members and the prototype are often quantified by the city block metric.

2.2. Prototype-distortion tasks and prototype abstraction

A particularly robust phenomenon is the *prototype enhancement effect*, which entails that the prototype is classified more easily than other category members during test, even though it has not been presented during training (Knowlton and Squire, 1993). Another common observation is that the rates of endorsing stimuli as category members (i.e. 'endorsement rates') decline gradually with growing distance to the prototype, a phenomenon also known as *typicality gradient* (Smith and Minda, 2002; Fig. 1 D.).

The prototype enhancement effect and the typicality gradient are often considered to be of particular theoretical interest, because they are thought to provide a window into prototype abstraction during learning (Posner and Keele, 1968). Indeed, in most studies on prototype abstraction in ASD, prototype effects and typicality gradients are taken as indicators of prototype abstraction. Somewhat counterintuitively, however, prototype effects do not necessarily require prototype

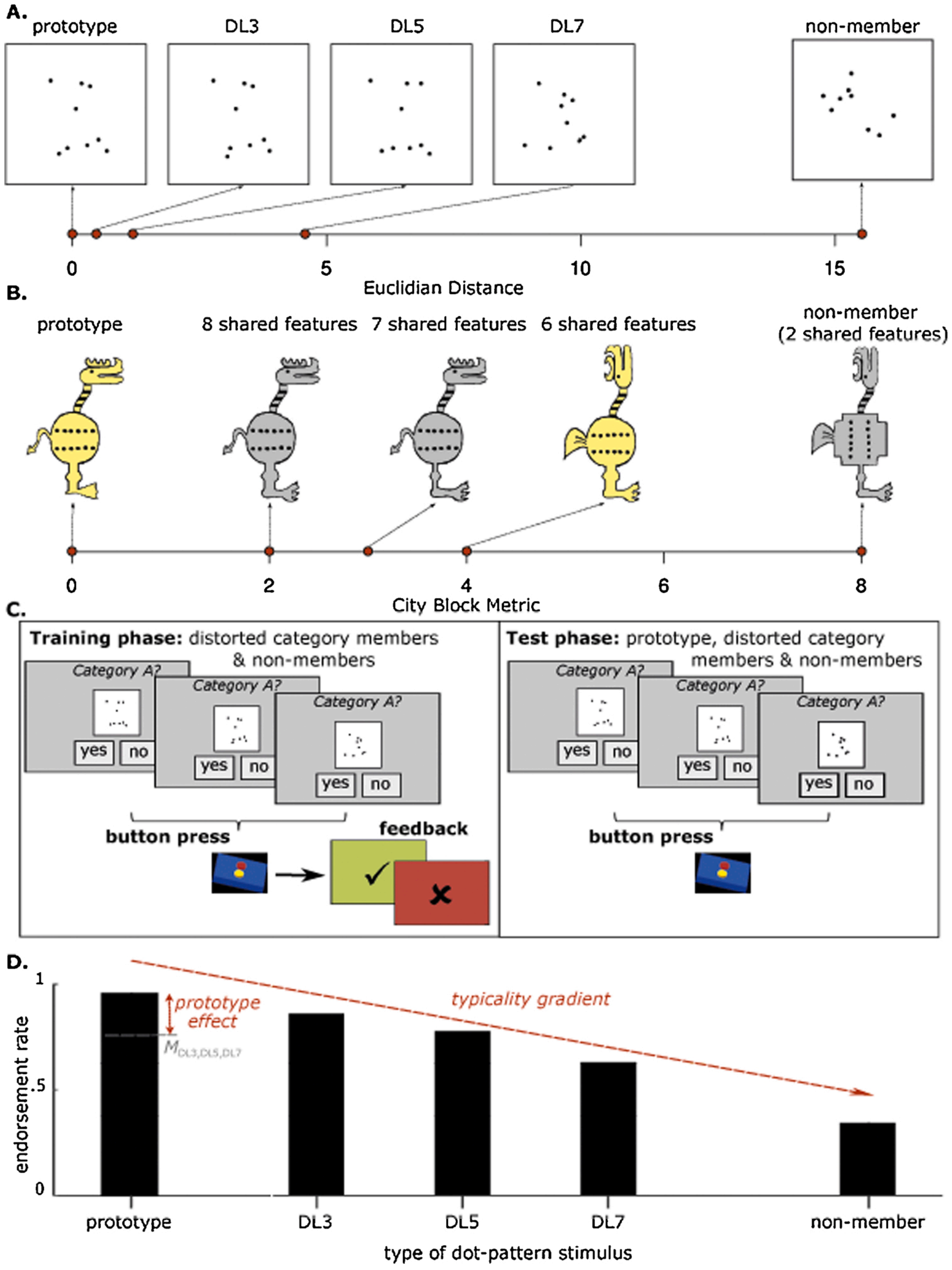

Fig. 1. A. Dot patterns representing the prototype of a category as well as examples for low (DL3), medium (DL5) and highly distorted (DL7) versions of the prototype, plus a random pattern unrelated to the category (non-member). The Euclidean distance of the stimulus to the prototype is marked by a red circle. B. Examples from a stimulus set with imaginary animals used by Bozoki et al. (2006) and Bowman and Zeithamova (2018). Stimuli differ with respect to 10 binary features. The city-block distance of the stimulus to the prototype is marked by a red circle. C. Dot patterns category learning task with a single category (A/notA task). In the training phase, participants learn via visual feedback which of the patterns belong to category A. The prototype is usually not shown during the training phase. In the test phase, participants are asked to classify old stimuli and new stimuli, including the prototype, without receiving any feedback. D. Theoretical rates of assigning stimuli to category A (endorsement rate) in the test phase. Most often, endorsement rates show a typicality gradient, in that they decline with growing distance from the prototype. Although participants never received feedback about the membership of the prototype is a category member, endorsement rates are often very high. Relatively higher endorsement rates for the prototype compared with averaged endorsement rates from previously seen and new distorted category members ($M_{DL3,DL5,DL7}$) is called the prototype enhancement effect.

abstraction, but can also be produced using an exemplar strategy (Shin and Nosofsky, 1992). Using formal analysis, it has been shown that the exemplar strategy can not only lead to reasonable accuracy levels, but can also result in a pattern of endorsement rates showing typicality gradients and prototype enhancement effects (Shin and Nosofsky, 1992; Smith, 2002). These effects are most pronounced when training exemplars are only mildly distorted versions of the prototype. Supplementary Fig. 1 illustrates an example of nearly indistinguishable prototype effects and typicality gradients for prototype- and exemplar-based categorization.

Recent research suggests that the exact implementational details of the prototype-distortion task affects the likelihood of adopting a prototype or an exemplar strategy (Bowman and Zeithamova, 2020). One task characteristic which is assumed to influence an individual's tendency to use a prototype over an exemplar strategy is the presence of either a single or multiple categories. Tasks with only one category (A/notA; Casale and Ashby, 2008) seem to be associated with a bias towards using a prototype strategy. A possible explanation for this observation is that the presence of competing categories moves the focus away from *within* category similarities, towards characteristics which show the biggest difference *between* categories (Davis and Love, 2010). This might interfere with the prototype strategy, as the processing of category-related similarities is crucial for prototype abstraction. Therefore, performance in A/notA tasks likely reflects prototype abstraction to a higher degree than performance in A/B tasks.

If this assumption is correct, one would expect that the performance of memory-impaired patients is more impacted in A/B tasks (where performance is more likely to be based on the memory-intensive exemplar strategy) than in A/notA tasks (where performance is more likely to be based on the memory-friendly prototype strategy). Indeed, it has been observed that memory-impaired patients show worse performance than controls in A/B tasks but not in A/notA tasks (Sinha, 1999; Zaki et al., 2003). Furthermore, there is evidence that activity in brain regions typically involved in episodic memory correlates with performing an A/B task, while performing an A/not A task is rather associated to activity in regions supporting non-declarative memory (Zeithamova et al., 2008).

