Perceptual and Conceptual Processing of Images: A Garner Interference Study Examining the Interaction between Image Format and Semantic Classification of Black-and-White, Greyscale, and Coloured Line-Drawings

Hayden J. Peel[1, 2] & Philippe A. Chouinard[2]

[1]Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

[2]Department of Psychology, Counselling and Therapy, School of Psychology and Public Health, La Trobe University, Victoria, Australia.

*Corresponding author:

Hayden J. Peel, Ph.D.
Brain Health Imaging Centre
Centre for Addiction and Mental Health
Toronto, Ontario
Canada
Phone: x 33157
E-mail: hayden.peel@camh.ca

Keywords: Garner paradigm; Garner interference; semantic processing; perceptual processing; line drawings

Abstract

We used the Garner paradigm to examine the relationship between the perceptual and conceptual aspects of an image for identification. Participants completed tasks that required them to selectively attend to semantic identity (e.g., manmade vs. natural) while ignoring image format, and vice versa, with the irrelevant dimension either held constant (baseline block) or varying simultaneously (filtering block). Experiment 1 compared black-and-white and greyscale drawings (BWvGS) and Experiment 2 compared coloured and greyscale drawings (CvGS). The former assessed the contribution of structural form cues to perception, while the latter assessed the importance of colour. Garner interference was found with accuracy (with more errors in filtering than baseline) in both the BWvGS and CvGS experiments, but this was asymmetric and limited to the semantic classification task. Participants' selective attention to semantic identity failed when the image format varied simultaneously, indicating that conceptual processing is not independent of perceptual processing. However, in the image format task, selective attention to image format was successful despite variation in semantic identity, demonstrating that perceptual processing can occur independently of conceptual analysis. Together, these results demonstrate that the semantic classification of images is influenced by the format of the stimulus.

3

Introduction

Requiring individuals to semantically classify drawings of objects is near ubiquitous in neuropsychological testing batteries, including the Boston Naming Test (Kaplan et al., 1983), Hooper Visual Organization Test (Hooper, 1983), Benton Visual Retention Test (Benton, 1945), and Montreal Cognitive Assessment (Nasreddine et al., 2005). However, our understanding of how we process the meaning and format of images is limited. A notable, paradoxical example of this comes from patients with visual form agnosia, who display impairments in perceptually recognising line drawings and basic shapes (Peel & Chouinard, 2023). Many of these patients are unable to judge the length of lines. This distortion, along with complaints of unclear vision, or things 'blurring together', might explain why recognition of figures composed solely of lines is so poor. Recognition of stimuli tends to improve when other cues, such as colour or texture, are present. For example, one patient could recognise a line drawing of a female lion as an animal but guessed that it was a horse or a dog and identified it as a lion when a yellow colour was added to the picture (Landis et al., 1982). Thus, it is conceivable that recognition might improve if other cues are present to provide diagnostic information, like surface characteristics or other structural cues. Moreover, patients retain knowledge about the objects depicted and can list their uses and functions once they are informed about the object's identity, thus demonstrating that their problems in recognition are perceptual, not conceptual. In contrast, patients with other conditions like associative agnosia demonstrate

4

opposite deficits, where perceptual recognition is intact, but semantically describing objects are deficient (Mendoza 2011). This raises the question, can these aspects of an image – their image format and their identity – be processed independently of one another?

In other words, how integral are the perceptual and conceptual aspects of an image for identification?

A typical view of object recognition implicates early visual areas, such as the primary visual cortex, with fine-grained featural analysis, and later visual areas, like the anterior temporal cortex, with conceptual aspects of processing (Kravitz et al., 2013). However, neuroimaging studies investigating how perceptual and conceptual processing interact in object recognition paint a more complicated picture (Chouinard et al., 2008; Chouinard & Goodale, 2012; Devereux et al., 2013; Martin et al., 2018; Victoria et al., 2019; Davis et al., 2020). For instance, Victoria et al. (2019) found that an extensive network of brain areas, including early visual areas, are involved with conceptual processing in a living/non-living task. The authors also found common activation to visual similarity in non-visual areas, such as the angular gyrus and the dorsomedial prefrontal cortex. This work demonstrates that the perceptual and conceptual aspects of image processing can interact in object processing in both early and later visual areas. Other neuroimaging work has highlighted that the effects of task demand and selective attention might modulate responses in both early and later visual areas (Bugatus et al., 2017; Çukur et al., 2013). In the present investigation, we extend this research by testing selective attention in a well-established psychophysical task that has not been

5

utilised in this context. We tested if participants could efficiently categorise Snodgrass & Vanderwart (1980) images according to their semantic identity (e.g., manmade or natural) while ignoring their image format (e.g., black-and-white vs. greyscale, or greyscale vs. colour), and vice versa, using a Garner speeded classification task (Garner, 1970).

The Garner paradigm examines people's ability to selectively attend to one aspect of a stimulus while a ignoring variation in another, much like the Stroop task (1935). In the Stroop task, people have more difficulty naming colour words when their font is in a different colour. For example, people are slower and make more errors when naming the word RED written in purple font compared to the word RED written in red font. The Garner paradigm exploits a similar effect by examining people's ability to selectively attend to one aspect of a multidimensional stimulus. The typical example of this effect is demonstrated by people being unable to selectively attend to the width of rectangles while ignoring their height (Garner, 1970). In one block of trials, termed 'baseline', people judge the width of rectangles that have a fixed height. In another block of trials, termed 'filtering', people again judge the width of rectangles, but the height of the rectangles can also vary from trial-to-trial. Garner interference is demonstrated with longer reaction times (RT) and / or more errors in the filtering compared to the baseline block. It is unclear from previous research if Garner interference can also occur with semantic identity and image format.

Two comparisons were of interest in the present investigation. The first compared black-and-white and greyscale images and the second compared colour and greyscale images. Previous work with the Garner paradigm by Cant and colleagues (2008; 2009) examined the interaction between form and different surface characteristics. They reported that colour and texture were processed independently of form as they found no evidence of Garner interference and thus considered them separable. However, this work considered form only by its metric size dimensions of height and width. At the level of conceptual judgements, there is evidence that colour can be diagnostic and facilitate recognition. For example, colour has been reported to facilitate object recognition in semantic categorisation tasks for natural objects compared to artifacts (Chouinard & Goodale, 2012; Bramão et al., 2011). Whether Garner interference occurs for conceptual judgements when the image format varies between colour and greyscale remains an outstanding question to be answered.

Moreover, there are other ways to conceptualise form. From an artistic perspective, a 2D image has form only if it gives an impression of a third dimension (Arnheim, 1974). Images of squares and cubes provide examples of images without and with form, respectively. The former depicts width and height two dimensionally while the latter depict width, height, and length three dimensionally. One can also say that line drawings also depict form to various degrees, ranging from simple to complex (Figure 1). Comparing black-and-white (i.e., line drawings) versus greyscale images is of interest in this regard. Line drawings have long been used by humans to depict objects in the real world dating back to 73,000-years-ago (Henshilwood et al., 2018). For the most part, these images tend to be simple – composed of lines conveying a third dimension

7

primarily through occlusion. We refer to these instances as having simple pictorial form. Greyscale images, like those in the updated version of the Snodgrass & Vanderwart (1980) stimulus set, have additional pictorial cues, such as texture and shading, to depict a third dimension, which means they have more complex pictorial form.

