Relating Categorization to Set Summary Statistics Perception

Noam Khayat and Shaul Hochstein

Life Sciences Institute and Edmond and Lily Safra Center (ELSC) for Brain Research, Hebrew University, Jerusalem, Israel 91904

Keywords: Categorization, prototype, boundary, summary statistics, ensemble, mean, range

contact information:

Prof. Shaul Hochstein,

Life Sciences Institute, Hebrew University, Jerusalem, 91904, Israel

shaulhochstein@gmail.com

+97254-421-8860

Abstract:

Two cognitive processes have been explored which compensate for the limited information that can be perceived and remembered at any given moment. The first parsimonious cognitive process is object categorization. We naturally relate objects to their category, assume they share relevant category properties, often disregarding irrelevant characteristics. Another scene organizing mechanism is representing aspects of the visual world in terms of summary statistics. Spreading attention over a group of objects with some similarity, one perceives an ensemble representation of the group. Without encoding detailed information of individuals, observers process summary data concerning the group, including set mean for various features (from circle size to face expression). Just as categorization may include/depend on prototype and inter-category boundaries, so set perception includes property mean and range. We now explore common features of these processes. We previously investigated summary perception of low-level features with a Rapid Serial Visual Presentation (RSVP) paradigm and found that participants perceive both the mean and range extremes of stimulus sets, automatically, implicitly, and on-the-fly, for each RSVP sequence, independently. We now use the same experimental paradigm to test category representation of high-level objects. We find participants perceive categorical characteristics better than they code individual elements. We relate category prototype to set mean and same/different category to in/out-of-range elements, defining a direct parallel between low-level set perception and high-level categorization. The implicit effects of mean or prototype and set or category boundaries are very similar. We suggest that object categorization may share perceptual-computational mechanisms with set summary statistics perception.

Introduction

Categorization is one of the most important mechanisms for facilitating perception and cognition, helping to overcome cognitive-perceptual bottlenecks (Cowan, 2001; Luck & Vogel, 1997) and perceive the "gist" of the scene (Alvarez & Oliva, 2009; Cohen, Dennet & Kanwisher, 2016; Hochstein & Ahissar, 2002; Hock, et al., 1974; Iordan et al. 2015, 2016; Jackson-Nielsen, Cohen & Pitts, 2017; Oliva & Torralba, 2006; Posner & Keele, 1970). Categorization follows and expands on the natural categories of objects in our environment, the intrinsic correlational structure of the world (Goldstone & Hendrickson, 2010; Rosch et al., 1976). There is long-term debate concerning the mechanisms and cerebral sites of categorization, with recent studies suggesting that there are multiple sites and processes of categorization. Thus, categorization itself may be categorized by task or goal (Ashby & Maddox, 2011), neural circuit (Iordan et al., 2015; Nomura & Reber, 2008), utility (Smith, 2014), and context (Barsalou, 1987; Koriat & Sorka, 2015, 2017; Roth & Shoben, 1983). The most common and accepted theoretical mechanisms for categorization are still rule-based, defining clear boundaries between categories (Goldstone & Kersten, 2003; Sloutsky, 2003; Davis & Love, 2010) and their cortical representations (Kriegeskorte et al. 2008; Iordan et al. 2015, 2016), and prototype- or exemplar-based, defining family resemblance (Ashby & Maddox, 2011; Goldstone & Kersten, 2003; Iordan et al., 2016; Maddox & Ashby, 1993; Medin, Altom & Murphy, 1984; Nosofsky, 2011; Rosch, 1973; Rosch et al., 1976; Posner and Keele, 1968).

In parallel, recent interest has focused on the perception of summary statistics of sets of stimulus elements. Observers have a reliable representation of the mean and range of sets of stimuli, even without reliable perception of the individual members of the presented set. Summary statistics, rapidly extracted from sets of similar items, presented spatially (Ariely, 2001; Alvarez & Oliva, 2009) or temporally (Corbett & Oriet, 2011; Gorea, Belkoura & Solomon, 2014; Hubert-Wallander & Boynton, 2015), include average, and range or variance of their size (Allik et al., 2014; Ariely, 2001; Corbett & Oriet, 2011; Morgan et al., 2008; Solomon, 2010), orientation (Alvarez & Oliva, 2009), brightness (Bauer, 2009), spatial position (Alvarez & Oliva, 2008), and speed and direction of motion (Sweeny, Haroz, & Whitney, 2013). Extraction of summary statistics appears to be a general mechanism operating on various stimulus attributes, including low-level information, as mentioned above, and more complex characteristics, such as facial expression (emotion) and gender (Haberman & Whitney, 2007, 2009; Neumann, Schweinberger & Burton, 2013), object lifelikeness (Yamanashi-Leib, Kosovicheva & Whitney, 2016), biological motion of human crowds (Sweeny, Haroz, & Whitney, 2013) and even numerical averaging (Brezis, Bronfman & Usher, 2015; for recent reviews see Bauer, 2015; Cohen et al., 2016; Haberman & Whitney, 2012; Hochstein et al., 2015). Examples of the methods used in these studies are shown in Figure 1.

We have suggested that these phenomena, categorization and set perception, may be related since they share basic characteristics (Hochstein 2016a,b; Hochstein et al., 2018a). In both cases, when viewing somewhat similar, but cetainly not identical items, we consider them as if they were the same, as a shortcut to prescribing an single appropriate response (Ariely, 2001; Medin, 1989; Rosch & Mervis, 1975; Rosch et al., 1976). We globally spread attention and see a flock of sheep in a meadow, a shelf of alcohol bottles at a bar, a line of cars in traffic, or a corpse of trees in a forest. Similarly, we see a set of circles (Ariely, 2001; Alvarez & Oliva, 2008; Corbett & Oriet, 2011) or line segments (Robitaille & Harris, 2011), or a set of faces in a crowd (Haberman & Whitney, 2007, 2009). All animals in the category "dogs" have four legs and a tail, but they may vary in color and size. All circles in a set are round, though they may vary in size or brightness. Categorization emphasizes relevant or common properties and de-emphasizes irrelevant or uncommon properties, reducing differences among category members (Fabre-Thorpe, 2011; Goldstone & Hendrickson, 2010; Hammer et al., 2009; Rosch et al., 1976; Rosch, 1999, 2002). Similarly, set perception captures summary statistics without noting individual values. Categorization, like ensemble perception, may depend on rapid feature extraction, to determine presence of defining characteristics of objects.

Set perception includes set mean and range (Ariely, 2001; Chong & Treisman, 2003, 2005), and categorization might rely on the related prototype (or mean exemplar) and/or inter-category boundaries (or category range). This conceptual similarity has been confirmed by the recent finding that set characteristics are perceived implicitly and automatically (Khayat & Hochstein, 2018), just as objects are categorized implicitly and automatically at their basic category level (Potter & Hagmann, 2015; Rosch et al., 1976). Finally, it has been suggested that determining whether a group of objects in a scene belong to the same category may actually depend on their characteristics that allow them to be seen as a set (Utochkin, 2015).

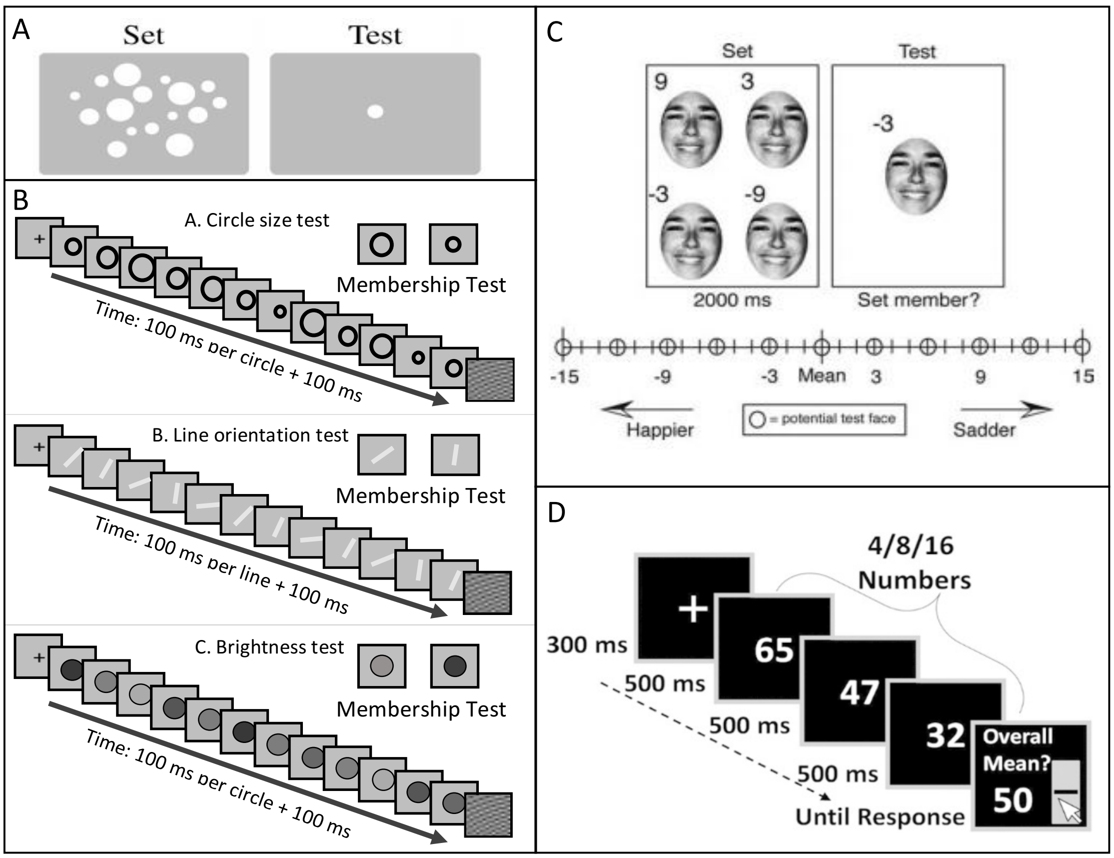

Figure 1. Previous study stimulus sets. **(A)** Ariely's (2001) schematic representation of the two intervals used in his experiment's trials. Observers were exposed for 500ms to a set of spatially disperesed circles differing by size, and then asked if a test stimulus was present, or smaller/larger than the set mean. **(B)** Khayat & Hochstein's (2018) RSVP sequences consisted of twelve circles, each presented for 100ms plus 100ms inter-stimulus interval (ISI), followed by a two alternative forced choice (2-AFC) memberhip test. Blocks contained circles differing in size, lines differing in orientation, or discs differing in brightness.

Observers were asked which of two test elements was present in the set. They were unaware that either test element could equal the set mean or the non-member could be outside the set range. **(C)** Haberman & Whitney's (2009) task included four faces (from a set of 4, 8, 12 or 16), differing in facial emotional expression, presented for 2s. Observers then indicated whether the test face was a member of the set. **(D)** Brezis, Bronfman & Usher's (2015) trials consisted of two-digit numbers sequentially presented in a rate of 500ms/stimulus. Set size was 4, 8 or 16. Participants reported set average.

The goal of the current research is to detail the similarity between set and category perception by applying to categories the very same tests that we used to study implicit set perception (Khayat & Hochstein, 2018). The following section briefly reviews the results of these previous tests. We note that there are important differences between categorization and set perception. Object categories are learned over a lifetime of experience, while set ensemble statistics are acquired on-the-fly. Different life experience may lead to individual differences in categorization and choice of object seen as the category prototype. Categorization may involve semantic processes, while set perception has been demonstrated for simple visual features, (though including face emotion). Thus, it would be difficult to claim that ensemble perception and categorization are identical, or take place at the same cortical site. However, their being different makes comparing them even more important, since if they share essential properties, they may depend on similar or analogous processes. This is the aim of the current study.