A second task characteristic which has been suggested to play a role for a bias towards the prototype or exemplar strategy is the memorizability of the training stimuli. Clearly, a smaller number of category members which are very distinct from each other are easier to remember than a large number of indistinguishable members. In contrast, a set of stimuli with high coherence make it easier to abstract a prototype (Bowman and Zeithamova, 2020; Homa et al., 2008b, 1981; Minda and Smith, 2001). The memorizability likely also differs across stimulus types with concrete stimuli, such as imaginary animals, being easier to remember than abstract shapes, such as dot patterns (Smith et al., 1990).

Taken together, though widely used, the validity of prototype effects and typicality gradients as indicators of prototype abstraction is questionable, as prototype effects are compatible with the use of an exemplar strategy. Past research has suggested that especially tasks employing a single large category are conducive to prototype based categorization.

2.3. *Investigating categorization strategies with formal modeling*

Several studies have focused on prototype effects and relatively course typicality gradients for distortion levels (Froehlich et al., 2012; Gastgeb et al., 2012; Vladusich et al., 2010), so that more fine grained information, such as exact distances to the prototype, were not taken into account. Formal modeling aims to provide a detailed description of the categorization processes, and takes mathematically derived distance measures for every stimulus into account. As such, it can be useful to disentangle the two strategies for single individuals (Bowman and Zeithamova, 2018). The following paragraphs will summarize the most commonly used formal implementations of the prototype and exemplar strategies.

In their canonical forms, prototype and exemplar models both contain a *sensitivity parameter* (c), quantifying the steepness of how *psychological similarity* (s) decays with distance. Similarity is related to distances using an exponential decay function (Minda and Smith, 2001; Nosofsky, 1988; Posner and Keele, 1968).

(1) similarity of the *i*th stimulus to prototype: $s_{iP} = e^{-cd_{iP}}$
(2) similarity of the *i*th stimulus to exemplar *j*: $s_{ij} = e^{-cd_{ij}}$

In both models, a high sensitivity parameter reflects a steep exponential decay, whereas a low sensitivity parameter indicates a gradual, more linear decay. With a high sensitivity parameter even close category members are considered dissimilar, whereas with a low sensitivity even category members that are far away are still considered similar.

The endorsement probability (r_{i}) for a to-be-categorized item i is based on the similarity to category A (s_{iA}) and to category B (s_{iB}). According to the prototype model, this similarity is simply the similarity to the prototype *P* of each category (i.e., $s_{\mathrm{iA}} = s_{\mathrm{i}P_A}$ and $s_{\mathrm{iB}} = s_{\mathrm{i}P_B}$). According to the exemplar model, this similarity equals the summed similarity to training exemplars of each category (*i.e.*, $s_{iA} = \sum_{j \in A} s_{ij}$ and $s_{iB} = \sum_{j \in B} s_{ij}$). The endorsement probability is then calculated using s_{iA} as the numerator and the sum of s_{iA} and s_{iB} as the denominator. In the exemplar model, but for identifiability reasons not in the prototype model, a response scaling parameter γ is included to allow for more deterministic responding than dictated by similarity alone (Ashby and Maddox, 1993; but the inclusion of the parameter is seen as controversial by some authors: Olsson et al., 2004; Smith and Minda, 2002; see Vanpaemel, 2016 for a discussion).

(3) endorsement probabilities: $r_i = \frac{s_{iA}^{\gamma}}{s_{iA}^{\gamma} + s_{iB}^{\gamma}}$ (prototype model: $\gamma = 1$)

In the case of an A/notA task (i.e. one target category plus random stimuli), s_{iB} is replaced by a free *criterion (k)* parameter (4). A high criterion parameter implies that an item has to have a relatively high similarity to category A in order to be endorsed as category member. When the criterion parameter is low, also items with a relatively low similarity have with a high endorsement probability.

(4) endorsement probabilities: $r_i = \frac{s_{iA}^{\gamma}}{s_{iA}^{\gamma} + k}$ (prototype model: $\gamma = 1$)

2.4. *Investigating neural processes underlying prototype-based categorization*

Prototype-distortion tasks evoke activity in a widespread network of brain regions, including occipital, (pre-)frontal and parietal regions but also subcortical areas like the striatum or the hippocampus (Bowman and Zeithamova, 2018; for a review, see Seger and Miller, 2010). The recruitment of specific brain areas in a given task is moderated by several factors. For example, hippocampal involvement appears to occur preferably during the first rather than later trials (Little et al., 2006) and is higher in A/B tasks compared with A/notA tasks (Zeithamova et al., 2008). Early visual areas are more involved in the categorization of simple shapes relative to more complex stimuli (Seger and Miller, 2010). Activity in the parietal lobe correlates more with receiving feedback than giving classification responses (Little et al., 2006). While these findings help to elucidate processes underlying similarity-based categorization in general, they do not permit to differentiate between neural correlates of the two categorization strategies.

Newer approaches combine functional magnetic resonance imaging (fMRI) with formal modeling, so that neural activity could be correlated to parameter estimates harvested from the two models. As expected, model-based fMRI speaks to differential neural correlates of the two strategies. In detail, similarity estimates derived from the prototype model correlated with activity in the ventromedial prefrontal cortex, the anterior hippocampus and the superior lateral occipital cortex (Bowman et al., 2020; Bowman and Zeithamova, 2018). Similarity estimates derived from the exemplar model correlated with activity in the posterior hippocampus, the lateral occipital cortex, and the parietal gyrus (Bowman et al., 2020; Bowman and Zeithamova, 2018).

Beyond investigating neural correlates of the two strategies, the combination of sophisticated multivariate fMRI analyses with formal modeling has been successfully used to discern prototype and exemplar strategies when behavioral modeling delivered inconclusive results (Mack et al., 2013). In this approach, adopted strategies are uncovered by assessing whether individual neuronal patterns either fit better with similarities predicted by a behavioral prototype or predicted by a behavioral exemplar model. In other words, while modeling of behavioral data relies only on endorsement probabilities, fMRI makes it possible to also take a more direct measure of category representation into account. A similar technique recently confirmed that individuals can also represent category structures in the form of a prototype and exemplars within a single task (Bowman et al., 2020).

Another exciting development in cognitive neuroscience is the usage of neural networks (NNs), which can be viewed as a specific class of formal models. The architecture of the NN mirrors a hypothesized chain of operations within biological boundaries and can mimic certain peculiarities in neuronal signaling. Single neurons are often represented as nodes within NNs, for which the behavior of their connections is modulated by learning rules (Yuste, 2015). In contrast to the formal modeling approaches for behavioral data from prototype distortion tasks described above, the application of NNs in this field does not aim at uncovering strategies underlying behavior, but focuses on understanding how specific neuronal characteristics (e.g., decreased synaptic plasticity) affect perceptual processing and learning in these tasks (Dovgopoly and Mercado, 2013; Mercado and Church, 2016). Moreover, while formal modeling relies on behavioral performances and distance measures, NNs are fed with mathematical descriptions of stimuli (e.g.

coordinates) and output simulated behavioural performances. Depending on the research goal, the validity of a network can be judged based on its abilities to solve certain classes of problems and/or to which degree the output matches performances of individuals from the general population or specific clinical groups. Therefore, without the acquisition of new empirical data, NNs can be used to test hypotheses on how peculiarities in neural signaling could explain patterns of behavioral performances from prototype-distortion tasks and to generate testable predictions what mechanisms could explain difficulties in prototype abstraction (Dovgopoly and Mercado, 2013; Mercado et al., 2015).

In sum, empirical studies with fMRI and theoretical approaches with NNs can provide insights into possible neural and cognitive processes underlying performance in prototype-distortion tasks. Moreover, model-based multivariate fMRI can be used as a highly sensitive tool to disentangle prototype and exemplar strategies on the individual level.

3. Performance on the prototype-distortion task among individuals with autism spectrum disorder

Having summarized the general research rationale and results about prototype based abstraction using prototype-distortion tasks, we now focus on studies that have used this task to study prototype-based abstraction in individuals with ASD.