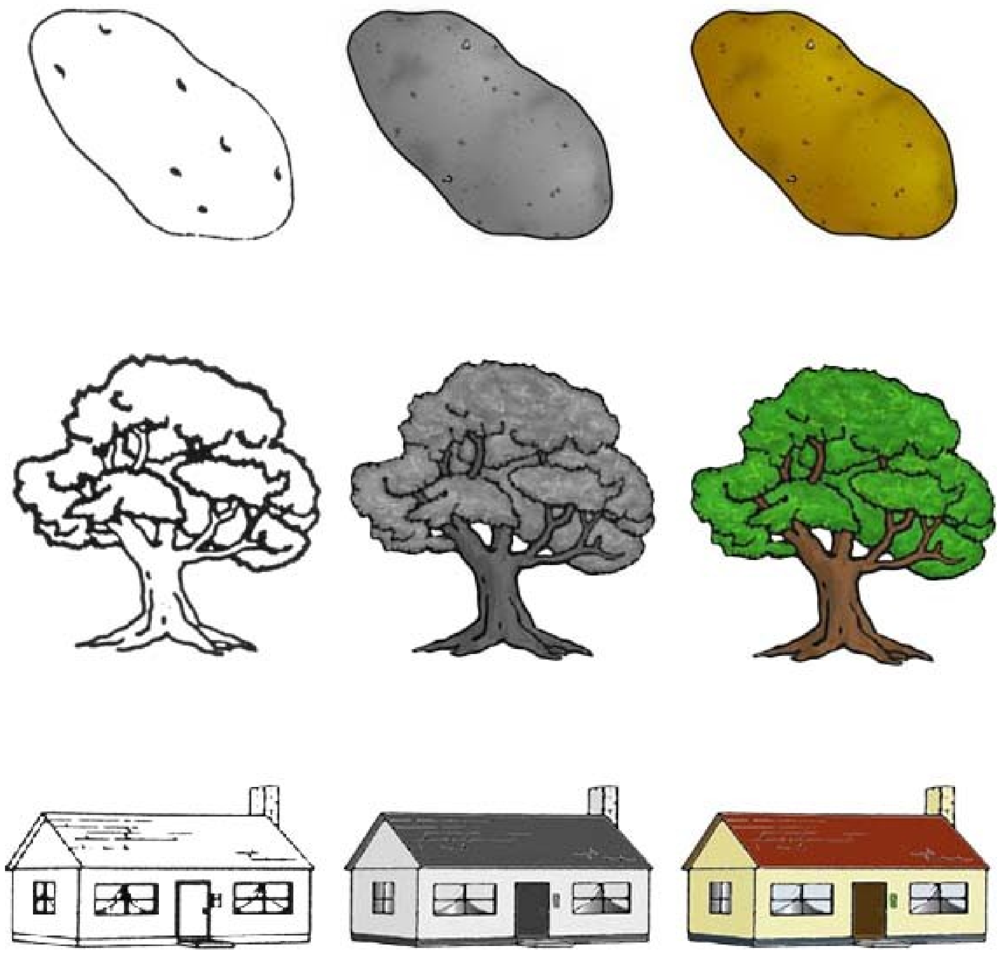

Figure 1. Examples of lines drawings ranging from simple to complex form – original black-and-white, greyscale, and coloured line drawings from left to right of Snodgrass and Vanderwart (1980) images updated by Rossion & Pourtois (2004) to contain coloured and greyscale versions.

In essence, this distinction between simple and complex pictorial form is one of structure. Complex pictorial form provides a richer simulation of real-world structure than simple pictorial form. The effects of

8

adding structural cues on object recognition is incompletely understood. Bonin and colleagues (2019) found that naming latencies in writing were fastest for colour images followed by greyscale images followed by black-and-white line drawings – demonstrating that image formats are processed differently when they are named in writing. To our knowledge, no previous research has examined the effects of image format in a semantic classification task using Garner interference.

We address this knowledge gap. Our primary interest was to test whether semantic classification can operate independently of variation in image format. Given the lack of precedence, we did not have specific predictions about whether Garner interference would be present or not. If interference was observed in the experiment using black-and-white and greyscale images (BWvGS), then that would be evidence for structural cues being important for identifying drawings. If interference was observed in the experiment using colour and greyscale images (CvGS), then that would indicate that colour is also integral for recognition.

Methods

Overview

Half (n = 18) the participants completed the black-and-white line drawings vs. greyscale images (BWvGS) experiment while the other half completed the colour vs. greyscale images (CvGS) experiment (n = 18). Each experiment comprised 6 blocks: 3 image format (2 baseline, 1 filtering) and

9

3 semantic category (2 baseline, 1 filtering) blocks that were completed in a counterbalanced order as determined by a Latin square. Eighteen participants also completed a control experiment verifying that our general procedures can replicate Garner interference with variations in the height and width of rectangles (2 baseline, 1 filtering). These participants completed the control experiment either before or after completing the BWvGS or CvGS experiment in a counterbalanced order. The entire session took between 30-60 minutes. The study was approved by the La Trobe University Human Ethics Committee and was carried out in accordance with the Declaration of Helsinki. All participants provided written informed consent.

Participants

Thirty-six right-handed participants completed the study (M_{age} = 25.87 years, age-range = 19-40, 24 females). We verified that all participants had at least 20/40 vision in each eye using the Snellen Eye Chart (Snellen, 1862) and were right-handed using the Edinburgh Handedness Inventory (Oldfield, 1971).

Stimuli and apparatus

All tasks were presented using E-Prime 3.0 software on a 24″ ASUS VG248QE monitor (Taipei, Taiwan) running at 120 Hz with a display resolution of 1920 × 1080. Participants viewed the stimuli at a fixed viewing distance of 57cm, with their head placed on a chin-and-head rest. Responses were recorded with button pressing on a Chronos Multifunction

10

Response and Stimulus Device (Model PST-100430, Psychology Software Tools).

We used the following procedure to generate a set of images for the experiments. Stimulus pairs (man-made and natural) were created from the Snodgrass and Vanderwart (1980) image set of 280-line drawings, which was more recently updated to contain greyscale and coloured versions of the same items (Rossion et al., 2004) (for examples, see Figure 1). We classified these items as either man-made or natural, with any ambiguous images being removed (e.g., a cartoon heart, a directional arrow etc.). Then, the stimulus set was further reduced according to the existing naming reaction time (RT), naming accuracy, familiarity, and complexity norms published in Rossion et al. (2004). Any item with < 90% line drawing naming accuracy was removed as well as any item where there was a substantial difference between line drawing and greyscale naming accuracy (e.g., the item ‘brush’ dropped from 100% accurate with line drawing to only 64% accurate with greyscale). Items were then clustered according to 100ms increments in RT (e.g., 600-700ms, 700-800ms, etc.) and selectively matched in manmade-natural stimulus pairs as best as possible according to line drawing familiarity, and line drawing and greyscale complexity ratings. Independent samples *t*-tests verified that there was no difference between the manmade and natural items (all $p > .215$) on any of these metrics. The result was a pool of 54 items or 27 pairs of stimuli (for full list, see Appendix 1). Stimuli image files were resized so that all subtended a visual angle of approximately 7.5. For the control task, 4 grey rectangles were created in Photoshop (Adobe Inc. 2019) that varied

11

in height (46 vs. 56mm) and width (44 vs. 52mm) factorially (Figure 2). These dimensions were chosen based on precedent use (Freud et al., 2013).

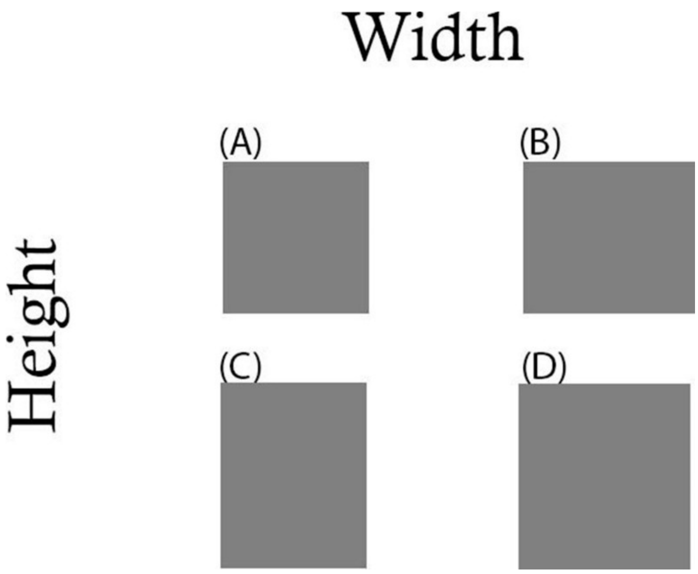

Figure 2. Depicts the 4 grey rectangles used in the control experiment. (A) and (B) are the ‘thin’ and ‘thick’ short rectangles, respectively. (C) and (D) are the ‘thin’ and ‘thick’ tall rectangles, respectively.

