The Availability Heuristic in a Symbolic-Connectionist Architecture

Aaron J. Hamer (ahamer@hawaii.edu)
University of Hawaii at Manoa, Department of Linguistics
1890 East-West Road, Honolulu, HI 96822, USA

Leonidas A. A. Doumas (alex.doumas@ed.ac.uk)
University of Edinburgh, Department of Psychology
3 Charles Street, Edinburgh EH8 9AD, UK

Abstract

Theories of how cognitive biases arise rely on heuristics that influence attention and reasoning. These heuristics serve as a post-hoc description of how attention and reasoning is weighted to produce the patterns of deviation in judgment. We are unaware of any process account of how these heuristics arise. In this article, we present several simulations of classic studies of the availability heuristic and describe how the availability heuristic arises through distributed symbolic representations and simple process interactions in an existing model of relational reasoning and concept development (*Discovery of Relations by Analogy*; Doumas, Hummel, & Sandhofer, 2008).

Keywords: availability heuristic, retrievability biases, symbolic-connectionism

Cognitive Biases and Heuristics

Tversky & Kahneman's (1974) seminal paper on cognitive biases (patterns of deviation in judgment) identified several heuristics (rules that constrain which hypotheses are entertained) that limit how people reason in uncertain situations. These cognitive biases have captured the imagination of scientists and laypeople alike, resulting in a massive body of research and a Nobel prize for Kahneman. Three major heuristics were identified in the Tversky & Kahneman (1974) paper: representativeness (the degree to which an event is prototypical of its class or the process that generates it), availability (the ease of recall of instances or properties), and adjustment and anchoring (early experiences serve as an anchor and adjustments due to subsequent experience fall short of the mark).

Despite the deep literature base and interest from fields ranging from behavioral economics to military science, the development and processing of heuristics held to account for cognitive biases remains largely unexplored. This gap in the literature seems strange given how much ink has been spilled on how such heuristics might operate, the cognitive biases that arise due to the interactions of various heuristics, and even the relative importance of each heuristic involved in a particular pattern of behavior.

In this paper we focus on the availability heuristic - in short, what is easily recalled has a large influence on reasoning, especially around assessments of frequency or probability. This heuristic is intuitively satisfying, as it is likely that instances of large classes (i.e., ones which occur frequently) are recalled more quickly and completely than instances of less frequent classes. Indeed, this phenomena fits well with accounts relying on frequency to mediate access (e.g., Jacoby & Dallas, 1981).

Recent theoretical work on the availability heuristic relies on the dual-process theory of reasoning (for a review, see Evans, 2008). Following Stanovich & West (2000), we shall refer to these processes as *System 1* and *System 2*. System 1 processes are characterized as fast, automatic, and reflexive whereas System 2 processes are characterized as slow, controlled, and reflective (e.g., Schneider & Schiffrin, 1977; Lieberman, Gaunt, Gilbert, & Trope, 2002). An integrated account of these processes has been identified as a key element in how people reason about relations differently than non-human animals (e.g., Hummel & Choplin, 2000; Doumas, Bassok, Guthormson, & Hummel, 2006). To our knowledge, no dual-process account explains how the availability (or any other) heuristic emerges without assuming it a priori as these accounts focus on a description of System 1 and System 2 at a computational level. We simulated several classic studies of the availability heuristic in Doumas, Sandhofer, and Hummel's (2008) *Discovery of Relations by Analogy* (DORA) model, a theory of concept development and relational reasoning which provides an account of how such heuristics might arise as schemas developed through experience.

Methods

In this section we describe two studies from Tversky & Kahneman's (1973) article on the availability heuristic, followed by a brief overview of the DORA model, and present task simulations.

Task Description

Tversky & Kahneman's (1973) article outlines 10 studies designed to investigate the existence of cognitive biases attributed to the availability heuristic. We simulated Studies 5 and 10.

In Study 5, adolescents (N = 118) enrolled in college-preparatory high schools provided estimates of the number of distinct combinations of committees of two to eight members that could be formed from a pool of ten candidates. Although the number of distinct committees of two and eight members are the same (as any committee of eight members drawn from a pool of ten candidates defines a unique unchosen group of two), the availability heuristic suggests that estimates of number of possible committees of two members will be higher than any other size due to the ease of generating and remembering distinct committees of two members as compared to larger committees. Median estimates of possible committees of each size from two to eight are shown in Figure 1 along with the correct values. Estimates of the number of possible committees decreased with committee size.