3.1. Search strategy

To identify relevant articles, a Pubmed search using the string ((’autism AND prototype AND (learning OR category OR categorization)’ OR (autism and category learning)) was performed in June 2019 and again in December 2020, together yielding 266 articles. From these, 236 articles were excluded which either were not focused on visual category learning or on ASD. From the remaining 30 articles, three articles were excluded because they used a category task in the learning phase, but focused on recognition memory rather than categorization during the test phase[1] (Gastgeb et al., 2011, 2009; Molesworth et al., 2005). Three additional articles were excluded on account of being review articles (Fields, 2012; Gastgeb et al., 2012; Mercado et al., 2020). From the remaining 24 articles, 11 did not relate endorsement probabilities to distance from the prototype or provide any other information suitable to infer whether participants could have used a prototype strategy. These articles are not part of the current review as they do not speak directly to prototype abstraction but are listed in Supplementary Table 1A for transparency. References from the 13 remaining articles were screened for further relevant research, yielding a book chapter reporting original findings not published elsewhere (Klinger et al., 2006), leading to a final total of 14 articles. A doctoral thesis (Meyer, 2014), referenced by the review article of Mercado et al. (2020), was not included but added to Supplementary Table 1, because it used a category task in the learning phase, but a recognition memory task in the test phase.

Table 1A details the articles included in our overview of prototype abstraction among ASD individuals. It is split up into three sections: Table 1A summarizes 11 experiments, based on 10 different articles, comparing behavioral (model free) performances of ASD and neurotypical (i.e. individuals without ASD; NT) individuals in prototype-distortion tasks, with one article contributing two experiments (Vladusich et al., 2010). Table 1B summarizes two articles which use formal modeling, both applied to the same data set reported by Church et al. (2010), which is detailed in Table 1A. Table 1C contains 5 experiments, based on 5 articles, addressing potential neuronal and cognitive mechanisms behind difficulties in prototype-based categorization in prototype-distortion tasks in ASD. The article of Church et al. (2015) has been included in Table 1C because it focuses exclusively on individuals with ASD and empirically tests a theoretical prediction stated by Dovgopoly and Mercado (2013). The article of Church et al. (2010) is included in Table 1A as well as 1B, because it includes model-free and model-based analyses. The article of Schipul and Just (2016) occurs in both Tables 1A and 1C, because it reports both behavioral and neuronal findings. The article of Mercado et al. (2015) is listed in Tables 1A and 1C, because it contains results from behavioral testing as well as simulations.

3.2. Results overview

3.2.1. Performance in the prototype-distortion task in ASD

As shown in Table 1A, articles differ with respect to the choice of outcome measures. Authors from the 10 articles reporting results from model-free analyses use different measures to investigate whether individuals with ASD experience more problems to perform prototype-based categorization. In particular, articles focused on prototype effects (Church et al., 2010; Gastgeb et al., 2012; Klinger et al., 2006; Klinger and Dawson, 2001; Mercado et al., 2015; Molesworth et al., 2008), typicality gradients (Church et al., 2010; Gastgeb et al., 2012; Mercado et al., 2015), overall performance (Church et al., 2010; Gastgeb et al., 2012; Mercado et al., 2015; Vladusich et al., 2010), and speed of learning (Schipul and Just, 2016; Vladusich et al., 2010). Some authors included additional measures, such as reaction times (Schipul and Just, 2016) and accuracy for medium and high distorted category members (Froehlich et al., 2012) or investigated gaze patterns (Gastgeb et al., 2012), but since none of them provides further insights into the ability to perform prototype-based categorization, there are included in our review for completeness only and will not be discussed further.

Table 1A shows that three experiments did not yield differences between ASD and NT groups for any of the primary outcomes considered by the authors (Froehlich et al., 2012; Tager-Flusberg, 1985; Vladusich et al., 2010, Experiment 2). In three experiments, there was evidence for behavioral differences for some but not all primary outcomes (Molesworth et al., 2008; Schipul and Just, 2016; Vladusich et al., 2010, Experiment 1). In the five remaining experiments, ASD participants showed different performance than TD participants for all included primary outcomes. Measures for which some experiments did not reveal differences between ASD and NT include prototype effects (Froehlich et al., 2012; Tager-Flusberg, 1985), typicality gradients (Froehlich et al., 2012; Vladusich et al., 2010, Experiments 1 & 2), total accuracy (Schipul and Just, 2016; Tager-Flusberg, 1985; Vladusich et al., 2010, Experiment 2) and learning speed (Vladusich et al., 2010, Experiment 2). Particularly remarkable is the inconsistent pattern of results in the two nearly identical dot patterns categorization tasks reported by Vladusich et al. (2010). Here, worse overall performance and slower learning in the first experiment could not be replicated in a second one with a partially overlapping autistic sample, only differing with respect to subtle changes in dot pattern construction.

Overall, the contradictory results in Table 1A paint a blurry picture of performance differences between both groups in the prototype distortion task. However, it might be possible to make sense of the mixed results by considering the task characteristics. As explained above, task characteristics biasing more towards a prototype relative to an exemplar strategy include the use of only a single category and a low memorizability of training exemplars. A closer look at Table 1A reveals several differences in experimental designs having the potential to affect whether individuals tend to prefer the prototype or the exemplar strategy. If single category designs and low memorizability (e.g., with a large

[1] In these articles, the test phase asked whether participants believed a stimulus such as the prototype has been shown during the training phase. The degree of prototype abstraction during training is assumed to be represented as the height of the false alarm rate for the prototype. However, newer findings question whether this is a valid measure of prototype abstraction, as most of the participants apparently rely on their memory for single training exemplars and not on an abstract representation of the unseen prototype (Bowman and Zeithamova, 2020).

Table 1A
Performance on prototype-distortion task among autistic and neurotypical individuals (in descending order by publication date).

Reference	Sample	Stimuli	Task type	Number of training exemplars	Distortion levels of training exemplars	Training phase	Outcome Measure(s)	Key Findings
Schipul and Just, 2016	$n = 16$ H F ASD & $n = 16$ N T adults; mainly male	dot patterns	A/ notA	5	6	passive viewing, alternating with test blocks	number of training blocks, error rates, reaction times	- number of training blocks: ASD > NT ASD - error rates: ASD ≈ NT
Mercado et al., 2015	$n = 13$ H F ASD & $n = 13$ TD children; mainly male (partly overlapping with sample of Church et al., 2010)	polygons created from dot patterns	A/ notA	Sit. A: 15−30 Sit. B: 15	3, 5, 7	Sit. A: individual testing; 4 modified task versions Sit. B: group testing; training similar to Church et al., 2010	endorsement of random shapes (’false alarms’), prototype effect, typicality gradient	ASD subgrouped according to endorsement of random shapes: ASD-I < 30 %, ASD-II > = 30 % - prototype effect: ASD-II < NT & ASD-I - typicality gradient: ASD-II < NT & ASD-I - no main effects or interactions for different task versions - substantial intra-individual variation across tasks in individuals with ASD
Gastgeb et al., 2012	$n = 20$ H F ASD & $n = 19$ N T adults; all male	dot patterns	A/ notA	40	5	passive viewing	prototype effect, typicality gradient, accuracy rates	- prototype effect: ASD < NT - typicality gradient: ASD < NT - total accuracy: ASD < NT
Froehlich et al., 2012	$n = 24$ H F ASD & $n = 25$ N T adults; all male	dot patterns	A/B/C	5	6	passive viewing, alternating with test blocks	prototype effect, typicality gradient, accuracy rates	- prototype effect: ASD ≈ NT - typicality gradient: ASD ≈ NT
Church et al., 2010	$n = 20$ H F ASD & $n = 20$ TD children; mainly male	polygons created from dot patterns	A/ notA	15	3, 5, 7	trial-and-error,	prototype effect, typicality gradient, accuracy rates	- prototype effect: ASD < NT - typicality gradient: ASD < NT - total accuracy: ASD < NT
Vladusich et al., 2010	Exp.1: $n = 15$ H F ASD & $n = 21$ N T Exp.2: $n = 13$ H F ASD (9 from exp. 1) & $n = 18$ N T adolescents & adults; all male	dot patterns	A/B	16	~5[1]	passive viewing of examples then trial-and-error, training until criterion reached	typicality gradient, number of training blocks, accuracy rates (prototype not shown during test phase)	Exp.1: - typicality gradient: ASD ≈ NT - number of training blocks: ASD > NT - total accuracy: ASD > NT Exp.2: - typicality gradient: ASD ≈ NT - number of training blocks: ASD ≈ NT - total accuracy: ASD ≈ NT - prototype effect: not significant in ASD group
Molesworth et al., 2008	$n = 20$ H F ASD & $n = 20$ TD children; mainly male	imaginary animals (6 continuous features)	A/B	6	–	trial-and-error	prototype effect, accuracy rates	- accuracy rates for individual stimulus classes: ASD ≈ TD - high heterogeneity among individuals with ASD
Klinger et al., 2006	$n = 50$ H F ASD & $n = 50$ TD children & teenagers; mainly male	imaginary animals (4 discrete features)	A/ notA	8	–	trial-and-error	prototype effect	- prototype effect: ASD < TD
Klinger and Dawson, 2001	$n = 12$ L F ASD, $n = 12$ DS & $n = 12$ TD children & adolescents; mainly male	imaginary animals (4 discrete features)	A/ notA	8	–	trial-and-error	prototype effect	- prototype effect: ASD & DS < TD - only TD above chance prototype endorsement
Tager-Flusberg, 1985[2]	$n = 14$ L F ASD, $n = 14$ ID & $n = 14$ TD children; mainly male	line-drawn pictures of everyday categories	A/B/C & A/ B/C/D	5	–	trial-and-error (matching to sample)	prototype effect, accuracy	- prototype effect: ASD ≈ ID ≈ TD - accuracy: ASD ≈ ID ≈ TD