Experimental Design

The two main experimental tasks, BWvGS and CvGS, were identical except that coloured images were presented in one while black-and-white line drawings were presented in the other. Before each task, participants were instructed to make their judgments as quickly and accurately as possible.

Baseline: There were four baseline blocks comprising two image format (IF) tasks and two semantic category (SC) tasks. In each of these, one aspect of the stimulus was held constant, while the other varied randomly from trial-to-trial.

12

Image Format (IF)*:* Participants were randomly presented with each image from only one semantic category at a time using E-Prime. For instance, in one block of trials, participants were shown all the "manmade" images, presented once in colour and once in greyscale in the CvGS experiment and once in black-and-white and once in greyscale in the BWvGS experiment. Each baseline condition consisted of 54 trials (27 colour image and 27 greyscale images in the CvGS experiment, 27 black-and-white images and 27 greyscale images in the BWvGS experiment). Participants were instructed to classify each image based on its format by pressing “1” for colour and “2” for greyscale in the CvGS experiment and “1” for black-and-and-white and “2” for greyscale in the BWvGS experiment. Each trial began with a 500-ms fixation point. This was followed by a 500-ms blank screen and then the stimulus. The stimulus remained on the screen for 3000-ms or until a response was made, which ever came first. To prevent expectancy effects, an inter-trial interval was randomly generated by E-Prime, ranging from 1200 to 2000-ms in increments of 200-ms.

Semantic Category (SC)*:* The design and procedures were identical as in the IF baseline, except participants were presented with stimuli in only one image format (i.e., coloured or greyscale in the CvGS experiment and black-and-white or greyscale in the BWvGS experiment) and the semantic category varied randomly from trial-to-trial. Participants classified images according to their semantic category by pressing ‘1’ for natural and ‘2’ for manmade.

13

Filtering: There were two filtering tasks. In the filtering tasks, participants were presented with each image in a similar manner as the baseline blocks. Trials began with a 500-ms fixation. This was followed by a 500-ms blank screen and then the stimulus. The stimulus was shown until a response was made or 3000-ms had elapsed, whichever came first. Each filtering block consisted of 108 total trials. Both stimulus features (image format and semantic identity) varied randomly from trial to trial. In one of the tasks, participants were instructed to focus on the image format while ignoring the semantic category. In the other task, the instructions were reversed with participants being instructed to focus on the semantic category while ignoring image format.

Control Experiment (HvW): The control experiment had the same logic. Participants completed two baseline tasks: One for the ‘short’ rectangles and the other for the ‘tall’ rectangles where only the width varied. The presentation procedures were the same, beginning with a 500-ms fixation, followed by a 500-ms blank screen, and then the stimulus appearing until a response was made or 3000-ms elapsed, whichever came first. In the ‘short’ baseline block, participants saw only the short rectangles (both 46 mm tall) that varied in width (44 vs. 52mm). Each rectangle was presented 27 times for 54 trials in total. Similarly, in the ‘tall’ baseline block, only the tall rectangles (both 56mm tall) were presented, varying in width (again 44 vs. 52mm) and each being presented 27 times for 54 trials in total. Participants also completed a filtering block, where height and width varied randomly from trial-to-trial. All 4 rectangles were presented 27 times each in a random order for a total of 108 trials.

Statistical Analyses

The data were analysed using the *lme4* package in R (R Core Team, 2022). The RT data for correct responses were logarithmically transformed to reduce skewness. Outliers were identified and removed based on the interquartile range (IQR) method. Specifically, the first and third quartiles (Q1 and Q3) were computed, and the IQR was calculated as the difference between Q3 and Q1. RT values were considered outliers if they were below Q1 minus 1.5 times the IQR or above Q3 plus 1.5 times the IQR. Data points falling outside this range were excluded from further analyses. The data were then analysed in two stages.

First, the data were entered into a series of linear mixed-effects models (LMMs) comparing each block to examine Garner interference (filtering > baseline1 = baseline2) and to test if the baseline tasks were significantly different. The latter point was important to verify. If the baseline tasks were mismatched, then unequal discriminability could cause a mandatory failure of selective attention and any further analysis could be inappropriate (Burns, 2014). Block (baseline1 vs. baseline2 vs. filtering) was entered into a model along with random intercepts and slopes for participant and item. These random factors were included given the potential for individual variability at both the participant and stimulus level when responding to the images. In all cases, we specified the 'maximal model' structure for random effects by including both intercepts

15

and slopes (Barr et al., 2013). However, as recommended by Matuscheck (2017), in cases where the model failed to converge, or where singularity of model fit was a concern due to near perfect correlations between slopes and intercepts, slopes were excluded to have the most parsimonious model capturing the data to avoid loss of power with the maximal model structure. The same analysis, with Block as a fixed factor and participant and item as random factors, was done for each version of the task (IF and SC) and each experiment (BWvGS and CvGS), resulting in four LMMs. The p–values for fixed effects were calculated from an F test and t test based on Statterthwaite's approximation. Any significant effects were further examined with *post hoc* pairwise comparison of means, with a Bonferroni adjustment to correct for multiple comparisons.

In stage 2, the baseline tasks were combined into a single variable (Baseline Combined) and entered into another LMM with random slopes and intercepts for participant and item, and Block (Baseline Combined vs. Filtering), Task (IF vs. SC), and their interaction as fixed effects. This was done for the BWvGS and CvGS experiments. Again, we specified the maximal model structure without loss of power, and all instances where models failed to converge are described where applicable. From there, p–values for fixed effects were calculated from an F test and t test based on Statterthwaite's approximation, and any significant effects were further examined with *post hoc* pairwise comparison of means with a Benjamini-Hochberg False Discovery Rate (FDR) correction.

As Garner interference is typically also inferred by increased error rates in the filtering compared to baseline blocks, we analysed accuracy in

16

a similar manner as the RT data. However, given the accuracy data was discrete and comprised scores of either 0 or 1, we performed a series of Generalised Linear Mixed Models (GLMMs). With the GLMM, we specified that the distribution of the dependent variable (i.e., accuracy) was binomial (Brown, 2021). GLMM fit by maximum likelihood (Lapace Approximation) was implemented, and the logit mixed effects model included random intercepts and slopes for participants and items. For the first stage of the GLMM, only Block (baseline1, baseline2, filtering) was included as a fixed effect. For the next stage, Block (Baseline Combined vs. filtering), Task, (SC vs. IF), and their interaction were included as fixed effects. The significance of fixed effects was calculated using Type II Wald Chisquare tests, and any interactions were further examined using *post hoc* pairwise comparison of means, which corrected for multiple comparisons with a Bonferroni and Benjamini-Hochberg FDR adjustment for the first and second stages, respectively.

Finally, the RT and accuracy data from the HvW control experiment were analysed using the same LMM and GLMM described earlier, with Block (Baseline Tall vs. Baseline Short vs. Filtering) as a fixed factor and participant and item as random factors. *Post hoc* pair-wise comparison tests, which corrected for multiple comparisons with a Bonferonni adjustment, were carried out to examine further any interactions and effects deemed significant in the LMM and GLMM.