Previous Study

We studied implicit perception and memory of set statistics by presenting a Rapid Serial Visual Presentation (RSVP) sequence of images of low-level items (circles of different size, lines of different orientation, discs of different brightness; see Figure 1B), and testing only memory of the members of the sequence (Khayat & Hochstein, 2018). Following set RSVP presentation, memory was tested by asking participants to choose which of two simultaneously presented image items was a sequence member. We did not inform participants that sometimes one test element – the member or non-member of the previously viewed sequence – was the mean of the sequence, and sometimes the non-member was outside the sequence element range.

We call these test-stimulus contingencies, trial subtypes (where A indicates test element in sequence range, B indicates out of range, and Am indicates sequence mean; first letter refers to test element member, second to non-member): Am-A (member test element equals set mean; both elements in range of sequence), A-Am (non-member equals mean), A-A (neither equals mean), Am-B and A-B (non-member is outside set range). As demonstrated in Figure 2A-D, we found a mean effect, for each of the three variables tested, circle size, line orientation and disk brightness: participants chose the test element that was equal to the mean more often (whether it was a member, Am-A, or not, A-Am, compared to A-A where neither was the mean).

We concluded that, since the stimulus sequence was quite rapid, participants had difficulty remembering all the members of the RSVP set, and maybe even any one of them. Instead, they automatically used their implicit perception of the sequence set mean and range to respond positively to test elements that matched or were close to the set mean. Thus, performance was more accurate for test member elements that equaled the mean – Am-A (Figure 2A,B middle bars of the left three; 2C,D left bars). When the non-member test element was equal to the set mean, it was frequently chosen as if it were a member, i.e. as if it had been seen in the set sequence. Participants actually chose this mean non-member even more frequently than the actual non-mean member – A-Am (Figure 2A,B leftmost bars; note that accuracy below 0.5 means that the non-member was chosen more frequently than the member.)

In addition, we found a *range effect*, i.e. participants rejected out-of-range non-members (Am-B and A-B) more frequently than in-range non-members (Am-A, A-A, A-Am). This is seen in Figure 2A,B, right two bars, and in Figure 2E,F, right bars, compared to left bars in each graph. The same effect was seen for response time (RT; Figure 2G) which was shorter for out-of-range than in-range non-members, indicating they were rejected more rapidly as well as more frequently.

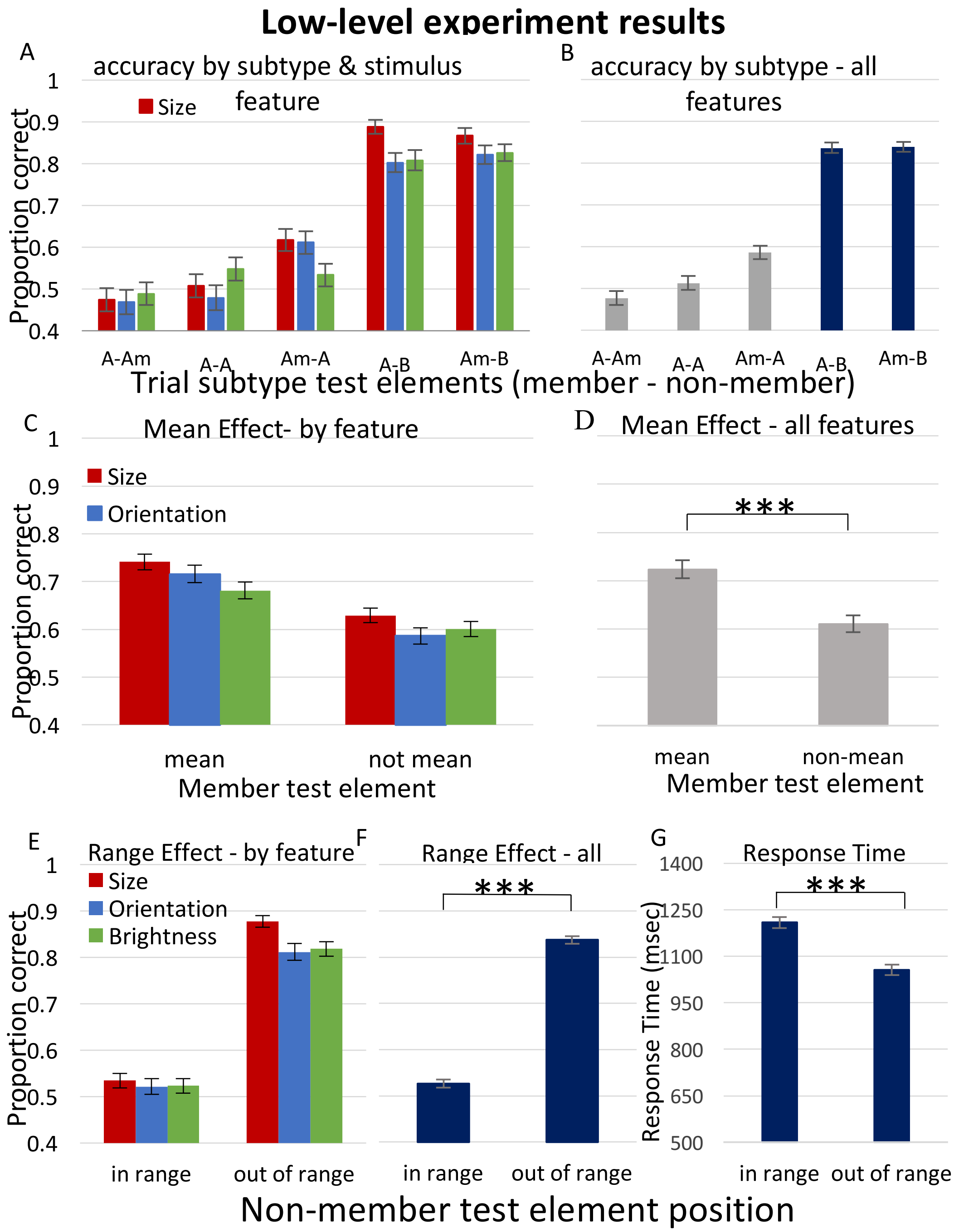

Figure 2. Low-level experiment results. (A) Accuracy rates for each trial subtype, i.e. their test elements; member-nonmember being equal to the set sequence mean (Am), being in the set range (A) or outside the range (B) and each stimulus feature (colored bars; see legend). Thus, trial subtypes include: Am-A

(member test element = mean; both test elements in sequence range); A-Am (non-member = mean; both in sequence range); A-A (neither = mean; both in sequence range); Am-B (member = mean; non-member outside sequence range); A-B (member not = mean; non-member outside sequence range). **(B)** Accuracy rates for each trial subtype, averaged across stimulus features. **(C)** Mean effect for each stimulus feature; accuracy rates for trials where the member test element equaled the set mean versus when it differed from the mean**.** Each comparison is significant $p<0.05$. **(D)** mean effect across features**.** $p<0.001$. **(E)** Range effect for each stimulus feature; accuracy rates for trials where the non-member test element is in-range versus out-of-range. Each comparison is significant $p<0.01$. **(F)** Range effect across features. $p<0.001$. **(G)** Range effect seen in Response Time, indicating this is not a accuracy-time trade off. $p<0.001$. All results from Khayat & Hochstein, 2018. Error bars here and in all folowing graphs represent between-participant standard error of the mean.

We concluded that participants automatically and implicitly determined the mean and range of the RSVP sequence even though they were not instructed to do so and even though this had no bearing on performance of the task at hand. Furthermore, they did so on-the-fly for each trial, independently, since each trial could have a different sequence mean and range.

Perception of set mean and range are not only implicit. In another study observers were asked to explicitly compare means of two arrays of variously oriented bars (mean comparison) or report presence of an outlier orientation among the array elements (outlier detection). It was found that mean comparison depended only the difference between the array means and outlier detection depended on the distance of the target from the array range edge (Hochstein, 2016a,b; Hochstein et al., 2018a,b). Thus, both set mean and range are perceived both explicitly and implicitly.

The goal of the current study is to test whether there are identical effects in the related perceptual phenomenon of categorization.

Prototypes as averages

We investigate here the non-trivial comparison between stimulus sets and object categories. The stimuli in previous studies of statistical perception were very similar, usually differing by a single varying feature (e.g., Ariely, 2001; Corbett & Oriet, 2011), or a combination of features forming a single high-level feature (e.g., facial expression; Habeman & Whitney, 2007, 2009). In contrast, categories might be thought of as a set of objects comprised of combinations of multiple features, with only some of these features necessarily present in each category exemplar, (where membership is defined by family resemblance). Thus, we compare the **mean** of the set elements with the **prototype** of category exemplars, based on the view that prototypes are the central or most common representations of a category (Goldstone & Kersten, 2003), possessing the mean values of its attributes (Langlois & Roggman, 1990; Reed, 1972; Rosch et al., 1976; Rosch & Lloyd, 1978; Rosch, Simpson & Miller, 1976). Note, however, that comparing these perceptual procedures does not depend on this definition of prototype, or even on prototype theory itself. Comparing categorization with set summary perception is valid simply because in both cases several stimuli are seen as belonging together, perhaps inducing the same response, because they share some characteristics and differ in others.

Similarly, we compare knowledge of category boundaries with perception of set range edges. As shown above, perceiving set range edges allows for rapid detection of outlier elements, and even unconscious perception of these edges allows for rapid rejection of out-of-range elements when trying to remember which elements were previously viewed. This was called the "range effect" (Khayat & Hochstein 2018). Similarly, knowing category boundaries allows for rapid separation of objects that belong to different categories, which we shall call a "boundary effect." Thus, we compare properties of set perception and categorization in terms of the implicit determination and knowledge of both the set mean and category prototype, as well as, the set range edges and the category boundaries. That is, having found that observers perceive rapidly and implicitly the mean and range of element sets, and use this information when judging memory of sequence stimuli, we now test if the same characteristics are present for object categories. Do observers of a sequence of objects determine automatically and implicitly their category and use the implied prototype (whether shown or not shown in the sequence) and the boundaries of the implied category, when later choosing images as having been seen in the sequence? These will be called the prototype and boundary effects, respectively. If we find similar characteristics in these processes, for categorization as for set perception, we will suggest that they may share basic cognitive mechanisms.

We note at the outset that there are important differences between perceiving set summary statistics and categorizing objects. We perceive the mean size, orientation, brightness, etc., of sets that we see just once, sets which are unrelated to any other sets seen before. Presented with a set of images, sequentially or simultaneously, we derive the mean and range of the size, orientation, brightness, etc., of that set, on-the-fly and trial by trial. Thus, presented with a single stimulus in isolation, it is logically inconsistent to ask to what set it belongs. In contrast, by their very nature, categories are learned over a lifetime of experience, and with this knowledge, we can know immediately to what category a group of objects, or even a single object belongs. In fact, one of the defining characteristics of "basic" categories is that these are the names given to single objects (e.g. cat, car, fork, apple, …). The situation with categorization is unlike that with sets, where we derive the set mean, on-the-fly, as we are presented with set members. Instead, when encountering an object (or group of objects belonging to a single category), we know the category to which it belongs, and we also know what is the prototype of that category; there is no need, and no possibility, of deriving anew the category and prototype of a group of familiar objects, (though we can learn new categories of unfamiliar objects). Furthermore, categories may be learned and recognized semantically, while the basic features of sets are often non-semantic. Nevertheless, and this is the basic argument of the current study, there may be similarities, if not identities, of mechanisms for representing set means/ranges and category prototypes/boundaries. We set out here to find the degree of similarity between these very different phenomena before endeavoring to uncover underlying mechanisms. Finding similarities, despite the differences enumerated above, would suggest that there are relationships between low-level and high-level representations of images, objects, categories, and concepts.

Methods

We present Rapid Stimulus Visual Presentation (RSVP) sequences of images of high-level category objects, conditions known to impair focused attention to each stimulus, but maintain statistical and categorical representations across time (Corbett & Oriet, 2011; Potter et al., 2014). We then present two images, one identical to one of the images in the sequence (the member) and the other a novel image (the non-member). Observer task is to choose the member – the image that was present in the sequence. This is a two-alternative forced choice (2-AFC) test, which is thus criterion-free, and has a chance guessing level of 50%. See Figure 3.

