SVDW Recognition Pilot: Preregistration

Kaitlyn Fallow (kmfallow@uvic.ca)

January 5, 2019

Contents

Background

Snodgrass and Vanderwart (1980)'s well-known stimulus set has been used in a number of recognition memory studies, perhaps most often in those exploring picture superiority effects. Although these studies have varied in their choice of dependent measure, with authors variously using hit rates, Br, the signal detection measure d', and so on to represent memory performance, response bias has been paid almost no direct attention.

Because much work in our lab has focused on materials-based differences in recognition memory response bias, and even more specifically on differences between stimulus sets comprising words and complex, visually rich images, the Snodgrass and Vanderwart (SVDW) set has long been of interest to us, being perhaps the most commonly used set of words and images in memory research. We are not aware of any consistent materials-based response bias differences having emerged from studies using these stimuli, although given the relative infrequency with which bias measures have historically been reported in this literature and the fact that other variables are often of primary interest and/or manipulated it is difficult to get a clear picture. Feenan & Snodgrass (1990), for example, reported a more conservative bias for SVDW words than for pictures, and C was numerically conservative across the board for single stimuli, but their primary interest was contextual change, which they accomplished by presenting some stimuli in pairs that were either maintained or rearranged at test. Healthy participants in Snodgrass & Corwin (1988)'s Exp. 2 showed a slightly conservative bias to SVDW pictures in the first study/test cycle that diminished to near-neutral in a second cycle. Karlsen & Snodgrass (2004) also reported roughly neutral c values for SVDW pictures, but in this case were interested in the role of familiarity and therefore restricted their image set to those at the extreme ends based on the familiarity-rating norms.

Response bias for the SVDW stimuli, and whether it differs between words and pictures, no doubt depends to some extent on characteristics of the experiment and sample, but whether or not materials-based differences are observed with these stimuli under "typical" recognition memory test conditions may tell us something about the boundary conditions of the effects we have observed. We have thus far found response bias to be significantly conservative for pictures, and significantly more conservative than that for words (which is typically neutral in words-only studies, and liberal in those intermixing words and pictures) in studies using paintings, faces, and photos of assorted scenes. Although these image types differ in a number of ways, all are visually complex and diverse; with the possible exceptions of faces and painted portraits, most of the images do not depict single objects and would be difficult to label succinctly. Also unlike the SVDW pictures, the images we have used do not map onto the word stimuli (3-8 letter English nouns of moderate-to-high frequency) we have used in our research, such that there are innumerable materials-based differences to which the response bias pattern could conceivably be attributed. Because the SVDW words and pictures refer to the same objects and the sets have been normed with respect to a number of variables thought to be important to memory (e.g., familiarity, visual complexity), the number of materials-based differences is more constrained. The presence or absence of a materials-based difference in recognition memory response

bias for these stimuli, then, may serve as a useful baseline in attempting to understand what drives the differences we have observed.

Methods

Participants

Participants will be UVic undergraduates who take part in research for bonus course credit. We will start by pilot testing 20-25 participants one at a time and regularly checking the data to get a sense of typical hit and false alarm rates for these stimuli under our standard procedure; the possible range, and therefore informativeness, of c and d' (and related statistics) decreases dramatically as these rates approach 0 or 1. If a lot of participants are making few or no mistakes (or – much less likely – few or no correct responses) of either type, procedural adjustments may be necessary to bring performance within an acceptable range.

Materials

Stimuli are 200 line drawings and corresponding labels from Snodgrass and Vanderwart (1980). These were drawn from the original set on the basis of a stimulus-piloting study (analyzed $N = 59$) wherein participants viewed a random selection of 208 of the 260 images and were given a maximum of 5 seconds to enter the first name that came to mind for each. This was done mainly for the purpose of excluding images depicting objects likely to be unfamiliar to many current students; exact name agreement was not of particular concern. We set an initial threshold of 90% approximate agreement across participants. 39 images met this threshold based on exact matches to the original name, and responses to all remaining images were reviewed by 3 raters with instructions to accept misspellings, incomplete responses that appeared likely to be correct, and approximate synonyms (e.g., "bunny" for "rabbit") as matches. Items for which the maximum, but not minimum, # of rater-determined matches would produce a total above the threshold were resolved by the researcher. The most common "problem" seemed to be over-general naming (e.g., "bird" for "eagle", "swan", etc.), likely reflecting imprecision in the original instructions, so such responses were generously interpreted. Images for which the corresponding label comprised 2 words were also removed for the sake of consistency. As the final number at the end of this process fell just short of the minimum desired total (192 + 6 buffers), the threshold was lowered to 85% yielding a total of 194, and the next highest 6 (minimum 84%) were assigned to be used as buffers.

Study and test lists will be randomized anew for each participant such that individual items will vary across individuals with respect to study status, study/test position, and whether they are presented as pictures or words. Exactly 50% of both studied and new items will be pictures, and the same item will never appear as both a picture and a word for a given participant. Similarly, items studied as pictures will never be tested as words or vice versa.

Procedure

Participants will begin by studying 96 items (50% pictures) bookended by 6 additional primacy/recency buffers (3 of each) for 1 s each with a 1-s ISI (including a 500-ms central fixation cross). Images will be presented in the centre of the screen at their original sizes (281 x 197 or 197 x 281 pixels) and words will be presented centrally in 48-pt Arial font. Participants will be informed at the beginning of the study phase that they are going to see a series of words and pictures representing an assortment of living and non-living things; that while they may see categorically related words and pictures, the same item will not be presented in both formats; and that their task is to try their best to remember each item. Following the study phase there is a short filler task wherein participants are asked to judge whether a series of numbers and letters, some of which have been rotated clockwise or counterclockwise to varying degrees, have been horizontally flipped or not. This task is designed to last approximately 5 min including time to read the instructions, with the actual task set to display as many trials as a given participant can complete in 3.5 min.

The recognition test comprises all studied items and an equal number of new items. Participants will be told they are going to see another series of words and pictures, some of which will be from the study list and some of which will be new, and that they will be asked to make studied/not studied judgments on a 6-pt confidence weighted scale (from 1 = definitely not studied to 6 = definitely studied). This scale will remain onscreen throughout the test, which participants will complete at their own pace. At the end of the test participants will be asked what % of words and pictures no the test they think were from the study list. The researcher will then ask participants some additional informal questions about their experience and for any comments they wish to share.

Hypotheses

We do not plan to test any hypotheses with data from this pilot sample, but the primary hypotheses we hope to test eventually are that, unlike the words and pictures we have used previously, the SVDW stimuli will *not* produce a significant materials-based difference in recognition memory response bias, nor will the bias for pictures be significantly conservative.

Planned analyses

Our primary dependent measures of interest will be hit and false alarm rates (determined by collapsing our 1-6 response scale into a binary one, with 1-3 = "new" and 4-6 = "old"), and corresponding signal detection theory (SDT)-based measures of response bias (c) and sensitivity (d'). Ceiling and floor rates will be replaced according to Macmillan and Kaplan (1985) to enable calculation of c and d'. At the pilot stage we do not plan to conduct any inferential analyses (or if they are conducted it will be purely exploratory), but will instead just look at means/distributions as data come in to get a sense of whether ceiling/floor rates and/or very poor performance (our usual exclusion criterion for d' is < 0.2) are frequent enough to merit concern.