Results

17

RT

Stage 1: LMMs included Block (Baseline1 vs. Baseline2 vs. Filtering) as a fixed factor and participant and item as random factors. However, only the participant random factor included both intercepts and slopes. The item slopes for Block were excluded from the model as they were highly correlated ($r > .99$) with the intercept for item. This resulted in an almost singular fit, where the model failed to converge, and was likely due to the images being simultaneously nested and crossed within conditions – i.e., only one kind of stimulus (e.g., manmade items) was presented in each baseline block, but both kinds of stimuli were presented in the filtering block. Thus, the final model for the BWvGS, CvGS, and control experiments were: RT ~ Block + (1 + Block | participant) + (1 | item).

The results of these analyses are reported in Table 1 and displayed in Figure 3. Figure 3 highlights that there were no differences between baseline blocks nor evidence of Garner interference in the BWvGS (Figure 3, panels A & B) or CvGS (Figure 3, panels C & D) experiments. Secondly, in the control experiment, Garner interference was observed, where RTs were slower in the filtering block compared to the baseline ones (Figure 3, panel E).

18

Table 1. *LMM's examining differences between baseline and filtering blocks for reaction time data*

Experiment	Task	Fixed Effect Tests	*Post Hoc* Comparison	
BWvGS	Image Format	$F_{(2, 20.01)} = 0.52$, $p = .600$	-	
	Semantic Classification	$F_{(2, 17.63)} = 0.16$, $p = .858$	-	
CvGS	Image Format	$F_{(2, 18.54)} = 1.15$, $p = .337$	-	
	Semantic Classification	$F_{(2, 17.23)} = 0.94$, $p = .410$	-	
Control (HvW)	-	$F_{(2, 2460.30)} = 279.50$, $p < .001$*	BaselineShort – BaselineTall	*z-ratio* = -1.63, $p = .327$
			BaselineShort - Filtering	*z-ratio* = -17.89, $p < .001$*
			BaselineTall - Filtering	*z-ratio* = -15.71, $p < .001$*

Asterisks (*) denote significant differences at the .016 level according to a Bonferroni adjustment

19

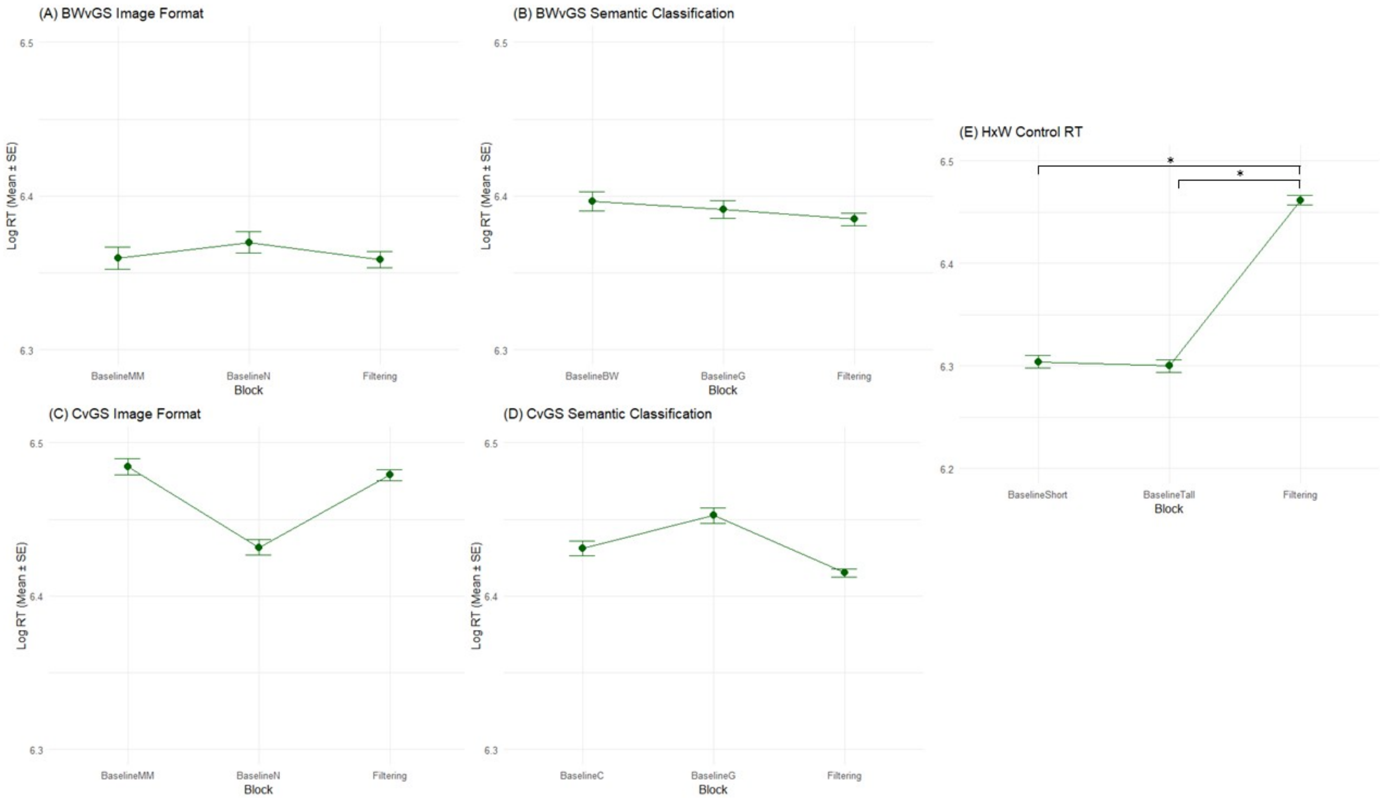

Figure 3. Line graphs depicting the Log RT mean ± SE for the stage 1 analyses. Panels A, B, C, D, and E depict the baseline1, baseline2, and filtering analyses for the BWvsGS, CvGS, and control experiments. As can be visualised, there were no significant differences between the baseline blocks, or between the baseline blocks and filtering for the BWvGS and CvGS experiment. Panel E depicts the typical Garner interference effect with rectangles, where there is no difference between

20

baseline blocks, and longer RT in the filtering condition. Asterisks (*) denote significant differences at the .016 level according to a Bonferroni adjustment.

21

Stage 2: Having verified that the RT in baseline blocks did not differ, we then combined them into a single variable for each task. For experiments 1 and 2, a LMM with Task (SC vs. IF) and Block (Baseline vs. Filtering) and their interaction as fixed factors, and participant and item as random factors was performed. For the BWvGS experiment, the model failed to converge when the intercepts and slopes for the random factors, including their interaction, were included. This was due to the high degree of correlation between the intercept for item and the slopes for Block (r > .99) and Task (r = -.82). A model with only the intercept for item was able to converge and thus was used as the maximal model. The same was true for the CvGS experiment. Thus, the final model was: RT ~ Block * Task + (1 + Block * Task | participant) + (1 | item).

No main effect of Block ($F_{(1, 16.97)} = 0.39$, $p = .543$) or a Block × Task interaction ($F_{(1, 17.00)} = 0.24$, $p = .629$) was observed with RT data in the BWvGS experiment (see *Figure 4*, panel A). However, the main effect of Task was approaching significance ($p = .060$), where participants were marginally faster in the IF compared to SC task. Thus, there was no evidence of Garner interference (Filtering >/ Baseline) in either the IF (Baseline $M = 6.37$, $SE = 0.05$, Filtering $M = 6.37$, $SE = 0.05$) or SC tasks (Baseline $M = 6.41$, $SE = 0.04$, Filtering $M = 6.40$, $SE = 0.05$).

Similarly, there was no main effect of Block ($F_{(1, 16.98)} = 0.88$, p =.418), Task ($p = .798$) or a Block × Task interaction ($F_{(1, 16.67)} = 2.11$, p =.174) in the CvGS experiment (see *Figure 4*, panel B). Again, there was no evidence of Garner interference (Filtering >/ Baseline) in the IF (Baseline

22

$M = 6.46$, $SE = 0.03$, Filtering $M = 6.47$, $SE = 0.03$) or SC tasks (Baseline $M = 6.45$, $SE = 0.03$, Filtering $M = 6.42$, $SE = 0.03$).