We do not inform observers that one of the imaged objects (either the member or the non-member) may be prototypical of the sequence category, and one (the non-member) may be outside the sequence category, i.e. belong to another category. Note that when non-member objects were chosen from a different category, still, they were purposely chosen to be not too distant from the sequence category, that is, from a relatively close category. We hypothesize that the influence of prototypes on implicit categorization and thus on memory will be similar to the influence of the mean when we tested set item memory (Khayat & Hochstein, 2018). Thus, we expect observers to accept prototypical objects as members more frequently, (whether or not they were in the sequence). Additionally, the presence in the test pair of an object outside the sequence category may aid in rejecting it as not seen in the sequence, just as items outside the set range were more easily rejected as non-members (Figure 2E,F).

Participants

Data of 15 in-house participants, students at the Hebrew University of Jerusalem were included in the analysis of Experiment 1 (age range = 20–27 years, mean = 23.4 years; 4 males, 11 females). We also have results for 226 Amazon Mechanical Turk participants for Experiment 3. Participants provided informed consent and received compensation for participation and reported normal or corrected-to-normal vision.

Stimuli and procedure

Procedures for Experiment 1 took place in a dimly lit room, with participants seated 50cm from a 24" Dell LCD monitor. We have less information as to their identity and precise experimental conditions of the Experiment 3 Amazon MTurks; (we excluded ~25% of these data for trials with RTs <200ms or >4s and for subjects with <33% remaining trials or <60% correct responses overall, thus including as many trials/subjects as possible, excluding data which are clearly not responses to the stimulus; e.g. Fabre-Thorpe, 2011). Stimuli were generated using Psychtoolbox version 3 for MATLAB 2015a (Brainard, 1997). MTurk testing used Adobe flash. Images, chosen from the "Google Images" database, were presented against a gray background (RGB 0.5, 0.5, 0.5).

Stimuli consisted of rapid serial visual presentation (RSVP) of a sequence of high-level objects or scene images presented in the center of the display, with a fixed size of 10.4cm high × 14.7cm wide, as demonstrated in Figure 3 (see also examples of images in Figure 8). Experiment 1 was divided into three blocks of 65 RSVP trials each, with a short break between them, to complete 195 trials total per participant; Experiment 3 had 60 trials total for MTurk; 1 session/participant.

A set of images (12 for in-house students; 9 for MTurks) was presented in each RSVP sequence, with 167ms Stimulus Onset Asynchrony (100ms stimulus + 67ms Inter-Stimulus Interval), followed by a 100ms masking stimulus. Then, after 1.5s, two images were presented side-by-side, simultaneously for the membership test; one, a member of the sequence and one a novel, non-member. Sequence member and non-member images were randomly placed to the left and right of fixation in the middle half of the width and height of the screen, and participants indicated position of the member by keypress. Images remained present until observer response. Since participants tend to perceive and remember better early and late elements, known as primacy and recency effects, in general and specifically in summary representations (Hubert-Wallander & Boynton, 2015), we excluded from the test member images the first and last two RSVP sequence images.

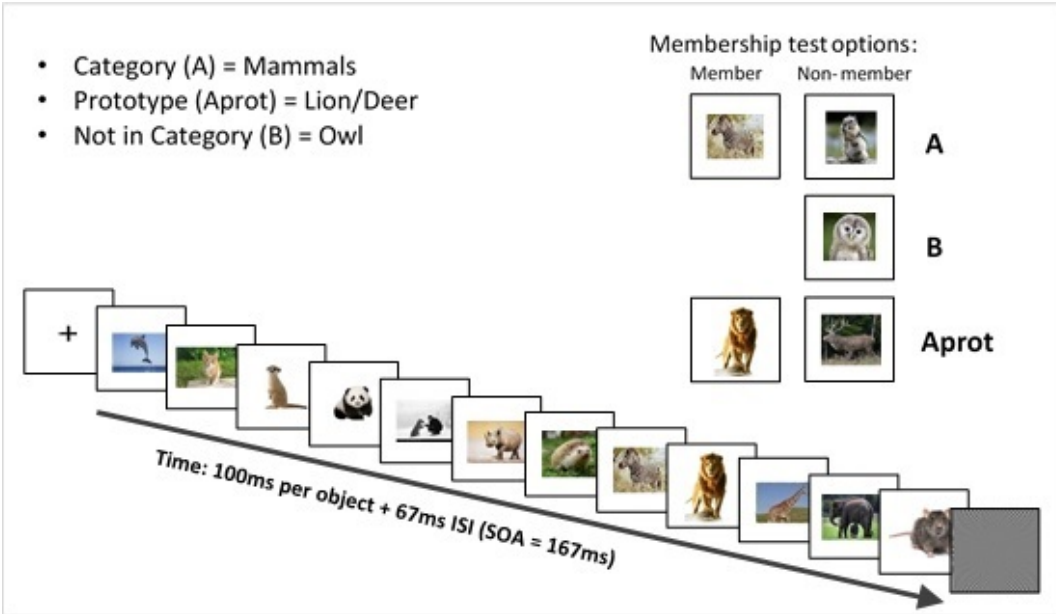

Figure 3. High-level categories RSVP membeship tests. Example RSVP trial with Mammals as the set category. On the membership test one of the optional subtype pair of images (Table 2) was presented for the member and the non-member images. The five trial subtypes for each of the 39 categories are designed by choice of the membership image. A member object image could be either, A (regular member of the category, not the prototype) or Aprot (category prototype), while the non-member object could also be B (belong to a different category).

		Category level		Exemplar types	
		Superordinate level	**Basic level**	**Typical exemplars (Prototypes & Common)**	**Non-Prototype exemplar**
Taxonomy	**Biological**	Plants*		Potted plant, Cactus	Watermelon plants, Vine
			Trees*	Oak, Olive	Sequoia, Baobab
		Fruits*		Apple, Orange	Pomegranate, Litchi
		Animals*		Dog, Deer	Mosquito, Octopus
			Reptiles	Python, Iguana	Legless Lizard, Commodore
			Birds*	Owl, Pigeon	Penguin, Pelican
		Mammals*		Cow, Lion	Whale, Bat
			Dogs*	German Shepard, Labrador	Chi Wawa, Bull-Terrier
	Non-biological	Food*		Pasta, Pancakes	Cake, Sushi
		Weapons*		Pistol, Riffle	Cannon, Molotov bottle
		Books		Harry Potter, The Bible	The hobbit, Comics
		Kitchen tools*		Whisker, Slicing Knife	Grater, Blender
		Toys*		Teddy bear, Rubik cube	Top, Plastic food
		Furniture*		Armchair, Sofa	Dresser, Stool
			Desks	Office desk, Writing desk	Reception desk, Cubicle desk
		Houses		Villa, Apartments	Igloo, Canoe
		Vehicles*		Car, Bus	Unicycle, Helicopter
			Cars*	Sedan, Hatchback	Formula 1, Model T
		Liquids		Water, Milk	Acetone, Soap
			Drinks*	Milk, Beer	Cognac, Sake
		Electronics*		TV screen, Laptop	Hair dryer, Shaver
		Clothes*		Shirt, Trousers	Socks, Gloves
	Abstract concepts	Games		Puzzle, Chess	Bowling, Super-Mario
		Music		Musical note, The Beatles	Mexican band, Accordion
			Sports*	Soccer, Basketball	Bowling, Billiards
		Religion		Jesus, Western wall	Buddha, Praying man
		Science*		Test tubes, Atom	Lecture, MRI
		Conflicts		Israeli-Palestinian	Random couple argument
		Symbols		Peace symbol, David star	Scouts symbol, Recycle symbol
		Occupations*		Judge, Policeman	Fisherman, Violinist
		Disasters		9.11 Plane crash, Tsunami	Volcano eruption, Avalanche
		Movies		The Godfather, Cinema & Popcorn	Cameraman, Script
			Horror	Wolf & full-moon, Hannibal Lecter	Scared face, Creepy doll
			Cartoons	Mickey mouse, The Simpsons	Scooby doo, Hello kitty
		Events		Wedding, Festival	Graduation ceremony, Parade
		Travel		Passport & Suitcase, Backpackers	Airport, Sunglasses
		Health		Heart-beat icon, Workout	Non-smoking, Granola
		Hazard		Slippery sign, Toxic (skull) sign	Unstable bridge, Medusa
		History		Martin Luther King, Hiroshima	Che Guevara, Mayan temples

Table 1. Categories with examples of their prototypes and other exemplars. The 39 categories used in the student experiment; 20 categories for MTurks, indicated by *. Categories are placed in the first or second column according to their being superordinate or basic level categories.

Thirty-Nine categories (20 for MTurks) were included in the experiment (Table 1), including manmade and natural objects (animate, inanimate and plants) and abstract conceptual scenes from different category levels. Each category was repeated in each trial subtype (see below), with entirely different images for each trial. For each category, we chose the 3 images that seemed to us to be closest to prototypical, and used them in the 3 test subtypes including a prototype (as non-member or as member versus nonmember same/different category). Of the 39 categories used for Experiment 1, 20 were later tested in Experiment 2 and only these were used in Experiment 3. For almost all the 20 categories, which were also tested in Experiment 2 (see below), high typicality was confirmed; we discarded data for the few discrepant images (<6% of trials). For the remaining 19 more conceptual categories, which were not tested in Experiment 2 (and not used in Experiment 3), we depended on examples from the literature (e.g. McCloskey & Glucksberg, 1978; Iordan et al., 2016) and experimenter judgement for in-house student participants (who came from the same cohort as experimenter NK). Note that if we err and choose non-typical images as prototypes, this would add noise and reduce results' significance; thus, the results themselves confirm our choice. For the entirely new MTurk tests, we used a different approach, depending on Experiment 2, as described below. We purposely chose both basic and superordinate categories, as well as conceptual categories, to broaden the potential impact of our results.

Trial subtypes

Trial subtypes were defined by the nature of the two test image objects vis-à-vis the sequence category (as in the low-level tests; see above Introduction and Khayat & Hochstein, 2018). Each member test image could be of an object from the RSVP sequence category (denoted 'A') or the prototype of this category (Aprot). The non-member test image could be of an object from the RSVP category (A) or even its prototype (Aprot), but, in either case, not actually presented in the sequence; or, the non-member could be an image of an object from a different category (B). Figure 3 illustrates these image types. Each pair of test images could be of one of five subtypes, listed in Table 2, (denoted Aprot-A, A-A, A-Aprot, A-B, or Aprot-B). Each subtype was tested for each category listed in Table 1.

Subtype number	Member Test Image (Correct)	Non-member Test Image (Incorrect)
1	Aprot	A
2	A	A
3	A	Aprot
4	A	B
5	Aprot	B

Table 2. Member recall test trial subtypes: Test image objects could be both from the RSVP sequence (A), one could be the prototype (Aprot) of that sequence (whether presented or not), and the non-member could be of another category (B). On every trial one image was a member and the other wasn't. Test pairs of subtype 1 have the member object being the category prototype, and the non-member another category exemplar not shown in the sequence. In subtype 2, both member and non-member are from the sequence but neither is the prototype. In subtype 3, the non-member is the category prototype, which was not shown in the sequence. In sub-types 4 and 5, the non-member is from another category, and the member is not the prototype (subtype 4) or is the prototype (subtype 5). Trial subtypes were presented in randomized order without observers knowing about this classification.

3.1.5. *Statistical tests* & data analysis

Analysis of variance (ANOVA) tests with repeated measures were conducted to verify that performance accuracy differences were due to the difficulty derived by effects emerging from the different trial subtypes, rather than within-participant differences in performance. For the 2-way repeated measure ANOVA, testing student participant effects of member typicality and non-member category, we combined data for non-member same category, whether prototypical or not. T-tests (one-tailed) between the averaged results of all participants for different subtype combinations were performed to investigate prototype and boundary representations effects. Since it is difficult to remember all the sequence images, we expect participants to prefer as members those test images with objects that are prototypes of the sequence category and to reject as members, those that are of a different category.