Table 1B
Formal modeling based analyses.

Reference	Sample	Stimuli	Task type	Number of training exemplars	Distortion levels of training exemplars	Training phase	Outcome Measure(s)	Key Findings
Voorspoels et al., 2018	see Church et al., 2010 in Table 1A						prototype model fit, mixture model fits, parameter estimates (criterion, sensitivity, latent group indicator)	Analysis 1 (Hierarchical prototype model): - model fit: ASD worse than TD - criterion: ASD > NT - sensitivity: ASD < NT Analysis 2 (Hierarchical mixture prototype plus guessing models): - latent group indicator: 1/20 individuals from NT, 8/20 from ASD assigned to guessing strategy - model fit/sensitivity: data not sufficiently informative to decide whether non-guessing individuals from TD and ASD groups differ in sensitivity
Church et al., 2010	see Church et al., 2010 in Table 1A						prototype model fit, parameter estimates (criterion & sensitivity)	- fit of prototype model: ASD worse than TD - criterion: ASD > NT - sensitivity: ASD < NT

number of stimuli) facilitate the adoption of a prototype strategy, and if ASD individuals are impaired on their prototype strategy but not on their exemplar strategy, one would expect to observe performance differences in experiments with these design choices, and less so in experiments where the exemplar strategy is more advantageous[2] .

Consistent with the conjecture that the presence of a single category makes it more likely to reveal performance differences between both groups, most consistent evidence for impaired performances in ASD relative to NT groups comes from A/notA tasks. A total of 6 out of 11 experiments investigating prototype-based categorization in ASD employed A/notA tasks, whereas the remaining 5 experiments used two or more contrasting categories (i.e. A/B, A/B/C or A/B/C/D tasks). Of the six A/notA experiments, five showed lower performance in individuals with ASD compared with NT groups on all included primary outcome measures (Church et al., 2010; Gastgeb et al., 2012; Klinger et al., 2006; Klinger and Dawson, 2001; Mercado et al., 2015). The one A/notA experiment conducted by Schipul and Just (2016) is an exception, which revealed a group difference in number of training blocks, but not in error rates. It is interesting to note that this study used the lowest number of training exemplars of all the included A/notA tasks, so that an exemplar strategy might have been quite useful. Turning to the 5 experiments employing multiple categories, three of these report no difference in performance in any of the employed outcome measures (Froehlich et al., 2012; Tager-Flusberg, 1985; Vladusich et al., 2010, Experiment 2), which is consistent with the conjecture about the role of the number of categories. One multiple-category study reports differences for some but not all primary outcomes (Vladusich et al., 2010, Experiment 1). Again, the mixed result for this experiment might be a reflection of the fact that this study used the largest category of all multiple category studies. A study less consistent with this reasoning is the study of Molesworth et al. (2008), where strong prototype effects were present in the group of typically developing (TD) children despite using an A/B design and a small category size. It remains unclear whether any other study characteristic (e.g. the specific stimulus set or the young age of the participants) could have led to an amplified tendency towards using a prototype strategy in the TD group or whether the 'prototype effect' is an artefact of stimulus set characteristics (e.g. saliency of prototypical stimuli). Assuming that the TD group indeed used the prototype strategy, the fact that children with ASD did not exhibit a significant prototype effect while showing similar accuracy rates as the TD group would go well with a successful compensation of difficulties in prototype abstraction by employing the exemplar strategy in the ASD group.

Regarding the memorizability of training exemplars, we have already noted that the only A/notA experiment where no performance difference has been detected employed the lowest number of training exemplars of all the included A/notA tasks (5 in Schipul and Just, 2016), so that an exemplar strategy might have been quite useful. Conversely, one of the two A/B designs in which group differences were found used the highest number of training exemplars of all the included multiple categories tasks, so that a prototype strategy might have been quite useful (16 in Vladusich et al., 2010, Experiment 1). Further factors which can moderate the memorizability such as the usage of more concrete stimuli like line-drawn every objects (Tager-Flusberg, 1985) and imaginary animals (Klinger et al., 2006; Klinger and Dawson, 2001; Molesworth et al., 2008) versus abstract patterns (Church et al., 2010; Froehlich et al., 2012; Gastgeb et al., 2012, 2012; Vladusich et al., 2010) do not show a clear picture. Therefore, while there is a suggestion that the usage of a small number of training exemplars could make it more unlikely to detect ASD-related difficulties in prototype abstraction, there is no indication for the choice of stimulus types to be influential.

The included experiments do not only differ with respect to the employed categorization tasks, but also with respect to sample characteristics such as age and functional levels. Of the 6 experiments where autistic and TD children took part (Church et al., 2010; Klinger et al., 2006; Klinger and Dawson, 2001; Mercado et al., 2015; Molesworth et al., 2008; Tager-Flusberg, 1985), all show at least one potential indicator of ASD-related difficulties in prototype abstraction. In the remaining 5 experiments with adolescent or adult samples (Froehlich et al., 2012; Gastgeb et al., 2012; Schipul and Just, 2016; Vladusich et al., 2010 Experiment 1 & 2), results are more mixed. Further, nearly all experiments focused on high-functioning people with ASD, and the two experiments which included low functioning individuals with ASD delivered mixed findings (Klinger and Dawson, 2001; Tager-Flusberg, 1985). Because experiments included either pure or predominantly male samples, our overview does not afford conjectures about the role of gender. In summary, it seems that difficulties in prototype-distortion tasks are more likely to occur at a younger age. Whether this can be better explained by developmental trajectories in prototype abstraction or the ability to recruit compensatory strategies remains unclear.

In sum, among the experiments which favor prototype-based categorization, most reveal at least some evidence for an ASD-related performance difference: Those experiments which used only a single

[2] This conjecture should be considered exploratory, as it was derived based on the collection of empirical results. We do not provide a confirmatory test for this conjecture.

Table 1C
Neuronal and cognitive processes underlying difficulties in prototype-based categorization.