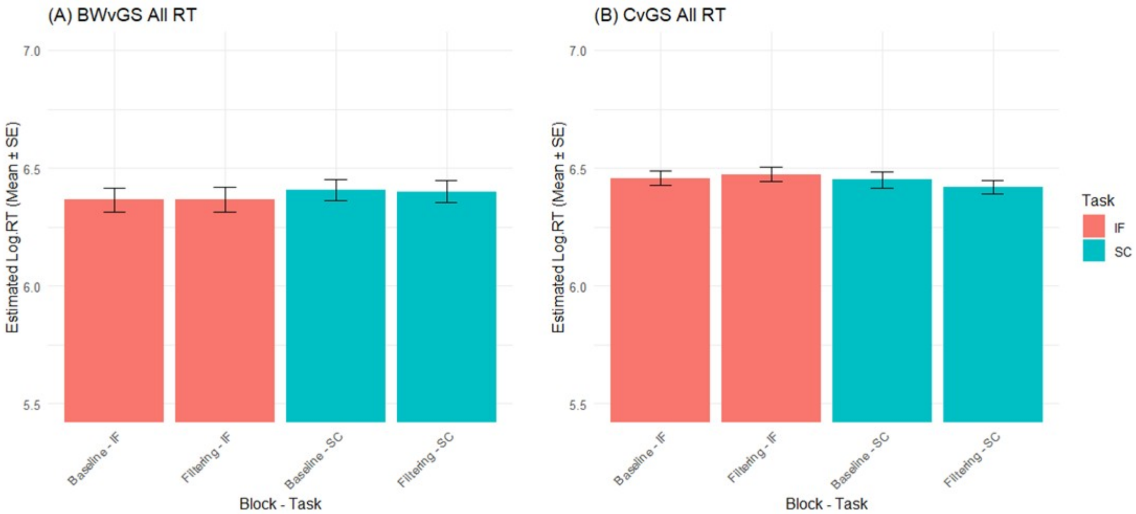

Figure 4. Bar graphs depicting the Log RT mean ± SE for the stage 2 analyses. No differences in RT were detected for the BWvGS (panel A) or the CvGS (panel B) experiments.

23

Accuracy

Stage 1: GLMMs were calculated with Block (Baseline1 vs. Baseline2 vs. Filtering) as a fixed factor and participant and item as random factors. However, the inclusion of the item and participant slopes in all models resulted in an almost singular fit, where the model failed to converge as the slopes for Block and participant were highly correlated with the intercepts for item and participant ($r > .99$). Therefore, the random slopes for both random factors were excluded, making this an intercept only GLMM. The final model was: Accuracy ~ Block + (1 | participant) + (1 | item).

The results of these analyses are reported in Table 2 and displayed in Figure 5. As depicted in Figure 5, there were no differences between the baseline and filtering blocks (Figure 5, panels A and C) in either IF tasks, as well as the BWvGS SC task (Figure 5, panel B). However, accuracy was lower in the filtering condition compared to the baseline greyscale in the CvGS SC task (*z.ratio* = 2.50, *p* = .037), although this did not survive correction for multiple comparisons (Figure 5, panel D). Finally, Garner interference was again observed (Figure 5, panel F) in the control experiment (see Table 2).

25

Table 2. *GLMM examining differences between baseline and filtering blocks for accuracy data*

Experiment	Task	Fixed Effect Tests	*Post Hoc* Comparison	
BWvGS	Image Format	$\chi 2_{(2)} = 2.21$, $p = .331$	-	
	Semantic Classification	$\chi 2_{(2)} = 3.16$, $p = .206$	-	
CvGS	Image Format	$\chi 2_{(2)} = 3.16$, $p = .206$	-	
	Semantic Classification	$\chi 2_{(2)} = 6.83$, $p = .033$*	BaselineC – BaselineGS	*z.ratio* = -1.16, $p = .740$
			BaselineC – Filtering	*z.ratio* = 1.00, $p = .951$
			BaselineGS – Filtering	*z.ratio* = 2.50, $p = .037$
Control (HvW)	-	$\chi 2_{(1)} = 135.73$, $p < .001$**	BaselineShort – BaselineTall	*z.ratio* = -0.24, $p = .968$
			BaselineShort – Filtering	*z.ratio* = 7.98, $p < .001$**
			BaselineTall – Filtering	*z.ratio* = 8.56, $p < .001$**

Note. Results are given on the log odds ratio (not the response) scale
Single asterisks (*) denote significant differences at the .05 level
Double asterisks (**) denote significant differences at the .016 level according to a Bonferroni adjustment