Results

The two basic measurements indicating observer performance are accuracy rates and response time (RT) for each trial subtype, as shown for student participants in Figure 4. The results by trial subtype roughly resemble those from the low-level experiment, demonstrated in Figure 2B, with some effects even more salient, as detailed below. Figure 5 presents averaged accuracy results across participants, sorted by subtype, isolating the three subtypes with both test image objects within the sequence category (subtypes Aprot-A, A-A, A-Aprot), for student (Figure 5A) and Mturk participants (Figure 5B).

We performed a two-way repeated measure ANOVA on the Figure 4 results. The overall prototype effect – the effect of one of the objects being the prototype of the category of the objects presented in the sequence – was significant ($F_{1,14}$=18.07, p<0.001); the boundary effect – the effect of the nonmember being of another category than the sequence objects – was highly significant ($F_{1,14}$=298.64, p<0.001) and the interaction between them was significant, as well, ($F_{1,14}$=13.36, p<0.005). The interaction effect suggests that the prototype effect may be larger in some cases, as we shall see in the following paragraph.

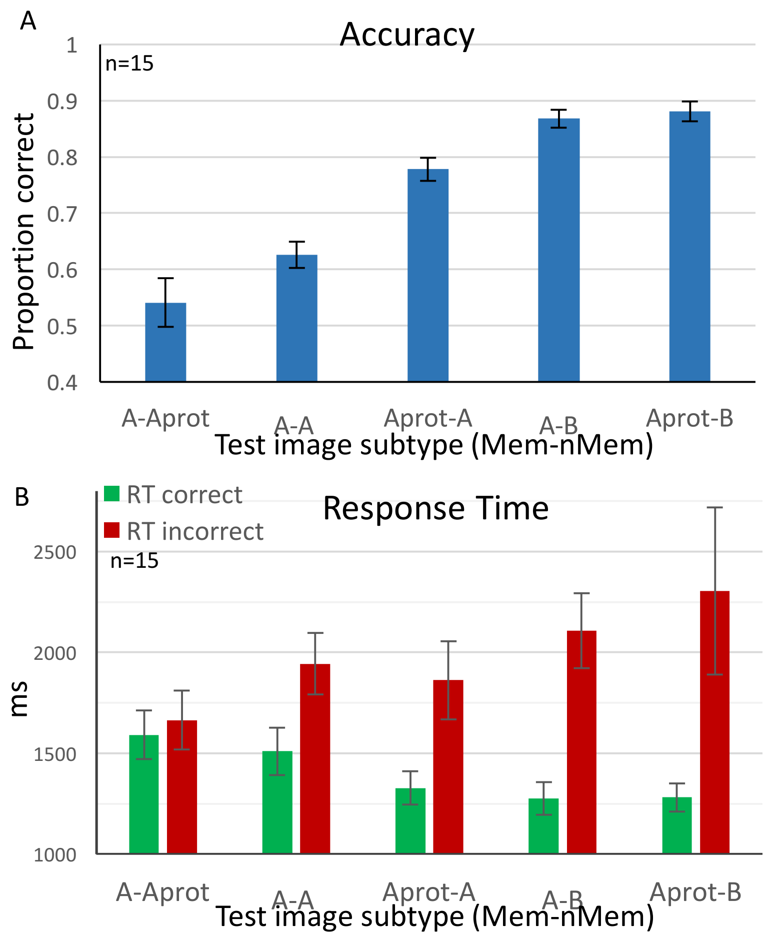

Figure 4. High-level image memory performance by RSVP trial subtype (Students). (A) Accuracy rates sorted by test image subtype (Mem=member, nMem=non-member). **(B)** Response Time measured for correct (green) and incorrect (red) responses, sorted by test image subtype.

Prototype Effect

The first factor to influence performance is the presence of category prototypical objects (prototypes and most common objects) in one of the test images. The presence of typical exemplars influenced accuracy (% correct responses) and RT, which together we call the *prototype effect*. As seen in the three left bars of Figure 4A and 5A,B prototype presence affected accuracy: accuracy Aprot-A > A-A > A-Aprot. Prototype presence affected response time (RT), as in Figure 4B: RT correct choice of member Aprot-A < A-A; RT incorrect choice of non-member A-Aprot < A-A).

It is possible that when including subtypes with non-member test images of an object of a different category (subtypes A-B and Aprot-B) in the above two-factor ANOVA calculation, the effect of the presence of a different category (B) reduces the prototype effect. Thus, to test the prototype effect alone, we conducted a one-way repeated measure ANOVA on the three subtypes with test image objects in the category boundaries (Figure 5). This one-factor ANOVA showed a significant prototype effect (students: $F_{2,28}=11.78$, $p<0.001$; MTurk: $F_{2,346}=26.96$, $p<0.001$). We conclude that, as predicted, when comparing trials containing only objects from the relevant category (subtypes Aprot-A, A-A, A-Aprot), the prototype had a major influence on observer response, which tended to attribute it as a member of the RSVP sequence regardless of whether it was or was not.

On the other hand, there is no significant difference between the case where the member is prototypical or not when the non-member is outside the category (Accuracy for Aprot-B = 0.88 versus for A-B = 0.86; $p=0.59$; see Figure 4A). The boundary effect overrides the prototype effect, (leading to the interaction effect in the two way repeated measure ANOVA above).

We conclude that, due to limited attentional resources, participants are unable to fully perceive and memorize all individual objects, but still succeed in having a good representation of the category itself. Therefore, they tend to relate the most representative object (the prototype) to the category (Figure 5A: students; Figure 5B: MTurks). We performed post-hoc T-tests between the different subtypes to find details of the effect, as shown in Figure 5A,B. The prototype effect is clearly present when comparing the relevant trial subtypes (Aprot-A, A-A, A-Aprot), which significantly differ from each other (students: $p<0.05$ for subtypes A-A versus Aprot-A or A-Aprot and $p<0.01$ for Aprot-A versus A-Aprot; MTurks: $p<0.001$ for all comparisons). These subtypes create a staircase shape from low performance of 0.54 ± 0.04 (MTurk: 0.64 ± 0.01; mean±s.e.) proportion correct for A-Aprot, via 0.63 ± 0.02 (0.7 ± 0.008) correct for A-A, to best performance of 0.78 ± 0.02 (0.76 ± 0.01) correct for Aprot-A. Surprisingly, even when the prototype was not present in the object sequence, it was often chosen as present when presented in the non-member test image. Nevertheless, when choosing between a non-prototypical member and a prototypical non-member (A-Aprot), membership is slightly more important than typicality (0.54 and 0.64 for students and MTurks, respectively; significantly > 0.50). We ask below if this is an all-or-none prototype-or-not prototype effect, or if it is a graded effect, as objects are more or less typical of the category.

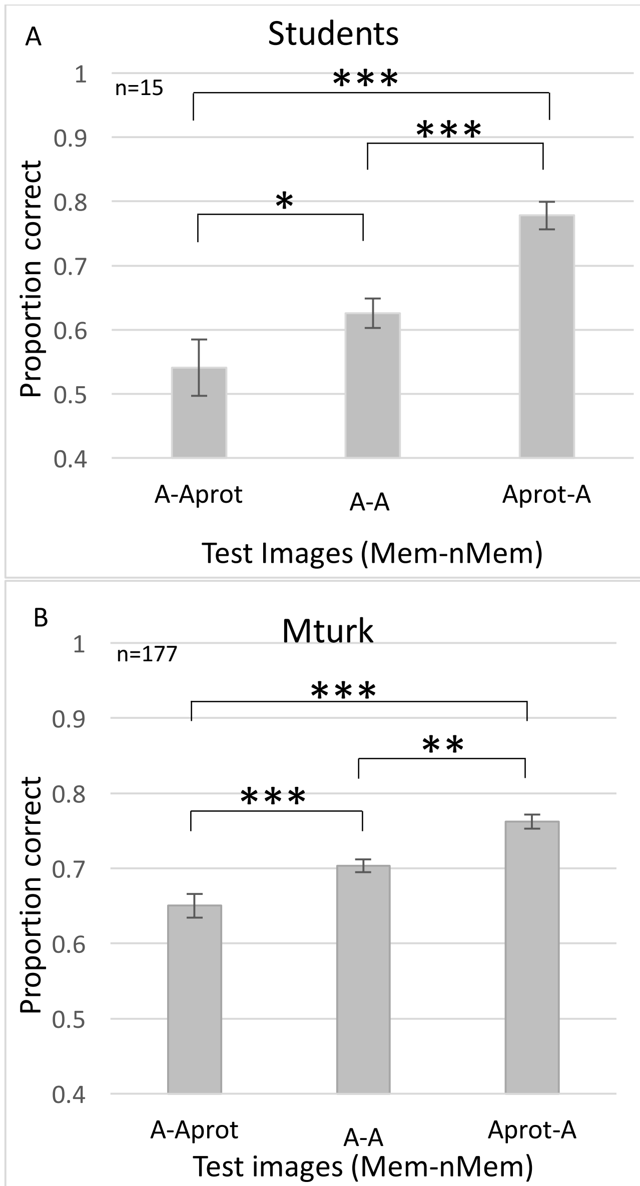

Figure 5. Category prototype object effect on accuracy. Proportion correct for subtypes with both test objects within the sequence category: A-Aprot, A-A and Aprot-A. T-tests among the subtypes show significant differences, indicating the expected prototype effect on observer judgment in membership-tests, with a

preference to choose the object which matches the category prototype (Mem=member, nMem=non-member). (A) Students. (B) Mturks. Significance indicated by: * $p<0.05$, ** $p<0.01$, ***$p<0.001$, here and in subsequent figures.

We conclude that with a failure of coding all the individual sequence images, due to brief image exposure times, the presence of prototype object images had a significant effect on the responses, whether they were members or non-members of the RSVP category. Along with these accuracy differences, an analysis of the response times (RT) provides additional support for the conclusion that participants perceive prototypes as ideal representatives of the category and "remember" these whether they were present or not. In Figures 6A-B, RT is classified into trials in which the non-member test image is correctly rejected (6A-B left green) or, incorrectly, chosen (6A-B right red) comparing when the non-member object is either a prototype or not. As expected, Figure 6A-B shows that correct responses (green) are made faster than incorrect responses (red), similar to the comparisons seen in Figure 4B. The details show further interesting comparisons, as follows. Analysis of the correct RTs indicate that when participants did correctly chose the non-prototype test member, they did so significantly slower when the non-member was a prototype (students: 1591±125ms; MTurk: 1364±28ms) than when the non-member was not a prototype (students: 1348±46ms; MTurk: 1319±23ms; T-test $p<0.05$), as displayed in Figure 6A-B, left diamonds. In other words, not only were they often manipulated to falsely pick the prototype as member (Figure 5), even when they did manage to choose a non-prototype member, their response was delayed, as if the presence of the non-member prototype (A-Aprot) affected their confidence. In addition, choosing the correct member object is faster when it is the prototype (Aprot-A versus A-A and A-Aprot, see Figure 4B, left 3 green bars). Furthermore, (Figure 6A-B, right red) choosing the non-member object, incorrectly, is faster when it is the prototype than when it is not (students: 1663±150ms versus 2015±130ms; MTurk: 1495±41ms versus 1557±31ms; T-test: $p=0.174$ (n.s.), $p<0.05$, respectively).

On the other hand, besides the prototype effect, there is still some degree of recognition of test object membership. Thus, as demonstrated in Figure 6C-D, choosing the prototypical object is faster when it is a sequence member (correct: Aprot-A; and

Aprot-B for students; students: 1304±50ms; MTurk: 1288±25ms) than when it is not (A-Aprot incorrect; students: 1663±150ms; MTurk: 1495±41ms; T-test: $p<0.05$, $p<0.001$. respectively). Even choosing the non-prototypical member is faster than choosing the typical non-member (Figure 6A-B middle two diamonds; T-test: $p=0.061$, $p<0.01$). This latter speed joins the greater accuracy (see above) to indicate it is not a speed-accuracy trade-off.