Reference	Sample	Stimuli	Task type	Number of training exemplars	Distortion levels of training exemplars	Training phase	Outcome Measure (s)	Key Findings
Mercado and Church, 2016	–	vectors describing stimuli from Mercado et al., 2015 & 3 novel polygon sets	A/ notA	–		category-dependent & randomly transformed vectors	reproduction of result patterns from Mercado et al., 2015	- training with category-dependent transformations: reliably good learning across stimulus sets similar to NT and ASD-I in Mercado et al., 2015 - training with random transformations: unpredictable, intra-individual fluctuations in learning across stimulus sets similar to ASD-II in Mercado et al., 2015
Schipul and Just, 2016	see Schipul and Just, 2016 in Table 1A						brain activity	- activity in parietal and occipital areas decreased over time in the NT group - activity in frontal, temporal, and parietal regions increased in the ASD group - functional connectivity between brain regions increased in the NT but not the ASD group
Church et al., 2015	n = 43 H F ASD children	polygons created from dot patterns	A/ notA	Prototype-training (PT): 1 Distortions-training (DT): 15	PT: 0 DT: 3, 5, 7	trial-and-error PT: prototype 15 times repeated DT: prototype not shown	typicality gradient, accuracy prototype model fit, parameter estimates (criterion & sensitivity)	ASD subgrouped according to endorsement of random shapes after DT: ASD-I < 30 %, ASD-II > = 30 % - typicality gradient in ASD-I: PT < DT - typicality gradient in ASD-II: PT > DT - accuracy in ASD-I: PT < DT - accuracy in ASD-II: PT > DT - model fit after PT: ASD-I ≈ ASD-II - model fit after DT: ASD-I better than ASD-II - sensitivity after PT: ASD-I ≈ ASD-II - sensitivity after DT: ASD-I > ASD-II
Mercado et al., 2015	- (see Mercado et al., 2015 Table 1A for the behavioral experiment)	vectors describing polygons from Church et al., 2010 & 4 novel polygon sets	A/ notA	–			reproduction of result patterns from behavioral experiment	- replication of findings from Dovgopoly and Mercado, 2013 - intra-individual fluctuations only reproducible by neuronal network assuming short-term variations in neuronal plasticity - training with repetitions of prototype instead of multiple distortions could aid low-performing ASD individuals
Dovgopoly and Mercado, 2013	–	vectors describing stimuli from Church et al., 2010 & Vladusich et al., 2010	A/ notA	–			reproduction of result patterns from Church et al. (2010) and Vladusich et al. (2010)	simulation of reduced neuronal plasticity reproduced behavioral patterns reasonably well when subgrouping ASD group into ASD-I and ASD-II (see Church et al., 2015)

Abbreviations: ASD = autism spectrum disorder. HF = high functioning. LF = low functioning. DS= down syndrome. TD = typically developing children. NT = neurotypical adults. ID = intellectually disabled. A/notA task: only one target category. A/B(/C/D) task: two or more categories. The column 'exemplar number' refers to the number of idiosyncratic stimuli shown during the study phase for each category.

Meaning of symbols in results column: Symbol '<' = smaller at a significance level of p <.05 or below. Symbol '>' = bigger at a significance level of p <.05 or below. Symbol '≈' = no statistically significant difference.

[1] The article describes the distortion level of the training set as 'medium', which likely equals a distortion level around 5. [2] Experiment 3, which investigated lexical knowledge, is not reported here.

category and a high number of training exemplars were more likely to show ASD-related difficulties in categorization. Further, among the experiments which favor exemplar-based categorization (those experiments which used multiple categories and a low number of training exemplars), most do not reveal evidence for such a difference. This observation is consistent with the assumption of impaired prototype abstraction in ASD but intact ability to use the exemplar strategy.

3.2.2. *Formal modeling based analyses to investigate prototype abstraction in ASD*

Prototype effects and typicality gradients in prototype-distortion tasks have often been taken as a reflection of the ability to abstract prototypes. However, as noted above, formal modelling has revealed that this line of reasoning is of limited value only. Rather than inferring people's strategy use from relatively coarse behavioral measures, an alternative approach involves testing different formal models that instantiate these different strategies based on detailed information about endorsement frequencies and distances to prototypes or exemplars.

Our search revealed no publication in which the prototype and exemplar models were directly compared against each other in individuals with ASD within the context of category learning. Table 1B summarizes the two articles in which a formal prototype model was compared between individuals with and without ASD (Church et al., 2010; Voorspoels et al., 2018), both based on the same data set. In the experiment from which the data set originated (Church et al., 2010), children (aged 7–12 years) with ($n = 20$) and without ASD ($n = 20$; TD) performed a prototype-distortion task with abstract shapes created from dot patterns. Via trial-and-error, participants had to learn which of 15 out of 30 stimuli belonged to a specific category ('cave ghosts'). The prototype, 30 new category members and 30 random stimuli were presented in a test phase with no feedback. As outlined above, model-free analyses yielded worse overall performance and lower endorsement rates for the prototype in autistic children. Additionally, Church et al. (2010) fit a prototype model to their data, and found that individual fits of the prototype model were significantly better for the TD compared to the ASD group. Moreover, relative to the ASD group, TD individuals showed lower criterion parameters, as well as a higher sensitivity to distances from the prototype. This was taken as support for the hypothesis that performance differences could be related to ASD-related deficits in prototype-based categorization.

These data were re-analyzed using a Bayesian hierarchical (latent mixture) approach by Voorspoels et al. (2018; Table 1B). Their first analysis, designed to closely mimic the analyses from Church et al. (2010), confirmed Church et al.'s conclusion, in that lower sensitivity parameters were found in the ASD compared to the TD group. In particular, the highest density interval for group differences in sensitivity parameters yielded a probability of 95 % for lower sensitivity to the prototype in the ASD group. Moreover, the Bayes factor for the comparison of models either assuming or not assuming group differences in sensitivity strongly favored the first model. However, while this analysis showed relatively homogenous sensitivity estimates in the TD group, the sensitivity estimates in the ASD group clustered into one subgroup with sensitivity estimates similar to the NT group and another subgroup with much lower sensitivity estimates (see Figure 3 in Church et al., 2010).

To further investigate the occurrence of potential ASD subgroups in sensitivity, a hierarchical latent mixture model was applied (Voorspoels et al., 2018; Analysis 2), allowing for discrete in-group differences (see also Bartlema et al., 2014). This approach incorporated a binary latent group indicator variable, which assigned participants either to a prototype or a simple guessing model. In other words, model estimation classified participants depending on whether their performance was better described by a prototype or by a guessing strategy in a data-driven manner. One out of 20 participants from the TD and 8 out of 20 participants from the ASD group were assigned to the guessing model, signaling that for these participants, the prototype model was most likely not an appropriate model, making interpretation of its parameters questionable, at best. Restricting estimates to those participants for which the prototype model was more appropriate than a simple guessing model, evidence for general differences in sensitivity between TD and the ASD group as a whole was much less clear. Although the 95 %-highest density interval for sensitivity differences spanned mostly negative values (indicating decreased sensitivity for ASD groups), it covered positive values as well, so that it remains uncertain whether meaningful group differences in sensitivity to the prototype exist. Moreover, the Bayes factor indicated that data were not sufficiently informative for deciding whether the model assuming sensitivity differences between the groups accounted better for the data.

3.2.3. *Potential neural and cognitive processes behind difficulties in prototype-based categorization in ASD*

Turning now to Table 1C, we see that a variety of methodological approaches have been employed to characterize potential mechanisms underlying ASD-related difficulties in prototype-distortion tasks. Schipul and Just (2016) studied neural activity associated to a prototype-distortion task in individuals with and without ASD using fMRI. In this experiment, participants passively learned category membership through presentation of category labels for each dot pattern and were then tested with corrective feedback in alternating encoding-test blocks. Training was performed outside the MRI and continued until participants reached a performance criterion of 70 %. A novel stimulus set was used for encoding-test blocks in the MRI. Encoding was associated to brain regions typically involved in performing prototype-distortion tasks (Seger and Miller, 2010), with no differences between groups. However, while the recruitment of parietal and occipital areas decreased through the course of learning in the NT group, the recruitment of frontal, temporal, and parietal regions increased over time in the ASD group. In addition, functional connectivity between brain regions increased in the NT but not the ASD group. The experiment was neither designed to discern categorization strategies nor to investigate group differences in neural activity associated to stimulus-wise characteristics such as distance to the prototype. Therefore, it remains unclear whether group differences in neural adaptation could be related to difficulties in prototype abstraction in the ASD group.

Dovgopoly & Mercado (2013) employed NNs to investigate which of the known neuronal characteristic of ASD might underlie difficulties in prototype-based categorization. NNs of visual cortical processing were modified to simulate reduced neural plasticity, reduced neural homeostasis (i.e. the regulation of the amount of synaptic change), increased neural noise and an increased number of mini-columns in the occipital cortex. Networks were fed with stimulus sets originally used by Church et al. (2010) and Vladusich et al. (2010), in order to investigate which of the modifications would be able to reproduce behavioral patterns. Analyses revealed that only the NN mimicking reduced neural plasticity showed a reasonable overlap with empirical data, particularly when ASD groups were split into low- and high performers. This finding could be replicated for another data set by a later simulation experiment (Mercado et al., 2015). Further simulations performed by Dovgopoly & Mercado (2013) suggest that individuals with ASD could benefit from showing only the prototypical stimulus during training, while NT individuals would learn better when being trained with multiple distorted versions of the prototype.