26

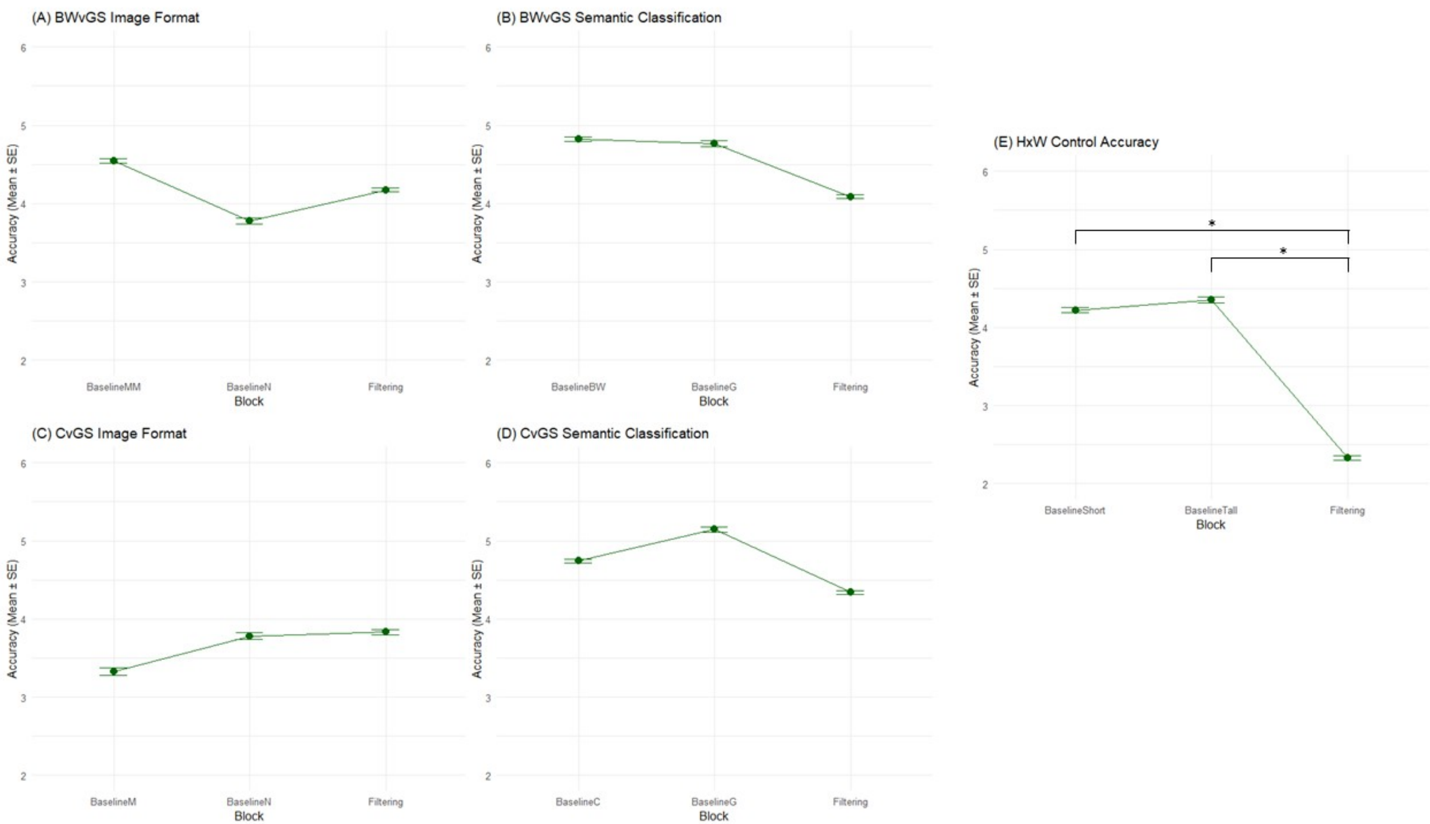

Figure 5. Line graphs depicting the logit scale accuracy mean ± SE for the stage 1 analyses. Panels A, B, C, D, and E depict the baseline1, baseline2, and filtering analyses for the BWvsGS, CvGS, and control experiments. As can be visualised, there were no significant differences between the baseline blocks, or between the baseline blocks and filtering for the BWvGS and CvGS experiment. Panel E depicts the control experiment, where they typical Garner interference effect can be seen with no

27

difference between baseline blocks, and lower accuracy in the filtering condition. Asterisks (*) denote significance at the .016 level according to a Bonferroni adjustment.

28

Stage 2: Having verified that baseline blocks did not differ, a GLMM was calculated with Block (Baseline Combined vs. Filtering), Task (SC vs. IF), and their interaction as fixed factors, and participant and item as random factors. The maximal model including all random factor slopes (i.e., Block, Task, and Block × Task) failed to converge and was singular. This was due to near perfect correlations between the slopes for Block and Task with the intercept for item (r = .98 and -.84, respectively). With these excluded, the model still failed to converge, due to extremely low correlations between the slopes for Block and Task and the intercept for participant (r = .06 and -.09, respectively). Therefore, an intercept only model was used for both BWvGS and CvGS experiments. The final model was: Accuracy ~ Block * Task + (1 | participant) + (1 | item).

For the BWvGS experiment (panel A, *Figure 6*), Garner interference was observed (filtering < baseline) in the SC task (*z.ratio* = 3.17, p =.003) and there was a Block × Task interaction ($\chi2_{(1)}$ = 7.19, p = .007). This was driven by higher accuracy in the SC baseline relative to the other three blocks. However, no Garner interference was detected in the IF task (*z.ratio* = -0.37, p = .778).

A Block × Task interaction was also observed in the CvGS experiment ($\chi2_{(1)}$ = 9.09, p = .002; panel B, *Figure 6*). Like in the BWvGS experiment, this was driven by Garner interference in the SC task (*z.ratio* = 2.46, p = .017) but not in the IF task (*z.ratio* = -1.77, p = .078). The interaction effect also seems to be driven by differences in accuracy between the tasks, where accuracy was lower in the IF compared to SC

29

task – IF baseline accuracy was lower than SC baseline and filtering, and IF filtering was lower than SC baseline and filtering (all $p < .001$).

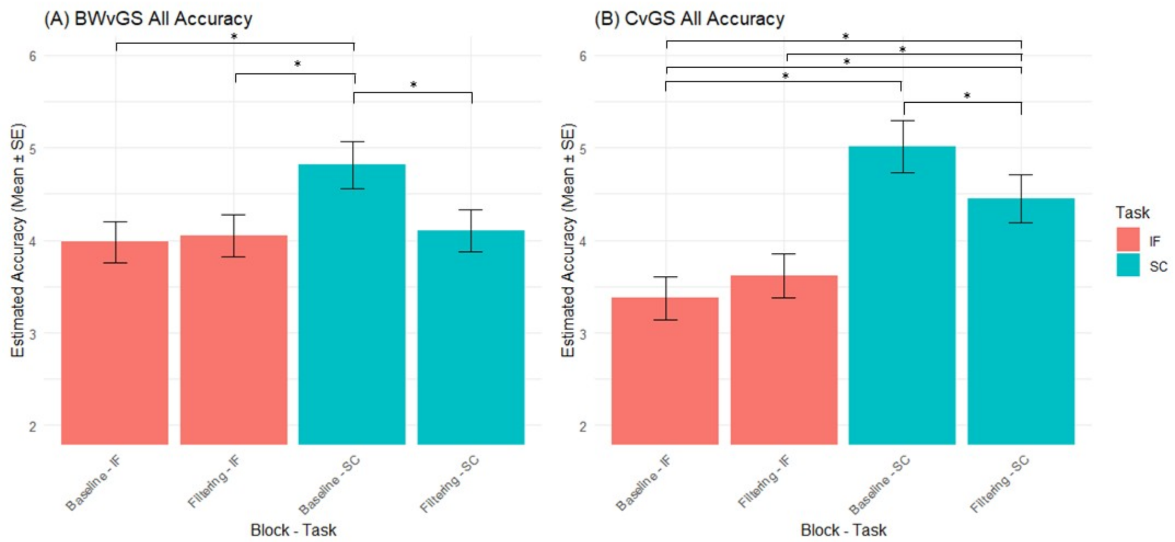

Figure 6. Bar graphs depicting the logit scale accuracy mean ± SE for the stage 2 analyses. Panels A and B depict the significant Block × Task interaction for the both the BWvGS and CvGS experiments, where Garner interference was detected in the SC, but not IF tasks. Asterisks (*) denote significance at the < .05 level.

Discussion

We investigated whether participants could selectively attend to image format while ignoring semantic category, and vice versa, by examining Garner interference. Garner interference was replicated in a control experiment examining selective attention to the width while ignoring the height of rectangles. In this task, participants made more errors and took

30

longer in the filtering compared to baseline blocks – confirming that the typical Garner interference effect can be observed with our general procedures. In both the BWvGS and CvGS experiments, no significant differences were observed in the RT data, but there was significant Garner interference with the accuracy data in the SC but not IF task. The most likely explanation for finding the same effects in both tasks is that the semantic classification of drawings is not independent of the image format, but, classification of image format can occur independently of the images' semantic category. The theoretical implications of these results are discussed.

Garner interference with accuracy but not RT data has been reported previously, although it is uncommon (Freud & Ganel, 2015). Freud and Ganel (2015) reported the effect when participants were completing a task that required the engagement of the dorsal and ventral visual streams for action and perception, respectively, in an interactive manner, as opposed to a typical perceptual task engaging the ventral visual stream, like the present study. It is unlikely that speed-accuracy trade-offs could account for us obtaining Garner interference with the accuracy data only, as averaging accuracy and RT scores across blocks and tasks for each participant and then correlating them (Wickelgren, 1977) revealed no association between the two measures (all $p > .348$). It is unclear why Garner interference was only observed with the accuracy data.

31

Nevertheless, Garner interference with accuracy demonstrates that classifying an image’s semantic identity is not separable from its image format. In both the BWvGS and CvGS experiments, our theoretical rationale for testing if selective attention to perceptual and conceptual aspects of the images was possible came from observations in form agnosia patients. People with form agnosia show the greatest impairment with processing simple, black-and-white line drawings, but show some improvement when identifying images that have more pictorial cues like the greyscale images used here (Farah, 2004; Peel & Chouinard, 2023). Our results support the notion that the difference in format between black-and-white and greyscale images is important when semantically classifying them. In neurotypical populations, previous naming experiments have demonstrated that the identification of black-and-white line drawings is slower than naming greyscale drawings (Bonin et al., 2019). However, we found no differences (both in RT and accuracy) between the baseline tasks when participants semantically classified items from only one format. This may relate to task differences. We had participants classify objects into semantic categories while Bonin et al. (2019) had participants name them. If this explanation is correct, then the processes involved in naming compared to semantically labelling items could be distinct. This is an interesting point – it is tempting to assume that one needs to identify a stimulus to semantically label it, but it appears that processing an image at the semantic level might be more automatic. Previous studies demonstrate that pictures are processed according to their semantic identity faster than their names (Carr et al., 1982; Catling &

32

Johnson, 2006). Further research is required to differentiate semantic processing from naming, but this interpretation nevertheless fits with anecdotal reports from participants in our study who reported that they were not identifying the stimulus according to its name before making their judgement (i.e., the sequence did not follow = that's a banana → that's a natural item → button press).