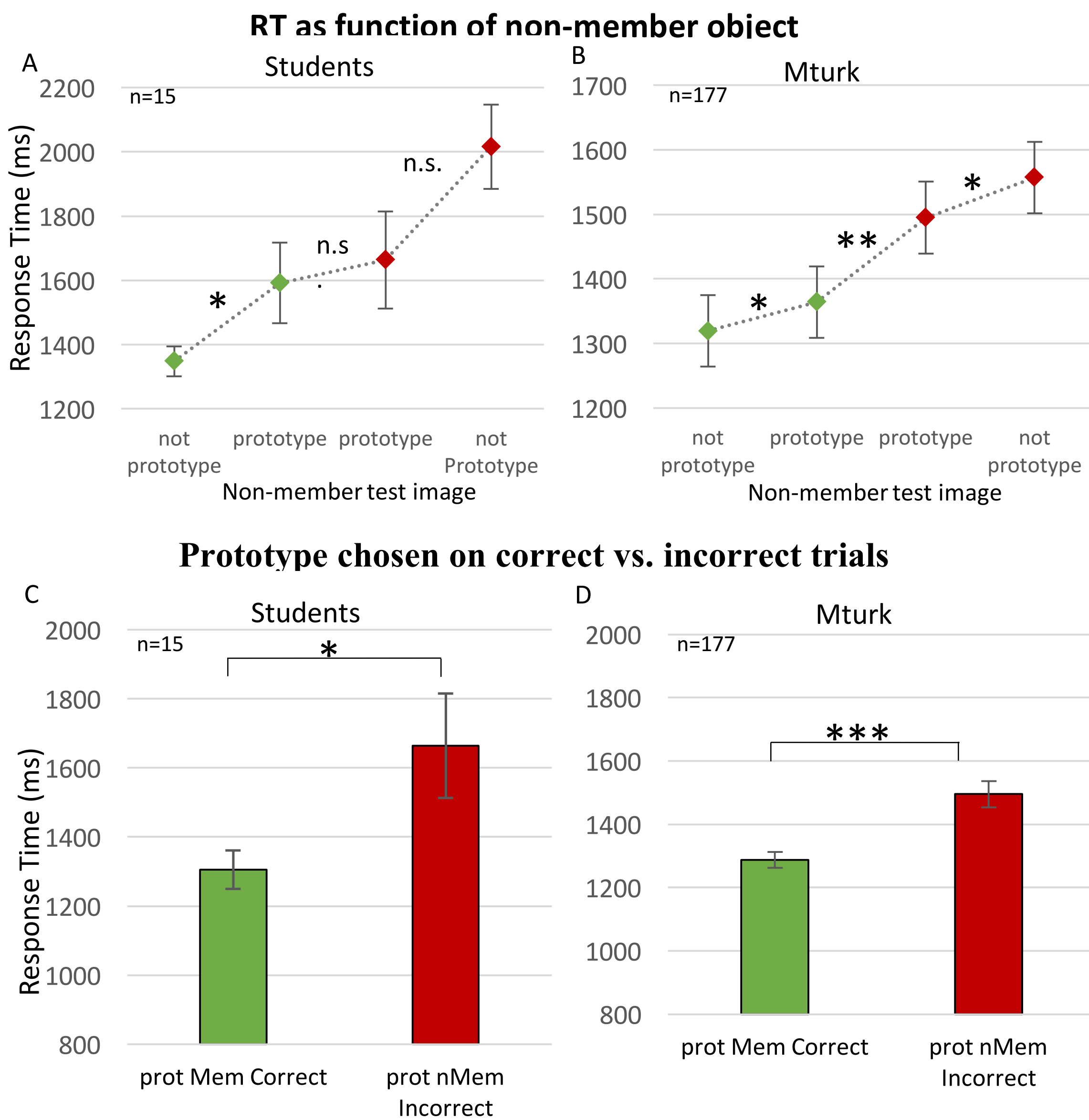

Figure 6: Response Time Prototype Effect. (A) Students: RT for each combination for the non-member test element as prototype, not prototype, correct and incorrect trials. Green and red diamonds represent correct and incorrect trials,

respectively. Left: RT compared for correct trials where the non-member test image is of the prototype of a category (A-Ap) versus all other trials where it is not of the prototype. Right: RT compared for incorrect trials where the non-member test image is of the prototype of a category (A-Ap) versus all other trials where it is not of the prorotype. Middle: RT compared for non-member object being the prototype and participants choosing this image, incorrectly, or the non-prototype member, correctly. (B) Similar graph for Mturk participants. (C) Students: RT comparison between trials with participants picking prototype object images correctly (green bar) versus incorrectly (red bar). (D) Similar graph for MTurk participants.

Range/Boundaries Effect

The second statistic found for low-level sets is the *range effect*, whose equivalent would be representation of category boundaries. A two-way repeated measure ANOVA was performed on accuracy and revealed a highly significant boundary effect, as shown above ($F_{1,14}$=298.64, p<0.001). As with low-level features, accuracy rates in trials of non-member objects outside of category boundaries (Aprot-B and A-B), i.e. non-member objects from a different category than the object sequence, were significantly higher (0.87±0.02) than in trials with both test objects within the category range (Aprot-A, A-A, A-Aprot; 0.65±0.02; p<0.001), as seen in Figure 7A.

This effect was observed also in response time measurements for correct responses, as shown in Figure 7B. Responses were significantly faster for trials where the non-member object was outside category boundaries, i.e., belongs to a different category (1279±54ms), than in trials where both test objects were from the category of the RSVP sequence (1476±65ms; p<0.01). Taken together, the increase in accuracy and decrease in RT indicate a consistent trend of reducing task difficulty by introducing non-member test objects from a different category, rather than a speed-accuracy tradeoff.

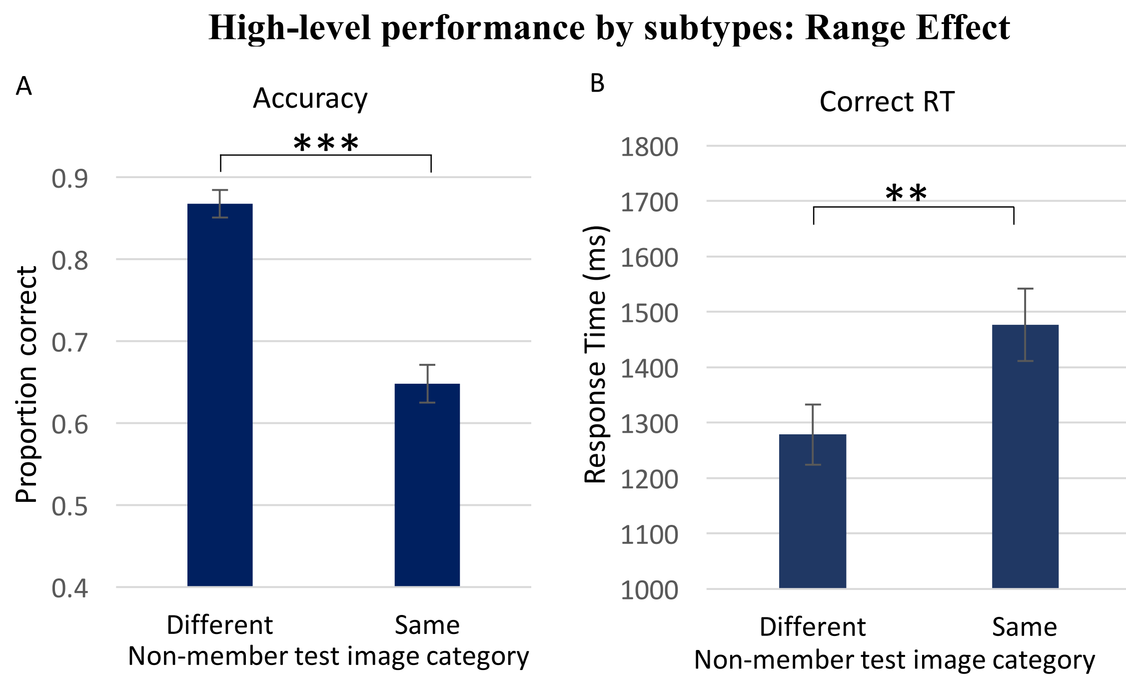

Figure 7. Range (category) Effect - Within vs. Between Category differentiation (Students). (A) Average accuracy for subtypes A-B and Aprot-B versus subtypes Aprot-A, A-A, and A-Aprot. Observers were more accurate when the non-mmber test object was from different category than that of the RSVP sequence. **(B)** RT of correct trials was significantly faster when the non-member object belonged to a different category than when both test objects belonged to the RSVP sequence category.

Experiment 2. Scoring object typicality.

So far, we have compared results for category and set sequence member recall and effects of prototype – mean, and boundaries – range edge on choice of member image in a 2-AFC task. In addition, Khayat & Hochstein (2018) measured how these mean and range effects are graded with the distance of the test item from the mean or from the range edge. To complete and quantify the comparisons, we would like to do the same for the prototype and category effects seen here. To this end, we need a measure of the distance of our test objects from their category prototype. (It would be nice to measure how far away from a category are objects from different categories, but this seemed too difficult for the present study.)

The current experiment was therefore designed to measure the subjective distance of objects from their category prototype – and to learn for each category which object is the prototype itself. To this end, we asked 50 MTurk participants to choose one of two image objects as a member of a previously named category, and used their response speed as a measure of the closeness of the object to the prototype. We will then use these results in Experiment 3 to measure the graded prototype effect. It has been well documented that responses are faster for prototypes than for non-prototypes (Ashby & Maddox, 1991; McCloskey & Glucksberg, 1979; Rips, Shoben, & Smith, 1973; Rosch, Simpson & Miller, 1976).

Methods

Stimuli and procedure

We present the name of a category in the middle of the screen for 1s, (font: Arial 32, white), followed, after 1.0s, by two test images, one of an object belonging to the named category, and one of a different category (attempting to choose objects that were from a different category but not too far from the named category). Images were presented to the left and right of the center of the display, in the middle half of the width and height of the screen. Images remained present until observer response.

Observer task was to choose, by key press, the image with an object that belongs to the named category. We hypothesize that the closer the object is to the category prototype, the faster will be the response, expecting participants to recognize prototypical objects as members of the named category quicker than they do atypical members. For example, participants will recognize an apple as a fruit faster than a kiwi, a cow as a mammal faster than a dolphin, and baseball as a sport faster than mountain climbing.

We tested 50 Amazon Mechanical Turk participants (MTurks). Participants performed two sessions of 300 trials/session. They were tested on 20 categories, as indicated in Table 1 (starred categories), 10 categories per session, with 30 test objects for each category.

Figure 8. Examples of category objects and their associated response times. Four example objects are shown for each of the categories of food, cars, birds,

animals and clothing, with the mean RT over 47 observers. We assume that shorter RTs are associated with objects that are closer to the prototype, and use the RT ranking of objects for each category as a measure for its typicality.

Results

As expected, response times varied among objects (maximum: 2.04s, minimum: 0.65s; mean range for 20 categories: 0.65s) and there was significant correlation among participants; (mean standard error between participants was 6% of the RT).

Examples of categories and their objects are shown in Figure 8. For each category, four objects are shown, and for each, the mean RT measured for our 50 MTurk observers.

We ranked the objects of each category (from 1-30) and computed the mean RT for each rank over all 20 categories. These average RTs were then normalized by: Normalized RT = (RT – minRT) / (maxRT – minRT), where minRT and maxRT are the minimum and maximum RTs for that category, and (maxRT – minRT) is the range of average (across participant) response times for each category. Figure 9 (blue symbols) demonstrates the average normalized RT for each category object rank. There is a high degree of across-category similarity, evidenced by the small s.e. among the categories. Interestingly, RT dependence on rank is steeper at the edges of the category objects, near the prototype (rank = 1) and far from it (rank = 30). We also measured the across-participant ranking, and found small standard deviations (Figure 9, red symbols). We shall now use this ranking as a typicality index for each item in its category, to measure the impact of typicality on object memory in the RSVP sequence test.