The idea that ASD-related deficits in performing prototype-distortion tasks can be overcome by presenting only the prototypical stimulus as a training exemplar has been addressed by an experiment conducted by Church et al. (2015). Children with ASD performed one prototype-distortion task with a standard training set of distorted category members, and a second task with a training set that consisted of repetitions of the prototype. Based on their performance in the standard training conditions, autistic children were subdivided into low- and a high-performers. Interestingly, only those autistic children who were poor learners in the standard task benefitted from the prototype-only training compared to the standard training regimen. Although a benefit of reducing training sets to only the prototype might seem counterintuitive at first sight, it makes sense when considering that the exclusive presentation of the prototype as the ideal member category A makes it obsolete to abstract prototypical features from multiple exemplars. These findings go well with the assumption that at least some individuals with ASD have difficulties to abstract prototypes and therefore rely to a great extent on their experience with exemplars.

While a good proportion of individuals with ASD has severe problems to solve prototype-distortion tasks, other autistic individuals perform similar to NT individuals (Church et al., 2015; Mercado et al., 2015; Molesworth et al., 2008; Voorspoels et al., 2018). Moreover, individuals with ASD show high intra-individual fluctuations in their performances (Mercado et al., 2015; Vladusich et al., 2010). Using a NN simulating variations in neural plasticity over relatively short periods, Mercado et al. (2015) were able to reproduce the high intra-individual variation observed in behavioral experiments. The authors state that although rapidly changing levels of neural plasticity in ASD might be a highly speculative assumption, it is consistent with the observation of dysfunctional cholinergic modulatory systems in ASD (Perry et al., 2001). In another simulation experiment, Mercado and Church (2016) hypothesized that individuals with ASD might encode visual stimuli in an idiosyncratic way, causing some random category-unrelated features to become highly salient. As a consequence, category learning can go well with one stimulus set but be heavily impaired for another, only slightly different stimulus set. To test this, a number of NN's, each representing the performance of an individual participant, were trained with two different transformations of four sets of abstract shapes. Category-dependent transformations, conducted via principal component analyses to reduce stimulus dimensionality but preserve category-relevant variations, were used to simulate learning in TD children. Idiosyncratic transformations, performed by random projection to amplify random features, was used to simulate learning in ASD children. As expected, TD NNs could easily and reliably learn to distinguish distorted category members from non-members in all different stimulus sets. In contrast, categorization accuracy for half of the ASD NNs varied unpredictably across stimulus sets, while the other half performed as well as the TD NNs in all sets. In sum, ASD-specific fluctuations in neural plasticity as well as idiosyncratic transformations are candidate mechanisms underlying the high heterogeneity in autistic samples. The representation of stimuli in an idiosyncratic way could be a plausible mechanism behind difficulties in prototype abstraction in ASD, as this requires to put an emphasis on within-category similarities and ignore irrelevant details. In contrast, idiosyncratic representations should not infer with an exemplar-based strategy.

In summary, neuroimaging data suggest that although individuals with and without ASD recruited similar brain regions during prototype-distortion tasks, neuronal activity in several brain areas as well as functional connectivity adapted differently in individuals with ASD. Simulations predict that an ASD-specific decrease in synaptic plasticity, short term fluctuations in synaptic plasticity as well as idiosyncratic representations of stimuli could underlie could difficulties in prototype-based categorization.

4. Discussion

Categorization is a vital cognitive skill, allowing us to structure the world and apply knowledge quickly to new situations. Several categorization strategies involve computing category-related similarities. An efficient strategy for similarity-based categorization is to abstract and store the categories central tendency, known as prototype. As an alternative strategy, the exemplar strategy involves storing category members as separate representations.

It has been suggested that prototype-based categorization, but not exemplar-based categorization is impaired among ASD individuals. The detail-focused cognitive style in individuals with ASD has been hypothesized to impair categorization based on prototype abstraction, because abstraction requires emphasizing the similarities associated with category membership and ignoring irrelevant differences between category members. In contrast, no such interference is assumed for memorizing exemplars, so that individuals with ASD could use the exemplar strategy to compensate deficits in prototype abstraction.

Further, research using prototype-distortion tasks in non-clinical populations has suggested that certain task characteristics of prototype-distortion tasks favor the adoption of one of the two strategies. Together, this means that one can expect performance differences between ASD and TD individuals for those designs which favor a prototype strategy, but not for those which favor an exemplar strategy.

Our review of 11 experiments using a prototype distortion task showed that most consistent evidence for ASD-related difficulties in prototype abstraction stems from experimental designs fostering a strong bias towards the prototype strategy.

Only a few studies have used formal modeling to understand behavior on prototype distortion tasks. They yielded clear evidence for severe difficulties in performing a prototype-distortion task in a subgroup of autistic children, but inconclusive findings for a remaining sample of children with ASD. As no published article yet compared the fits between the prototype and exemplar models in ASD, it remains unclear to which extent ASD-related difficulties in prototype abstraction could be masked by a compensatory exemplar strategy. Moreover, the usage of new approaches such as model-based multivariate fMRI could further increase the sensitivity of discerning the two strategies and deliver important insight into underlying neuronal mechanisms.

Existing studies which aimed at elucidating potential neuronal and cognitive processes underlying ASD-related difficulties in prototype abstraction are scarce. The only experiment on neural correlates of performing a prototype-distortion task found atypical patterns of neuronal adaptation and functional connectivity in the course of learning in individuals with ASD. Simulations by NNs propose that decreased neural plasticity might explain potential difficulties in prototype-based categorization and short-term fluctuations in neural plasticity could account for inconsistent performances observed in autistic samples (Dovgopoly and Mercado, 2013; Mercado et al., 2015). In fact, alterations in the regulation of synaptic plasticity have been evidenced in mouse models and some of the known ASD-risk genes are involved in synaptic plasticity (Bourgeron, 2015; Hansel, 2019). While the direct assessment of indicators of neural plasticity is largely not feasible in humans, one could investigate whether autistic individuals with and without difficulties in prototype abstraction specifically differ with respect to genes coding for synaptic plasticity. Moreover, the combination of transcranial magnetic stimulation with fMRI can been used to study neuronal plasticity in humans (Pascual-Leone et al., 2011). On the cognitive level, NN simulations indicate that hyper-specific, idiosyncratic representations of stimuli could impair prototype abstraction in an unpredictable and unreliable manner (Mercado and Church, 2016). This could, for instance, be further investigated by assessing neuronal representations via multivariate fMRI approaches such as representational similarity analyses (Mack et al., 2013), having the potential to uncover differences in the representation of category structures between autistic versus NT individuals not observable on the behavioral level.

It is a common observation that samples with autistic individuals exert a high degree of inter-individual or even intra-individual variability in their ability to perform prototype-distortion tasks (Mercado et al., 2015; Molesworth et al., 2008; Voorspoels et al., 2018). Accordingly, NN simulations showed the best overlap with empirical data when ASD groups were sub grouped into good and poor learners (Church et al., 2015; Dovgopoly and Mercado, 2013; Mercado et al., 2015). Although some attempts have been made to find stable characteristics which distinguish autistic individuals with good and poor performances in prototype-distortion tasks, there is little consensus in the literature.

Candidate characteristics, for instance, are the non-verbal IQ (Vladusich et al., 2010), the verbal IQ (Klinger and Dawson, 2001) or a general difficulties in processing the typicality of stimuli (Molesworth et al., 2008). Discerning prototype and exemplar strategies at the individual level could be a fruitful venue to investigate whether stable individual characteristics directly impair prototype abstraction and/or influence the use of a compensatory exemplar strategy in ASD.