There is also evidence that colour perception is largely normal in visual form agnosia patients (Peel & Chouinard, 2023), and, given that colour can be diagnostic and facilitate recognition (Chouinard & Goodale, 2012; Bramão et al., 2011), we wanted to examine whether Garner interference could be observed with structurally identical objects that differed only by colour. Indeed, interference was found in the CvGS experiment in the SC task, suggesting that image format colour is integral with semantic classification. Some research has reported that colour and form are processed independently. In an electroencephalography study, Proverbio et al. (2004) report different underlying mechanisms are at play when colour and form are being recognised in everyday objects, where there is a primacy of global shape over colour in object recognition. Likewise, in other Garner interference studies, Cant et al. (2008; 2009) demonstrated that shape and colour are processed independently. The present result builds from this earlier work by showing that colour and form may be independent, while colour and semantic classification are not. Why might this be the case?

33

For both experiments, the observed Garner interference for the SC task but not IF suggests asymmetric interference. In other Garner interference studies, asymmetric interference has been observed with emotion and identity judgements with face perception (Schweinberger et al., 1999), and two broad explanations for asymmetric Garner interference have been put forth by Atkinson et al. (2005). The first relates to the differences in processing speed of the two types of information that are under investigation where faster tasks tend to interfere with slower ones, but not vice versa. There is trending evidence that this may be the case in the BWvGS experiment, as the difference in RT between tasks was approaching significance, but the same cannot be said in the CvGS experiment. The second explanation for asymmetric interference relates to a reliance of one process over another that is unidirectional as opposed to bidirectional. We believe this best explains the effect observed in both experiments. Algom and Fitousi (2016) describe this effect as follows – given a pair of dimensions, A and B, A can be separable with respect to B, but B is integral with respect to A. Moreover, if A exists then B also exists, but if B exists A may or may not exist. For example, a phoneme must have a pitch, but a pitch can exist without any linguistic property like a phoneme. In our case, it would be the processing of featural information facilitating judgements about semantic identity but not the other way around. That is, a drawing can have a semantic identity, but it must simultaneously possess an image format to exist (if A exists then B also exists) whereas an image format does not have to possess a semantic identity (if B exists A may or may not exist).

34

This observed relationship between conceptual and perceptual processes aligns with neuroimaging results. Dirani and Pylkkanen (2024) found magnetoencephalography evidence that semantic representations are not strictly amodal but contain visual components. Davis and colleagues (2020) also demonstrated with fMRI that visual representations predict perceptual memory in visual areas, but also facilitate conceptual and general memory in more anterior regions, highlighting the importance of visual properties in influencing conceptual judgements. Other fMRI research from Victoria et al., (2019) showed that perceptual features help organize mental categories throughout the object processing hierarchy, demonstrating that visual similarity impacted adaptation in nonvisual brain regions like the dorsomedial prefrontal cortex. The neural correlates of Garner interference have not been investigated to the authors' knowledge, but whole brain analyses have shown that fronto-parietal attention networks are key to selective visual attention (Salo et al., 2017). Future research could seek to elucidate the relative importance of early visual areas to each task investigated here to clarify our interpretations of early visual processing impacting later processing, but not vice versa.

Methodological considerations

Compared to the typical Garner effect that assesses the height and width of objects, the stimuli in our experiment were more complex and differed by a host of other factors including configuration, size in reality, and so on. This potential limitation means there may be other factors, beyond the natural

35

vs. manmade, or colour vs. greyscale distinction that influenced processing speed and accuracy owing to target image variability. To statistically account for this possibility, we included the random effects associated with item in our analyses. Indeed, we found the same pattern of effects in both the BWvGS and CvGS analyses experiments, which supports the reliability of the asymmetric interference between conceptual and perceptual processing.

Future research could examine more precisely what kinds of featural information might be producing Garner interference in the semantic classification task. One possibility is line curvature. Natural objects comprise more *organic* irregular lines and manmade items contain more *inorganic* straight lines (Arnheim, 1974). In our stimulus set, all natural objects had irregular contours while our manmade objects contained a mixture of straight and irregular lines (e.g., wheel, television, and whistle). These perceptual differences between categories might normally facilitate their semantic classification and influence Garner interference.

It is also apparent from our analyses that there may have been differences between the tasks in terms of difficulty. In the BWvGS experiment, there was a trend for participants being faster to classify objects on the IF compared to SC task, but no differences in terms of accuracy. This effect did not reach statistical significance, but it is suggestive of more rapid classification of items as being black-and-white or greyscale, compared to judging if they are manmade or natural. In the CvGS experiment, the opposite was true, where participants had higher

36

accuracy for the SC task but no differences with RT. This suggests that semantic classification was easier than classifying stimuli as coloured or greyscale. Although stimuli from the ‘colour’ block had colour and were qualitatively different from achromatic grey, it is conceivably more difficult to classify images of elephants and metallic pliers, which are ordinarily grey, as being greyscale or coloured compared to images of strawberries and tennis balls, which are ordinarily red and yellow, respectively.

Indeed, the data suggests some variability in the participants’ ability to classify images as either grayscale or colour. Half the participants performed with accuracy levels of over 95% while the other half had more difficulty (range: 89 – 94%), which may relate to differences in basic visual functions (e.g., contrast sensitivity). Including participant and item as random factors in our analyses makes us more confident that the difference in difficulty between the IF and SC tasks is real and not the result of confound in our choice of stimuli or driven by a sampling bias. Conceptually discriminating images was perhaps a more familiar task than perceptually labelling them as being in one format versus another. The former is conceivably done more frequently in daily life – rarely are we asked, or required, to classify images based on their format. In any case, this observed difference between the tasks, while interesting, was not central to our research question assessing Garner interference, and does not affect our conclusions that semantic classification is not separable from image format.

Another methodological consideration is our choice of stimuli. We used the Snodgrass and Vanderwart drawings given they have been extensively used in research, and as such, have many norms associated with them that made it possible to match them according to various criteria as we did here. We did not use photographs or other more salient image formats because (1) we were specifically interested in how we process drawings, and (2), there are existing norms that made them possible to match into manmade / natural categories. Although photographs and more salient images contain more features that people use to recognise objects in the real world, they are also two dimensional like our stimuli. In contrast, real-world objects are three-dimensional. So, further investigation of how this difference might translate to real-world settings with solid objects is required.

One suggestion for future research would be to perform similar Garner interference experiments in virtual reality (VR) using more salient stimuli than those we used in this study. VR affords the researcher ample control of how visual stimuli are presented, with potentially the same high levels of internal validity as in experiments using flatscreen monitors, but better emulates real world viewing by presenting images with stereoscopic cues to give the impression of a third dimension. Our results point to the importance of structural form cues in guiding conceptual judgements so it would be informative to better understand how changes to the metric features of stimuli in a three-dimensional environment might affect such processes. For example, viewpoint rotations or manipulations of stimulus size could be one avenue for future research to pursue using VR.

38

Conclusions

In closing, we demonstrate that image format and semantic classification are not independent, although the relationship is asymmetric. In both the BWvGS and CvGS experiments, participants could not selectively attend to semantic identity when image format varied, highlighting that random variation in image format cues interfered with the ability to conceptually classify images. However, the same was not true when the task required perceptual classification of image format when semantic identity varied from trial to trial, highlighting that image format could be processed independently of its conceptual identity.