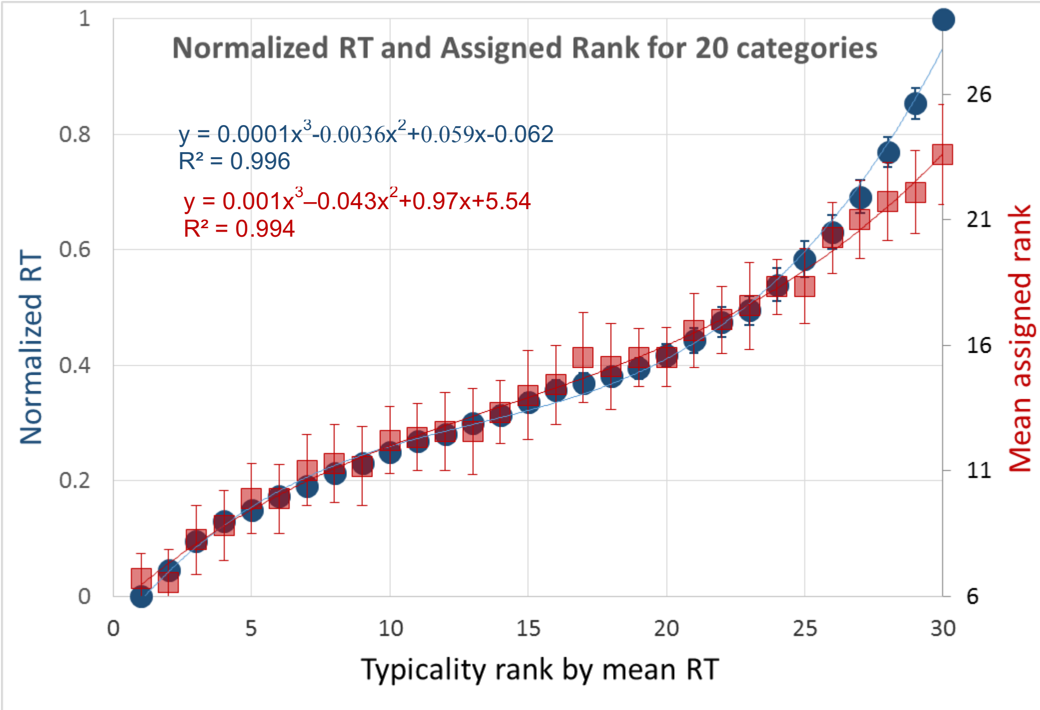

Figure 9. Average normalized RT for each category object rank. Objects were ranked from 1-30 for each category according to the RT in the scoring object typicality test, where observers simply indicated which of two objects belonged to a previously-named category. We then normalized the actual RTs, averaged over participants, and compare the result with the ranking (blue). The fit of the two measures is very good with good agreement among participants. Also shown is the mean and standard deviation across partocipants of the rank assigned to each objects (red). The results match closely the mean RT data, and the across stamdard deviation is small, confirming the methodology.

Experiment 3

Having derived a measure of the distance of each object from its category prototype – the typicality index – we now use this index to measure the impact of typicality on memory of objects in a previously seen sequence. For low-level objects (Khayat and Hochstein, 2018), it was easy to measure the distance of each element from the mean of the sequence since the elements differed by a measurable feature (orientation, brightness, size; see Figure 1B). We found there, as shown in Figure 10B,D, that the mean effect is graded. That is, as the member element is closer to the mean, so it is preferably chosen as the member (Figure 10B). Similarly, as the non-member is further from the mean, so it is rejected as not being the member (Figure 10D). We now ask if this same rule applies to category objects. We have seen the prototype effect in Figure 5 as a preference to choose as member, objects that are exactly the prototype of the category. Is this effect also graded?

For Experiment 3, we tested MTurk participants (see above, Methods) with the 20 starred categories in Table 1 and tested in Experiment 2. We use the mean across-participant RT found in Experiment 2 as the basis for the typicality ranking of objects for Experiment 3. Note that different MTurk participants were tested in Experiments 2 and 3; (Experiment 1 was with in-house student participants). For Experiment 3 all objects presented in the test pairs were from the same category as the previously presented sequence, so that we are now testing the graded prototype effect, and not the range effect (seen in Experiment 1; Figures 4 and 7).

Figure 10 displays the graded prototype effect. We measure the proportion correct, that is the probability of choosing the member object as having been seen in the category sequence, as a function of the typicality index of the member object (Figure 10A). Typicality is ranked from 1 to 30, where 1 is the closest to the prototype, i.e. the shortest average RT measured in Experiment 2. Note the gradual decrease in choosing the member as it is further from the prototype. Similarly, as the non-member is gradually further from typical, i.e. the mean RT to this object was greater in Experiment 2, so this object is more often rejected, and is less often chosen as the member (Figure 10C).

Despite the Experiment 2 nonlinear dependence of typicality rank on image RT, Figure 10A,C data fit well a linear regression. This may be due to the near linearity of the Figure 9 curve, except at its extremes, and because Figure 10A averages over nonmember rank, Figure 10C over member rank; (Fig. 11A over both).

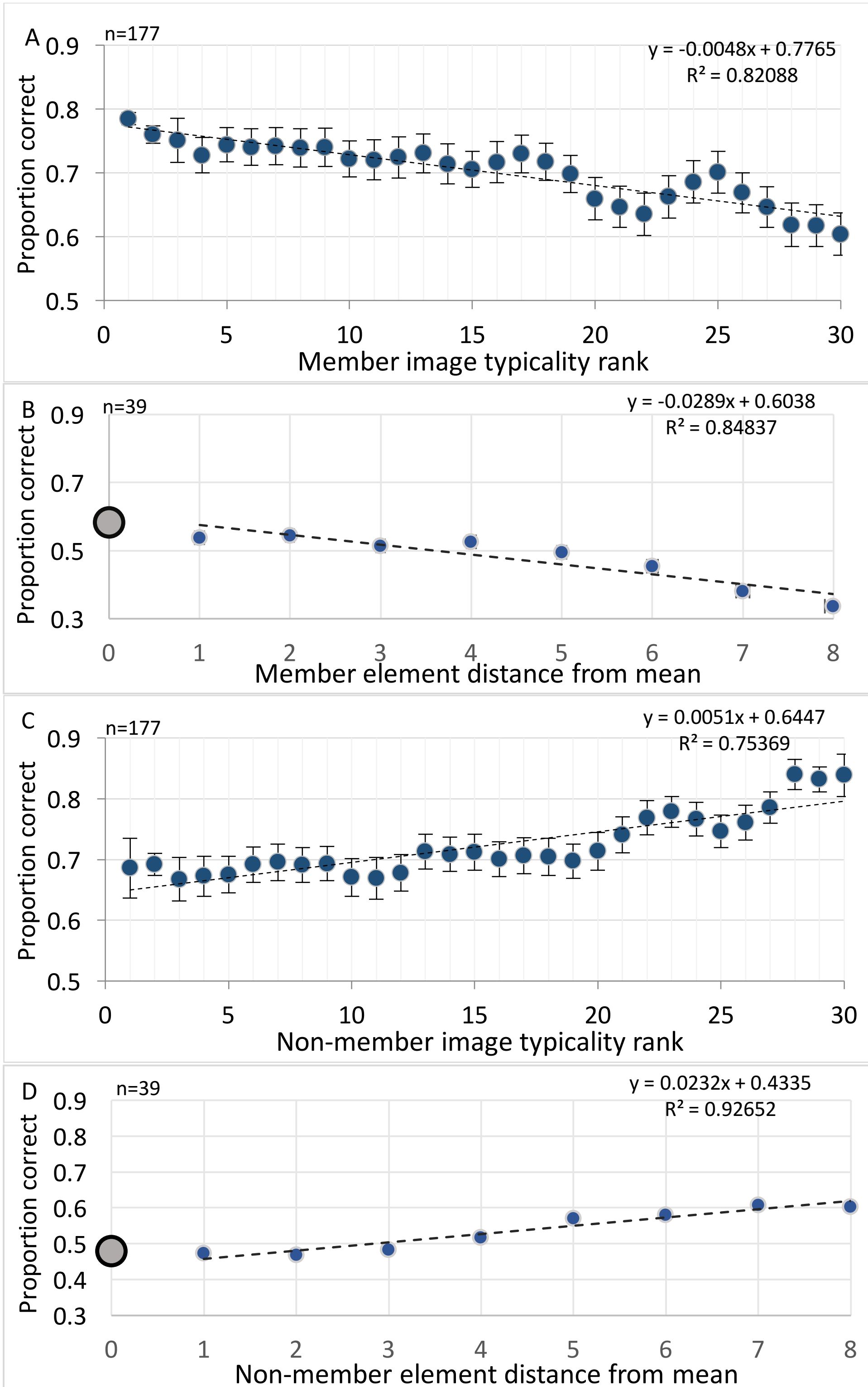

Figure 10. Graded prototype effect. (A) Proportion correct as a function of the typicality index of the member test object where typicality is ranked from 1 to 30

(1 closest to prototype). **(B)** Similar graph for low-level feature experiment (from Khayat & Hochstein, 2018). Proportion correct as function of member test-element distance from set mean. Note similar gradual decrease in probability of choosing the member as it is further from the prototype/mean. **(C)** Proportion correct as a function of the typicality index of the non-member test object as it is gradually further from typical, so that this object is more easily and more often rejected, i.e. less often chosen as the sequence member. **(D)** Similar graph for low-level feature experiment. Note similarity between low-level feature and high-level categorization effects.

The choice of an image is not dependent only on that image, however, since there are always two images displayed and we ask participants to choose between them. Thus, the relative measure between the two images should determine which image participants choose. Having found that sequence member object closeness to the prototype, and sequence non-member distance from the category prototype both add to correct choice of the member, we now plot choice accuracy as a function of the difference between the distances of the non-member and the member. This is shown in Figure 11A, where we also show the parallel graph for low-level features (11B; from Khayat & Hochstein, 2018).

These graphs, including the high-level categorization graphs, are not without noise. Noise comes from the random second image in the membership tests, from inter-participant differences, and from the very nature of our using RT as a determinant for typicality. Nevertheless, the good fit to a single trendline suggests that our conclusion is well founded, as follows. When viewing a sequence of objects belonging to a single category, observers often fail to recall the identity of each object seen, and instead, when asked which of two objects was included in the sequence, depend, on recognition of the category seen, knowledge of the prototypical object, and estimation of the distance of the two test objects from the category prototype.