Beyond individual difference in basic cognitive abilities among autistic individuals, a factor contributing to the high variability might be that rather of having a deficit in prototype abstraction, individuals with ASD might be simply biased against using the prototype strategy despite being capable of using so. Depending on (potentially interacting) factors such as individual motivation or stress level, task design and instructions, this bias might manifest in behavior or not. This matter has been also debated with respect to the theory of weak central coherence, in that the detail-focused processing of individuals with ASD might not reflect a limited ability but a specific cognitive style (Happé and Frith, 2006). Future studies could directly test, whether it could be possible to specifically instruct or train subjects to use the prototype strategy and, as a next step, whether this could be useful for everyday functioning.

In conclusion, the current review highlights that although some evidence exists that individuals with ASD might be specifically impaired in prototype-based categorization, this claim warrants further investigation using methodological tools such as formal modeling, developed by basic research, to uncover strategies usually hidden in conventional analyses. Deeper knowledge of underlying cognitive deficits could provide testable predictions how individuals with ASD could be aided to overcome potential learning difficulties and inform the optimization of trainings and therapeutic interventions.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This work was supported by the German Research Foundation [DFG BA 6601/2-1].

References

Aizenstein, H.J., MacDonald, A.W., Stenger, V.A., Nebes, R.D., Larson, J.K., Ursu, S., Carter, C.S., 2000. Complementary category learning systems identified using event-related functional MRI. J. Cogn. Neurosci. 12, 977–987. https://doi.org/10.1162/08989290051137512.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Antzoulatos, E.G., Miller, E.K., 2011. Differences between neural activity in prefrontal cortex and striatum during learning of novel, abstract categories. Neuron 71, 243–249. https://doi.org/10.1016/j.neuron.2011.05.040.

Ashby, F.G., Maddox, W.T., 1993. Relations between prototype, exemplar, and decision bound models of categorization. J. Math. Psychol. 37, 372–400.

Bartlema, A., Lee, M., Wetzels, R., Vanpaemel, W., 2014. A Bayesian hierarchical mixture approach to individual differences: case studies in selective attention and representation in category learning. J. Math. Psychol. 59, 132–150.

Bourgeron, T., 2015. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. 16, 551–563. https://doi.org/10.1038/nrn3992.

Bowman, C.R., Iwashita, T., Zeithamova, D., 2020. Tracking prototype and exemplar representations in the brain across learning. eLife 9, e59360. https://doi.org/10.7554/eLife.59360.

Bowman, C.R., Zeithamova, D., 2018. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. J. Neurosci. 38, 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018.

Bowman, C.R., Zeithamova, D., 2020. Training set coherence and set size effects on concept generalization and recognition. J. Exp. Psychol. Learn. Mem. Cogn. https://doi.org/10.1037/xlm0000824.

Bozoki, A., Grossman, M., Smith, E.E., 2006. Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia 44, 816–827. https://doi.org/10.1016/j.neuropsychologia.2005.08.001.

Casale, M.B., Ashby, F.G., 2008. A role for the perceptual representation memory system in category learning. Percept. Psychophys. 70, 983–999.

Church, B.A., Krauss, M.S., Lopata, C., Toomey, J.A., Thomeer, M.L., Coutinho, M.V., Volker, M.A., Mercado, E., 2010. Atypical categorization in children with high-functioning autism spectrum disorder. Psychon. Bull. Rev. 17, 862–868. https://doi.org/10.3738/PBR.17.6.862.

Church, B.A., Rice, C.L., Dovgopoly, A., Lopata, C.J., Thomeer, M.L., Nelson, A., Mercado, E., 2015. Learning, plasticity, and atypical generalization in children with autism. Psychon. Bull. Rev. 22, 1342–1348.

Cook, R.G., Smith, J.D., 2006. Stages of abstraction and exemplar memorization in pigeon category learning. Psychol. Sci. 17, 1059–1067. https://doi.org/10.1111/j.1467-9280.2006.01833.x.

Davis, T., Love, B.C., 2010. Memory for category information is idealized through contrast with competing options. Psychol. Sci. 21, 234–242. https://doi.org/10.1177/0956797609357712.

Dovgopoly, A., Mercado, E., 2013. A connectionist model of category learning by individuals with high-functioning autism spectrum disorder. Cogn. Affect. Behav. Neurosci. 13, 371–389. https://doi.org/10.3758/s13415-012-0148-0.

Etcoff, N.L., Magee, J.J., 1992. Categorical perception of facial expressions. Cognition 44, 227–240.

Fields, C., 2012. Do autism spectrum disorders involve a generalized object categorization and identification dysfunction? Med. Hypotheses 79, 344–351. https://doi.org/10.1016/j.mehy.2012.05.032.

Froehlich, A.L., Anderson, J.S., Bigler, E.D., Miller, J.S., Lange, N.T., DuBray, M.B., Cooperrider, J.R., Cariello, A., Nielsen, J.A., Lainhart, J.E., 2012. Intact prototype formation but impaired generalization in autism. Res. Autism Spectr. Disord. 6, 921–930. https://doi.org/10.1016/j.rasd.2011.12.006.

Gastgeb, H.Z., Rump, K.M., Best, C.A., Minshew, N.J., Strauss, M.S., 2009. Prototype formation in autism: can individuals with autism abstract facial prototypes? Autism Res. 2, 279–284. https://doi.org/10.1002/aur.93.

Gastgeb, H.Z., Wilkinson, D.A., Minshew, N.J., Strauss, M.S., 2011. Can individuals with autism abstract prototypes of natural faces? J. Autism Dev. Disord. 41, 1609–1618. https://doi.org/10.1007/s10803-011-1190-4.

Gastgeb, H.Z., Dundas, E.M., Minshew, N.J., Strauss, M.S., 2012. Category formation in autism: can individuals with autism form categories and prototypes of dot patterns? J. Autism Dev. Disord. 42, 1694–1704. https://doi.org/10.1007/s10803-011-1411-x.

Hansel, C., 2019. Deregulation of synaptic plasticity in autism. Neurosci. Lett. 688, 58–61. https://doi.org/10.1016/j.neulet.2018.02.003.

Happé, F., Frith, U., 2006. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J. Autism Dev. Disord. 36, 5–25. https://doi.org/10.1007/s10803-005-0039-0.

Homa, D., Sterling, S., Trepel, L., 1981. Limitations of exemplar-based generalization and the abstraction of categorical information. J. Exp. Psychol. Hum. Learn. 7, 418–439. https://doi.org/10.1037/0278-7393.7.6.418.

Homa, D., Proulx, M.J., Blair, M., 2008a. The modulating influence of category size on the classification of exception patterns. Q. J. Exp. Psychol. 61, 425–443. https://doi.org/10.1080/17470210701238883.

Homa, D., Proulx, M.J., Blair, M., 2008b. The modulating influence of category size on the classification of exception patterns. Q. J. Exp. Psychol. 61, 425–443. https://doi.org/10.1080/17470210701238883.

Klinger, L.G., Dawson, G., 2001. Prototype formation in autism. Dev. Psychopathol. 13, 111–124.

Klinger, L.G., Klinger, M.R., Pohlig, R.L., 2006. Implicit learning impairments in autism spectrum disorders: implications for treatment. New Developments in Autism: The Future Is Today. Jessica Kingsley Publishers.

Knowlton, B.J., Squire, L.R., 1993. The learning of categories: parallel brain systems for item memory and category knowledge. Science 262, 1747–1749.

Koriat, A., Sorka, H., 2015. The construction of categorization judgments: using subjective confidence and response latency to test a distributed model. Cognition 134, 21–38. https://doi.org/10.1016/j.cognition.2014.09.009.

Little, D.M., Shin, S.S., Sisco, S.M., Thulborn, K.R., 2006. Event-related fMRI of category learning: differences in classification and feedback networks. Brain Cogn. 60, 244–252. https://doi.org/10.1016/j.bandc.2005.09.016.

Liu, R., Holt, L.L., 2009. Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. J. Cogn. Neurosci. 23, 683–698. https://doi.org/10.1162/jocn.2009.21392.

Mack, M.L., Preston, A.R., Love, B.C., 2013. Decoding the brain's algorithm for categorization from its neural implementation. Curr. Biol. 23, 2023–2027. https://doi.org/10.1016/j.cub.2013.08.035.

Medin, D.L., Schaffer, M.M., 1978. Context theory of classification learning. Psychol. Rev. 85, 207.