Acknowledgements

We thank Professor Pauleen Bennett for her feedback on the manuscript and Joshua Egeland for his assistance with data collection. This work was financially supported by La Trobe University's School of Psychology and Public Health.

Declarations

Open Practices Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. Materials for the experiments reported here is available here from Michael Tarr's stimulus set

39

(https://sites.google.com/andrew.cmu.edu/tarrlab/stimuli). None of the experiments was preregistered.

- Funding – this research received in kind support from La Trobe University
- Conflicts of interest/Competing interests – the authors report no conflict of interest
- Ethics approval - The study was approved by the La Trobe University Human Ethics Committee HEC20104
- Consent to participate – Written informed consent was obtained from the participants
- Consent for publication – Participants provided signed informed consent to publish their data
- Availability of data and materials – The data will be made available upon reasonable request, and the materials are available at Michael Tarr’s stimulus set repository (https://sites.google.com/andrew.cmu.edu/tarrlab/stimuli)
- Code availability – The code is available in text (see Results section)

References

Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability-integrality distinction. *Psychological Bulletin*, *142*(12), 1352-1383. https://doi.org/10.1037/bul0000072

Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye.* University of California Press.

Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception & Psychophysics*, *67*(7), 1199-1213. https://doi.org/10.3758/BF03193553

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Benton, A. L. (1945). A visual retention test for clinical use. *Archives of Neurology and Psychiatry, 54,* 212–216.

Bonin, P., Méot, A., Laroche, B., Bugaiska, A., & Perret, C. (2019). The impact of image characteristics on written naming in adults. *Reading and Writing*, *32*(1), 13-31. https://doi.org/10.1007/s11145-017-9727-3

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, *138*(1), 244-253. https://doi.org/https://doi.org/10.1016/j.actpsy.2011.06.010

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351.

41

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage*, *155*, 437-449. https://doi.org/10.1016/j.neuroimage.2017.03.06

Burns, D. M. (2014). The many faces of Garner Interference. *Published Thesis*.

Cant, J. S., Large, M.-E., McCall, L., & Goodale, M. A. (2008). Independent Processing of Form, Colour, and Texture in Object Perception. *Perception*, *37*(1), 57-78. https://doi.org/10.1068/p5727

Cant, J. S., & Goodale, M. A. (2009). Asymmetric interference between the perception of shape and the perception of surface properties. *Journal of Vision*, *9*(5), 13-13. https://doi.org/10.1167/9.5.13

Chouinard, P. A., Morrissey, B. F., Köhler, S., & Goodale, M. A. (2008). Repetition suppression in occipital–temporal visual areas is modulated by physical rather than semantic features of objects. Neuroimage, 41(1), 130-144.

Chouinard, P. A., & Goodale, M. A. (2012). FMRI-adaptation to highly-rendered color photographs of animals and manipulable artifacts during a classification task. Neuroimage, 59(3), 2941-2951.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763-770. https://doi.org/10.1038/nn.3381

Davis, S. W., Geib, B. R., Wing, E. A., Wang, W.-C., Hovhannisyan, M., Monge, Z. A., & Cabeza, R. (2020). Visual and Semantic Representations Predict Subsequent Memory in Perceptual and Conceptual Memory Tests. *Cerebral Cortex*, *31*(2), 974-992. https://doi.org/10.1093/cercor/bhaa269

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects. *The Journal of Neuroscience*, *33*(48), 18906-18916. https://doi.org/10.1523/jneurosci.3809-13.2013

42

Dirani J, Pylkkänen L. MEG Evidence That Modality-Independent Conceptual Representations Contain Semantic and Visual Features. J Neurosci. 2024 Jul 3;44(27):e0326242024. doi: 10.1523/JNEUROSCI.0326-24.2024. PMID: 38806251; PMCID: PMC11223456.

Freud, E., Avidan, G., & Ganel, T. (2013). Holistic processing of impossible objects: Evidence from Garner's speeded-classification task. *Vision Research*, *93*, 10-18. https://doi.org/https://doi.org/10.1016/j.visres.2013.10.001

Freud, E., & Ganel, T. (2015). Visual control of action directed toward two-dimensional objects relies on holistic processing of object shape. *Psychonomic Bulletin & Review*, *22*(5), 1377-1382.

Garner, W. (1970). The stimulus in information processing. *American Psychologist*, *25*, 350-358.

Henshilwood, C. S., d'Errico, F., van Niekerk, K. L., Dayet, L., Queffelec, A., & Pollarolo, L. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature*, *562*(7725), 115-118. https://doi.org/10.1038/s41586-018-0514-3

Hooper, H. E. (1983). Hooper Visual Organization Test Manual. Los Angeles, CA, Western Psychological Services.

Kaplan, E., Goodglass, H., & Weintrab, S. (1983). The Boston naming test. Philadelphia: Lea & Febiger.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends in cognitive sciences, 17(1), 26-49.

Landis, T., Graves, R., Benson, F., & Hebben, N. (1982). Visual recognition through kinaesthetic mediation. *Psychological Medicine, 12*(03). doi:10.1017/s0033291700055616

43

Lu, X., Suryanarayan, P., Adams, R. B., Jr., Li, J., Newman, M. G., & Wang, J. Z. (2012). On Shape and the Computability of Emotions. Proc ACM Int Conf Multimed, 2012, 229-238. doi:10.1145/2393347.2393384

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L. Y., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife*, *7*, e31873. https://doi.org/10.7554/eLife.31873

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language*, *94*, 305-315.

Mendoza, J.E. (2011). *Associative Visual Agnosia.* In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_708

Miller J. Reaction time analysis with outlier exclusion: bias varies with sample size. Q J Exp Psychol A. 1991 Nov;43(4):907-12. doi: 10.1080/14640749108400962. PMID: 1775668.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*(4), 695-699. doi: 10.1111/j.1532-5415.2005.53221.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Peel, H. J., & Chouinard, P. A. (2023). A review of the impairments, preserved visual functions, and neuropathology in 21 patients with visual form agnosia–A unique defect with line drawings. *Neuropsychologia,* 108666.

44

Proverbio, A. M., Burco, F., del Zotto, M., & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research*, *18*(3), 288-300.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217-236. https://doi.org/10.1068/p5117

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* URL https://www.R-project.org/.

Salo E, Salmela V, Salmi J, Numminen J, Alho K. Brain activity associated with selective attention, divided attention and distraction. Brain Res. 2017 Jun 1; 1664:25-36. doi: 10.1016/j.brainres.2017.03.021. Epub 2017 Mar 28. PMID: 28363436.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, *61*(6), 1102-1115.

Snellen, H. (1862). *Letterproeven, tot bepaling der gezigtsscherpte* (Vol. 1). J. Greven.

Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174-215.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Victoria, L. W., Pyles, J. A., & Tarr, M. J. (2019). The relative contributions of visual and semantic information in the neural representation of object

45

categories. *Brain and Behavior*, *9*(10), e01373. https://doi.org/https://doi.org/10.1002/brb3.1373

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, *41*(1), 67-85.

Zimmerman, D. W., & Williams, R. H. (2000). Restriction of Range and Correlation in Outlier-Prone Distributions. Applied Psychological Measurement, 24(3), 267–280. https://doi.org/10.1177/01466210022031741

46

Appendix

Natural	Manmade
Umbrella	Tree
Table	Carrot
Windmill	Kangaroo
House	Butterfly
Basket	Pineapple
TV	Chicken
Shoe	Cat
Tie	Mushroom
Key	Flower
Bottle	Banana
Guitar	Bear
Pram	Monkey
Suitcase	Swan
Whistle	Strawberry
Canon	Leopard
Traffic light	Camel
Wheel	Caterpillar
Violin	Ostrich
Nail	Potato
Cigar	Pepper
Kettle	Duck
Wrench	Onion

47

Harp	Grasshopper
Roller-skate	Bee
Sled	Fly
Pliers	Asparagus

890