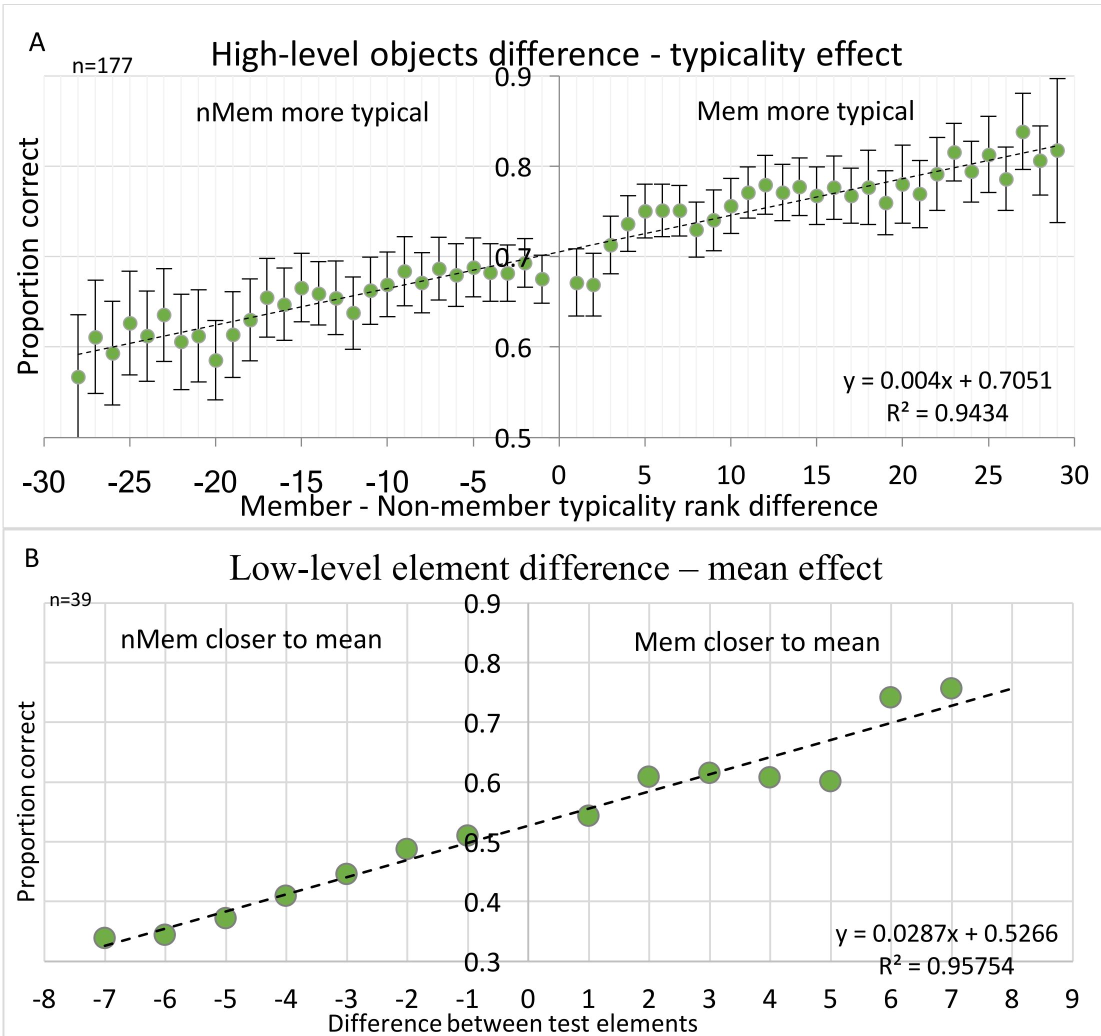

Figure 11. (A) Accuracy as a function of the difference between the distances of the non-member and the member objects from typicality (considering that correct choice in the membership test depends on both the member and the non-member image distances from the prototypical. **(B)** Parallel graph for low-level features (Khayat & Hochstein, 2018).

Discussion

The current results confirm and extend those of recent studies suggesting that statistical representations generalize over a wide range of visual attributes, from simple features to complex objects, giving accurate summaries over space and time (Alvarez & Oliva, 2009; Ariely, 2001; Attarha & Moore, 2015; Chong & Tresiman, 2003; Gorea, Belkoura & Solomon, 2014; Haberman & Whitney, 2009; Hubert-Wallander & Boynton, 2015). This result is now extended to object categories, as well. These efficient representations overcome severe capacity limitations of perceptual resources (Alvarez & Oliva, 2008; Robitaille & Harris, 2011), they are formed rapidly and early in conscious visual representations (Chong & Treisman, 2003), without focused attention (Alvarez & Oliva, 2008; Chong & Treisman, 2005) and without conscious awareness of individual stimuli and their features (Demeyere et al., 2008; Pavlovskaya et al., 2015). Thus, their underlying computations play a fundamental role in visual perception and the rapid extraction of information from large and complex sources of data. In particular, we propose that categorization mimics set summary statistics perception processes, sharing its characteristics. Note that rapid gist perception does not imply low cortical level representation, on the contrary, it is the result of rapid feedforward computation along the visual hierarchy (Hochstein and Ahissar, 2002)

Regarding high-level categories, we revealed two phenomena that match those found for low-level features, by using a similar experimental design for the two experiments: a RSVP sequence followed by a 2-AFC experiment test of image memory. (1) Typicality effect – the typicality level of an object was well represented, as it biased participants' decision towards choosing the more typical exemplar (of the presented category) as the member of the RSVP sequence. The typicality effect led to faster and more accurate responses for member test items, and also to choice of the incorrect item, when it had superior typicality (Figures 4-6; 10-11). Thus, the more typical object was chosen as present in the sequence, whether it was or was not actually present there. The typicality effect is similar to the *set mean value effect* found for low-level features. (2) Boundary effect – categorical boundary representation assisted participants in rejecting images with objects which do not belong to the category of the RSVP sequence, therefore correctly choosing the member image and achieving higher performance levels in these trials (Figures 4, 7). This effect is similar the *set range edges effect.*

Furthermore, using a dedicated response time test to rank the typicality of items within their category, we find that the typicality effect is graded, similar to the set mean value effect (Figures 10-11). The degree to which observers preferentially choose category items as having been members of the trial sequence is directly related to the degree of typicality of the test items. Both member and non-member items are chosen more frequently as they are closer to prototypical; member items, correctly, and non-member items, incorrectly. In particular, the relative typicality of the member test item versus the non-member test item strongly affected observer choice of which item they reported as member of the sequence (Figure 11). Participants associated the more typical object to the displayed RSVP sequence, regardless of whether the prototype actually was or was not a member of the set. It is as if, when viewing the sequence of objects, they perceived the category but had only a poor representation of its individuals. This is exactly what was found for set perception (Khayat & Hochstein, 2018; Ward, et al., 2016; but see Usher, et al., 2018).

We propose that participants unconsciously considered prototypes as better representatives of the categories than less typical exemplars and correspondingly chose them as members of the sequence, perhaps because prototypes usually contain the most common attribute values shared among the category members (Goldstone & Kersten, 2003; Rosch & Mervis, 1975).

As in the low-level experiment, participants were not informed about the categorical content of the RSVP sequences and so they had no knowledge concerning the involvement of prototypes, categories etc., and they only followed the instructions of an image memory task. The similarity of the effects emerging from the two experiments implies that statistical and categorical representations are cognate phenomena that share perceptual characteristics, and perhaps are generated by similar computations.

Note that both the category prototype and boundary effects are based on participants' implicit categorization, extracted from the images in the RSVP sequences.

The results indicate that they adjusted their responses towards the relevant category, even though they were not guided to take category information into consideration in the alleged memory test. While participants concentrated on the RSVP images, themselves, it seems that category context extraction overcame the cognitive abilities of memorizing the objects or scenes presented by the images.

Nevertheless, we note that accuracy in this experiment was superior to that in our previous set summary statistics experiment (compare Figures 4 and 2; Khayat & Hochstein 2018). This may well be due to accurate memory of some sequence items, which is easier for object images than for abstract items (circles, disks or line segments), which differ only in size, brightness or orientation. This result also confirms that participants are trying to recall the actual objects displayed in the sequence – they sometimes succeed in remembering them – and they are not consciously trying only to categorize the images.

Categorical perception is often influenced by context (Barsalou, 1987; Cheal & Rutherford, 2013; Joubert et al., 2017; Koriat & Sorka, 2015, 2017; Roth & Shoben, 1983). *Water*, for example, may be associated with different categories, depending on context. It is a drink, a liquid for bathing or cleaning, or the medium of marine animals. Thus, the category to which participants associated each sequence object would naturally be affected by other sequence objects. We conclude that the current categorization processes occurred rapidly and intuitively, based on the variety of sequence objects, but also on earlier processing of interactions between objects and their contexts (Barsalou, 1987; Joubert et al., 2007; Koriat & Sorka, 2015, 2017; Roth & Shoben, 1983).

Differences between low-level parameter sets and high-level categories

There are several differences between the low-level and the high-level results that should be pointed out. For the low level, we measured not only the graded mean effect, but also the graded range effect, i.e. the gradual effect of the distance of the presented non-member element from the edge of the range of the presented sequence. This range effect has its equivalent in the boundary effect seen in Figure 4, above. To extend this to a graded effect would require measuring the distance between an object of one category from the "edge" of a different category. This is beyond the scope of the current study.

A second difference to be noted is that it is easier to remember particular pictures of objects than specific elements in a sequence that differ only in a low-level feature (orientation, size or brightness). Thus, as mentioned above, performance in the high-level test is superior overall. (Note performance axis difference between Figure 10A,C and B,D.)

Another significant difference between testing the low-level set features and the high-level category objects is that the set of low-level elements, and their range and mean, are determined on-the-fly for each trial, by the sequence of stimuli actually presented. In contrast, the high-level categories are, of course, learned from life experience, and their prototype and boundaries are known immediately when seeing the first object in the sequence, (or first few if the category is ambiguous). Categorization is thus pre-determined, and not a result of the experience in the experiment itself. At the same time, there may well be inter-participant differences in the way they categorize objects, and, in particular, in the specific objects that they consider prototypical.

Related to the latter two differences is another. Categories are often denoted and remembered by their name, introducing a semantic element to the association of a variety of objects to a single category. This is not so for the low-level features studied previously. Nevertheless, recall that the world contains, naturally and intrinsically, objects that cluster separately in feature space, and thus categories that are language-independent (Goldstone & Hendrickson, 2010; Rosch et al., 1976).

Implications for categorization processes

There is ongoing debate concerning category representation in terms of the boundaries between neighboring categories, in terms of a single prototype (category members resemble this prototype more than they resemble other categories' prototypes), or in terms of a group of common exemplars (new objects belong to the same category as the closest familiar object). Our finding that participants respond on the basis of both the mean and range of sets, and similarly on the basis of the prototype and boundary of object categories may suggest a hybrid process model.

Concerning the single prototype versus multiple exemplar theories, our results may support prototype theory, since we find that participants choose test objects that are more prototypical, rather than recalling viewed exemplars. Nevertheless, category prototypes may be a secondary read-out of fuzzy representations of multiple exemplars (see below). Our results resemble the Deese-Roediger-McDermott (DRM; Roediger & McDermott, 1995) finding that when presented with a list of related words, participants recall a non-presented "lure" word with the same frequency as the presented words. In the DRM paradigm, participants study lists of words (e.g. tired, bed, awake, rest, dream, night, blanket, doze, slumber, snore, pillow, peace, yawn and drowsy) that are related to a non-presented lure word (e.g. sleep). On a later test, participants often claim that they previously studied the related lure words. Similarly, it was found that after learning a set of distortions of a random dot pattern, participants learn the undistorted pattern – the prototype – more easily than a new distortion, though only after a first viewing (Posner & Keele, 1968). These results may be added to the ensemble and categorization results, relating different situations – semantic and perceptual – where perceiving related items induces representation and recall of the mean or prototype processes, suggesting that similar processes may underlie them.

Such recall is referred to as "false" memories, since false recognition of the related lure words is indistinguishable from true recognition of studied words (Gallo 2006; Schacter & Addis, 2007). Our results, too, are "false" memories, since participants indicate recall of items that were not presented in the sequence. This is equally true for our study of category prototype recall and our studies, and those of many others, of set ensemble presentation and recall of the set mean – even in its absence from the presented sequence. Nevertheless, the term "false memory" is generally used in reference to recall of events and narratives that did not occur or were not related.

Perceiving category exemplars in terms of the category prototype may be the source of categorical priming (e.g. Ray, 2008; Fazio, Williams, & Powell, 2000), whereby responses to unseen exemplars (and in particular to the category prototype) are faster when primed by previously perceiving another category exemplar. Interestingly, similar effects have been found for sets (Marchant & de Fockert, 2009), and there is even negative priming for unconscious viewing of single unusual shapes (DeSchepper & Treisman, 1996).