Medin, D.L., Altom, M.W., Murphy, T.D., 1984. Given versus induced category representations: use of prototype and exemplar information in classification. J. Exp. Psychol. Learn. Mem. Cogn. 10, 333–352. https://doi.org/10.1037/0278-7393.10.3.333.

Mercado, E., Church, B.A., 2016. Brief report: simulations suggest heterogeneous category learning and generalization in children with autism is a result of idiosyncratic perceptual transformations. J. Autism Dev. Disord. 46, 2806–2812. https://doi.org/10.1007/s10803-016-2815-4.

Mercado, E., Church, B.A., Coutinho, M.V.C., Dovgopoly, A., Lopata, C.J., Toomey, J.A., Thomeer, M.L., 2015. Heterogeneity in perceptual category learning by high functioning children with autism spectrum disorder. Front. Integr. Neurosci. 9, 42. https://doi.org/10.3389/fnint.2015.00042.

Mercado, E., Chow, K., Church, B.A., Lopata, C., 2020. Perceptual category learning in autism spectrum disorder: truth and consequences. Neurosci. Biobehav. Rev. 118, 689–703. https://doi.org/10.1016/j.neubiorev.2020.08.016.

Meyer, A.T., 2014. Visually Guided Prototype Learning in Children with Autism Spectrum Disorder (PhD Thesis). The University of North Carolina at Chapel Hill.

Minda, J.P., Smith, J.D., 2001. Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. J. Exp. Psychol. Learn. Mem. Cogn. 27, 775–799. https://doi.org/10.1037/0278-7393.27.3.775.

Molesworth, C.J., Bowler, D.M., Hampton, J.A., 2005. The prototype effect in recognition memory: intact in autism? J. Child Psychol. Psychiatry 46, 661–672. https://doi.org/10.1111/j.1469-7610.2004.00383.x.

Molesworth, C.J., Bowler, D.M., Hampton, J.A., 2008. When prototypes are not best: judgments made by children with autism. J. Autism Dev. Disord. 38, 1721–1730. https://doi.org/10.1007/s10803-008-0557-7.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., Burack, J., 2006. Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. J. Autism Dev. Disord. 36, 27–43. https://doi.org/10.1007/s10803-005-0040-7.

Nosofsky, R.M., 1988. Exemplar-based accounts of relations between classification, recognition, and typicality. J. Exp. Psychol. Learn. Mem. Cogn. 14, 700.

Nosofsky, R.M., Little, D.R., James, T.W., 2012. Activation in the neural network responsible for categorization and recognition reflects parameter changes. PNAS 109, 333–338. https://doi.org/10.1073/pnas.1111304109.

Nosofsky, R.M., Zaki, S.R., 1998. Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. Psychol. Sci. 9, 247–255. https://doi.org/10.1111/1467-9280.00051.

Olsson, H., Wennerholm, P., Lyxzén, U., 2004. Exemplars, prototypes, and the flexibility of classification models. J. Exp. Psychol. Learn. Mem. Cogn. 4, 936–941. https://doi.org/10.1037/0278-7393.30.4.936.

Pascual-Leone, A., Freitas, C., Oberman, L., Horvath, J.C., Halko, M., Eldaief, M., Bashir, S., Vernet, M., Shafi, M., Westover, B., 2011. Characterizing brain cortical plasticity and network dynamics across the age-span in health and disease with TMS-EEG and TMS-fMRI. Brain Topogr. 24, 302.

Perry, E.K., Lee, M.L., Martin-Ruiz, C.M., Court, J.A., Volsen, S.G., Merrit, J., Folly, E., Iversen, P.E., Bauman, M.L., Perry, R.H., 2001. Cholinergic activity in autism: abnormalities in the cerebral cortex and basal forebrain. Am. J. Psychiatry 158, 1058–1066.

Plate, R.C., Wood, A., Woodard, K., Pollak, S.D., 2019. Probabilistic learning of emotion categories. J. Exp. Psychol. Gen. 148, 1814–1827. https://doi.org/10.1037/xge0000529.

Posner, M.I., Keele, S.W., 1968. On the genesis of abstract ideas. J. Exp. Psychol. 77, 353–363. https://doi.org/10.1037/h0025953.

Posner, M.I., Goldsmith, R., Welton Jr., K.E., 1967. Perceived distance and the classification of distorted patterns. J. Exp. Psychol. 73, 28.

Schacter, D.L., 1990. Perceptual representation systems and implicit memory. Toward a resolution of the multiple memory systems debate. Ann. N. Y. Acad. Sci. 608, 543–567. https://doi.org/10.1111/j.1749-6632.1990.tb48909.x discussion 567-571.

Schipul, S.E., Just, M.A., 2016. Diminished neural adaptation during implicit learning in autism. Neuroimage 125, 332–341. https://doi.org/10.1016/j.neuroimage.2015.10.039.

Seger, C.A., Miller, E.K., 2010. Category learning in the brain. Annu. Rev. Neurosci. 33, 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546.

Shin, H.J., Nosofsky, R.M., 1992. Similarity-scaling studies of dot-pattern classification and recognition. J. Exp. Psychol. Gen. 121, 278–304. https://doi.org/10.1037//0096-3445.121.3.278.

Sinha, R.R., 1999. Neuropsychological Substrates of Category Learning. ProQuest Information & Learning, US.

Smith, J.D., 2002. Exemplar theory's predicted typicality gradient can be tested and disconfirmed. Psychol. Sci. 13, 437–442. https://doi.org/10.1111/1467-9280.00477.

Smith, J.D., Minda, J.P., 2002. Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. J. Exp. Psychol. Learn. Mem. Cogn. 28, 800–811.

Smith, A.D., Park, D.C., Cherry, K., Berkovsky, K., 1990. Age differences in memory for concrete and abstract pictures. J. Gerontol. 45, P205–P210. https://doi.org/10.1093/geronj/45.5.P205.

Smith, J.D., Zakrzewski, A.C., Johnson, J.M., Valleau, J.C., Church, B.A., 2016. Categorization: the view from animal cognition. Behav. Sci. 6, 12. https://doi.org/10.3390/bs6020012.

Tager-Flusberg, H., 1985. Basic level and superordinate level categorization by autistic, mentally retarded, and normal children. J. Exp. Child Psychol. 40, 450–469. https://doi.org/10.1016/0022-0965(85)90077-3.

Vanpaemel, W., 2016. Prototypes, exemplars and the response scaling parameter: a Bayes factor perspective. J. Math. Psychol. 72, 183–190. https://doi.org/10.1016/j.jmp.2015.10.006. Bayes Factors for Testing Hypotheses in Psychological Research: Practical Relevance and New Developments.

Vladusich, T., Olu-Lafe, O., Kim, D.-S., Tager-Flusberg, H., Grossberg, S., 2010. Prototypical category learning in high-functioning autism. Autism Res. 3, 226–236. https://doi.org/10.1002/aur.148.

Voorspoels, W., Rutten, I., Bartlema, A., Tuerlinckx, F., Vanpaemel, W., 2018. Sensitivity to the prototype in children with high-functioning autism spectrum disorder: an example of Bayesian cognitive psychometrics. Psychon. Bull. Rev. 25, 271–285. https://doi.org/10.3758/s13423-017-1245-4.

Yuste, R., 2015. From the neuron doctrine to neural networks. Nat. Rev. Neurosci. 16, 487–497. https://doi.org/10.1038/nrn3962.

Zaki, S.R., Nosofsky, R.M., Jessup, N.M., Unverzagt, F.W., 2003. Categorization and recognition performance of a memory-impaired group: evidence for single-system models. J. Int. Neuropsychol. Soc. 9, 394–406. https://doi.org/10.1017/S1355617703930050.

Zeithamova, D., Bowman, C.R., 2020. Generalization and the hippocampus: more than one story? Neurobiol. Learn. Mem. 175, 107317. https://doi.org/10.1016/j.nlm.2020.107317.

Zeithamova, D., Maddox, W.T., Schnyer, D.M., 2008. Dissociable prototype learning systems: evidence from brain imaging and behavior. J. Neurosci. 28, 13194–13201. https://doi.org/10.1523/JNEUROSCI.2915-08.2008.