Conclusions:

We conclude that while observing the projected images, participants first, implicitly generalized them into a category. Then, at the membership test, they use this categorical context to classify the probability of presence within the sequence of the test images. That is, when visual memory capacity is insufficient, then this implicit categorical context affects their judgment. If indeed categorizations are executed by similar computations as in statistical perception of the visual system, then it is possible that these are only particular embodiments of a general system, which efficiently determines our perception and behavior. It is especially poignant that set mean perception and categorization, which help behaving in a too-rich and too-complex environment by applying shortcuts to perception, may share perceptual-computational mechanisms, perhaps at different cortical levels. We have suggested that the neural mechanism used is a population code (Georgopoulos et al. 1986) that encodes both the mean and the range of the stimulus set (Hochstein, 2016a,b; Pavlovskaya et al., 2017a, b; see also Brezis et al., 2016, 2018). Using a population code to determine set mean answers the question of how the visual system computes mean values without knowing values for each element separately, (whether represented when viewed and forgotten, or never explicitly represented). Due to broad tuning and overlap of neuron receptive-fields, a population code is necessarily used for perceiving individual element values and may be used directly, with a broader range of neurons over space and time, to perceive set mean values. We now suggest the same type of population code may be used for categorization. Category prototype and boundaries could be the read out of fuzzy representations of multiple exemplars. It has already been suggested that ensemble summary statistics might serve as the basis for rapid visual categorizations (Utochkin 2015).

A distinction was made between automatic, intuitive global-attention scene gist perception, using vision at a glance, versus explicit, focused-attention vision with scrutiny (Hochstein & Ahissar, 2002). Gist is acquired automatically and implicitly, by bottom-up processing, and details are added to explicit perception by further, top-down guided processes. The current study demonstrates that even when it is observers' intention to detect and remember the details of each image in a sequence – an intention that in this case often leads to failure – nevertheless, the automatic, implicit process of gist perception succeeds in acquiring both set and category information.

A question that still needs to be addressed is the cerebral correlates of mechanisms underlying these processes. An investigation using physiological techniques (fMRI or EEG), while participants perform behavioral tasks, as in the current study, might indicate brain regions or electrophysiological patterns of activity, which are specific to systems that generate these automatic representations. Such a study might also test the notion that similar sites perform set mean and range perception as well as categorization.

Acknowledgements

We thank Yuri Maximov, the lab's talented programmer, and lab co-members Safa' Abassi and Miriam Carmeli. Thanks to Merav Ahissar, Udi Zohary, Israel Nelken, Robert Shapley, Howard Hock and Anne Treisman (of blessed memory), for helpful discussions of earlier drafts of this paper. This study was supported by a grant from the Israel Science Foundation (ISF).

The data for the experiments reported will be made available on-line, and none of the experiments was preregistered.

References:

1. Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, **101**, 34-40.
2. Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, **19**(4), 392-398.
3. Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences USA*, **106**(18), 7345-7350.
4. Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, **12**(2), 157-162.
5. Ashby, F. G., & Maddox, W. T. (1991). A response time theory of perceptual independence. In: Doignon, J.-P. & Falmagne, J.C., eds. *Mathematical Psychology: Current Developments*, pp. 389-413, Springer, New York, NY.
6. Ashby, F. G., & Maddox, W. T. (2011). Human Category Learning 2.0. *Annals of the New York Academy of Science,* **1224**, 147–161.
7. Attarha, M., & Moore, C. M. (2015). The perceptual processing capacity of summary statistics between and within feature dimensions. *Journal of Vision*, **15**(4), 9.
8. Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In: Neisser, U., ed., *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, pp 101-140, Cambridge U Press, Cambridge, UK.
9. Bauer B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, **59**(2), 171.
10. Bauer, B. (2015). A selective summary of visual averaging research and issues up to 2000. *Journal of Vision,* **15**(4):14, 1–15
11. Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433–436.

12. Brezis, N., Bronfman, Z. Z., Jacoby, N., Lavidor, M., & Usher, M. (2016). Transcranial direct current stimulation over the parietal cortex improves approximate numerical averaging. *Journal of Cognitive Neuroscience* **28**(11), 1-14.

13. Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, **5,** 10415.

14. Brezis, N., Bronfman, Z., & Usher, M. (2018). A perceptual-like population-coding mechanism of approximate numerical averaging. *Neural Computation*, **30**, 428–446.

15. Cheal, J. L., & Rutherford, M. D. (2013). Context-dependent categorical perception of surprise. *Perception*, **42**(3), 294-301.

16. Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, **43**(4), 393-404.

17. Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Attention, Perception, & Psychophysics*, **67**(1), 1-13.

18. Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, **20**(9), 324-335.

19. Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta psychologica*, **138**(2), 289-301.

20. Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, **24**(1), 154-176.

21. Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science*, **21**, 234–242.

22. Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, **46**(11), 2861-2864.

23. DeSchepper, B., & Treisman, A. (1996) Visual Memory for Novel Shapes: Implicit Coding Without Attention, *Learning, Memory, and Cognition*, **1996**, 22 (1), 27-47.

24. Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, **2,** 243, 1-12.

25. Fazio, R. H., Williams, C. J., & Powell, M. C. (2000) Measuring associative strength: Category-item associations and their activation from memory. *Political Psychology*, **21**(1), 7-25.

26. Gallo, D. A. (2006). *Associative illusions of memory*. New York, NY: Taylor & Francis.

27. Georgopoulos, A. P., Schwartz, A.B., & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science,* **233**, 1416–1419.

28. Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, **1**(1), 69-78.

29. Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. In, Weiner, I. B. ed., *Handbook of Psychology*, pp. 597–621. Wiley, Hoboken, NJ.

30. Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, **14**(9), 22, 1-14.

31. Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, **17**(17), R751-R753.

32. Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, **35**(3), 718-734.

33. Haberman, J. & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In, Wolfe, J. & Robertson, L., eds. *From Perception to Consciousness: Searching with Anne Treisman*, pp 339–349, New York, NY: Oxford University Press.

34. Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: What makes the difference? *Cognition*, **112**(1), 105-119.

35. Hochstein, S. (2016a) The power of populations: How the brain represents features and summary statistics. *Journal of Vision* **16** (12), 1117.

36. Hochstein, S. (2016b) How the Brain Represents Statistical Properties. *Perception* **45**: 272.

37. Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, **36**(5), 791-804.

38. Hochstein, S., Khayat, N., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2018a). Set Summary perception, outlier pop out, and categorization: A common underlying computation? *European Conference on Visual Perception* 2018, Trieste, Italy, online.

39. Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision* **15** (4), 7, 1-17.

40. Hochstein, S., Pavlovskaya, M., Bonneh, Y., & Soroker, N. (2018b). Comparing Set Summary Statistics and Outlier Pop Out in Vision. *Journal of Vision*, 18(13), 12, 1-13.

41. Hock, H. S., Gordon, G. P., & Whitehurst, R. (1974). Contextual relations: the influence of familiarity, physical plausibility, and belongingness. *Perception. & Psychophysics.*, **16**, 4–8

42. Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, ***15***(4), 5, 1-12.

43. Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2015). Basic level category structure emerges gradually across human ventral visual cortex. *J Cognitive Neuroscience,* **27**(7), 1–29.

44. Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category representations in object-selective cortex, *Neuroimage,* **134**, 170–179.

45. Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness & Cognition*, **48**, 149-160.

46. Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, **47**(26), 3286-3297.

47. Khayat, N., & Hochstein, S. (2018). Perceiving Set Mean and Range: Automaticity and Precision. *Journal of Vision*, 18: 23: 1-14.

48. 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.

49. Koriat, A., & Sorka, H. (2017). The construction of category membership judgments: Towards a distributed model. In Cohen, H. & Lefebvre, C., eds. *Handbook of Categorization in Cognitive Science,* 2nd Edition, pp. 773-794, Elsevier, Amsterdam, Netherlands.

50. Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky. H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron,* **60**(6):1126–1141.

51. Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, ***1***(2), 115-121.

52. Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390**(6657), 279-281.

53. Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Attention, Perception, & Psychophysics*, **53**(1), 49-70.

54. Marchant, A. P., & de Fockert, J. W. (2009). Priming by the mean representation of a set, *The Quarterly Journal of Experimental Psychology*, **62**:10, 1889-1895

55. McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, **6**: 462-472.

56. McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, **11**(1), 1-37.

57. Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**(12), 1469-1481.

58. Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**(3), 333-352.

59. Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11), 1-8.

60. Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, **128**(1), 56-63.

61. Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews,* **32**(2), 279-291.

62. Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In, Pothos, E. M., & Wills, A. J., eds., *Formal Approaches in Categorization,* pp. 18-39. Cambridge University Press, New York, NY.

63. Oliva, A. & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. In, Martinez-Conde, S., Macknik, S.L., Martinez, L.M., Alonso, J.-M., & Tse, P.U., eds. *Progress in Brain Research, Visual Perception, Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception,* **155B**, 23-36.

64. Pavlovskaya, M., Soroker, N., Bonneh, Y. S., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, **27**(7), 1397-1411.

65. Pavlovskaya, M., Soroker, N., Bonneh, Y., & Hochstein, S. (2017a). Statistical averaging and deviant detection in heterogeneous arrays. 40th European Conference on Visual Perception Abstracts, 40, 160.
66. Pavlovskaya, M., Soroker, N., Bonneh, Y., & Hochstein, S. (2017b). Statistical averaging and deviant detection may share mechanisms. 2017 Neuroscience Meeting, Washington, DC: Society for Neuroscience. 365:4
67. Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77(3), 353-363.
68. Posner, M. I., & Keele, S. W. (1970) Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304-308.
69. Potter, M. C., & Hagmann, C. E. (2015). Banana or fruit? Detection and recognition across categorical levels in RSVP. *Psychonomic Bulletin & Review*, **22**(2), 578-585.
70. Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13ms per picture. *Attention, Perception, & Psychophysics*, **76**(2), 270-279.
71. Ray, S. (2008). An investigation of time course of category and semantic priming, *Journal of General Psychology*, **135**, 2, 133-148;
72. Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**(3), 382-407.
73. Rips, L. J., Shoben, E. J., & Smith, E. E. (1973) Semantic Distance and the Verification of Semantic Relations, *Journal of Verbal Learning and Verbal Behavior,* **12**:(1), 1-20
74. Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, **11**(12), 18, 1-8.
75. Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**(4), 803-814.
76. Rosch, E. (1973). Natural categories. *Cognitive Psychology*, **4**(3), 328-350.

77. Rosch, E. (1999). Reclaiming cognition: The primacy of action, intention and emotion. *Journal of Consciousness Studies*, **6**(11-12): 61-77.

78. Rosch, E. (2002). Principles of categorization. In, Levitin, D. J., ed., *Foundations of Cognitive Psychology: Core Readings,* pp. 251-270. Cambridge, MA, US: MIT Press. (Reprinted from 1978/1999, pp 189-206).

79. Rosch, E., & Lloyd, B. B. (Eds.). (1978). *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

80. Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**(4), 573-605.

81. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**(3), 382-439.

82. Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance*, **2**(4), 491-502.

83. Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, **15**(3), 346-378.

84. Schacter, D. L., & Addis, D. R. (2007) The cognitive neuroscience of constructive memory: remembering the past and imagining the future, *Philosophical Transactions of the Royal Society B* **362**, 773–786.

85. Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, **7**(6), 246-251.

86. Smith, J. D. (2014) Prototypes, exemplars, and the natural history of categorization. *Psychonomics Bulletin Review,* **21**, 312–331

87. Solomon, J. A. (2010) Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, **10**(14): 19, 1-16

88. Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, **39**(2), 329-337.

89. Usher, M., Bronfman, Z. Z., Talmor, S., Jacobson, H., & Eitam, B. (2018). Consciousness without report: insights from summary statistics and inattention 'blindness'. *Philosophical Transactions of the Royal Society B,* **373**: 20170354.
90. Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, **15**(4), 8, 1-14.
91. Ward, E. J., Bear, A., Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals? The role of statistical perception in determining whether awareness overflows access. *Cognition* **152**, 78–86.
92. Yamanashi-Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, **7**, 13186, 1-10.