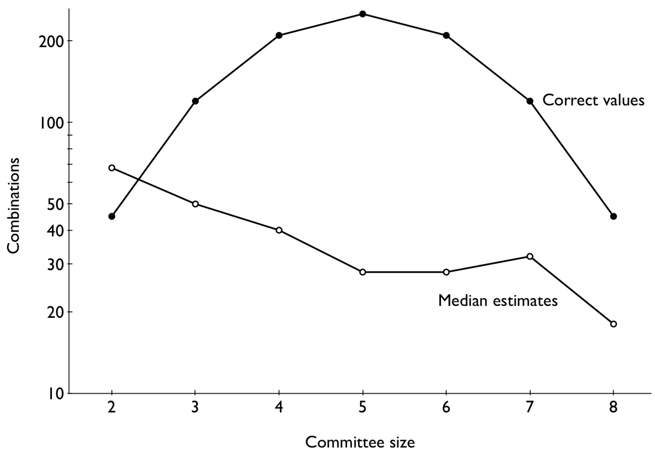

Figure 1. Results from Tversky & Kahneman's (1973) Study 5. Correct values and median estimates of committee size are graphed on a logarithmic scale.

In Study 10, undergraduate students (N = 62) were asked to recall paired personality traits after a training session. The personality trait pairs were standardized and grouped into two sets, a highly-related group (e.g., *kind-honest*) and a unrelated group (e.g., *humble-messy*), in which each of the highly-related pairs were rated as having a higher probability of co-occurrence than any of the unrelated pairs. During training, participants listened to an audio recording consisting of trait pairs occurring one to four times. Equal numbers of highly-related and unrelated trait pairs occurred at each level of frequency. The order of trait pairs was randomized. We focus on the *recall* task, in which participants were given a list of one trait from each pair and asked to recall the corresponding trait. Mean correct recall rates were 41% in the highly-related condition and 19% in the unrelated condition.

The DORA Model

Doumas, Hummel, and Sandhofer's (2008) DORA model of concept development and relational reasoning attempts to explain how structured representations of concepts arise from unstructured examples in the world. Many computational models of analogical and relational reasoning have been developed (for reviews, see Doumas & Hummel, 2005; Holyoak, 2005, 2012); however, DORA is the only model to date that provides an account for how the structured (i.e., symbolic) predicate representations of object properties and relations upon which models of relational reasoning rely can be learned from unstructured input (although see Lu, Chen, & Holyoak, 2012, for an alternate Bayesian account of how representations of weight vectors that support many forms of relational reasoning can be learned from simple feature lists). We focus here on the elements of the model that are important for our simulations of the availability heuristic - knowledge representation, retrieval, concept development, and relational mapping. For a complete description of the model in all its gory detail see Doumas et al. (2008).

Knowledge Representation DORA is a symbolic connectionist architecture (i.e., a computational model based on traditional connectionist computing elements which represents and manipulates structured representations) descended from Hummel and Holyoak's (1997, 2003) LISA model. Propositions in DORA are represented by a hierarchy of units (see Figure 2). At the bottom, semantic units represent the features of objects and roles in a distributed fashion. At the next level, these distributed representations are connected to localist units (POs) representing individual predicates (or roles) and objects. Localist role-binding units (RBs) link role and object units into specific role-filler sets. At the top of the hierarchy, localist P units link RBs into relational propositions.

To represent the proposition, 'Alex is messy and humble', PO units (triangles and large circles in Figure 2) represent the object Alex and the predicates *messy* and *humble*. POs are are connected to semantic units describing their features. For instance, Alex might be represented by features such as 'human', 'male', etc.; *messy* by 'high-entropy', 'disorganized', etc.; and *humble* by 'reticent', 'shy', etc. While we label semantic units for expositional clarity, these units need not have any intrinsic representational meaning. Rather, semantic units instantiate distributed representations in the traditional connectionist sense (e.g., Rumelhart, McClelland, & PDP Research Group, 1986).

Propositions in DORA's working memory (WM) are stored in two mutually exclusive sets. The *driver* represents DORA's current focus of attention, and controls the sequence of firing. The *recipient*, or active memory in Cowan's (2001) terms, contains propositions available for mapping to the propositions in the driver. Specifically, one to three proposition(s) become active in the driver (i.e., enter working memory). When a proposition enters working memory, role-filler bindings must be represented dynamically on units that maintain role-filler independence (i.e., POs and semantic units; see Hummel & Holyoak, 1997, 2003). In DORA, objects are dynamically bound to roles by systematic asynchrony of firing with units for roles and their fillers firing in direct sequence (see Doumas et al., 2008).

For example, to bind Alex to the *messy* predicate, the units representing Alex fire followed by the units representing *messy*. Separate role-filler bindings fire sequentially to form relational structures (e.g., Alex + *messy* fires followed by Alex + *humble*). As units in the driver become active, they impose patterns of activation on the semantic units. Units in the recipient compete via lateral inhibition to respond to the pattern of firing imposed on the semantic units. Grossly, this driver to recipient flow of activation corresponds to the process of comparison, where propositions in the driver will tend to activate propositions in the recipient to which they are most similar both structurally (through shared role semantics) and featurally (through shared object feature semantics).

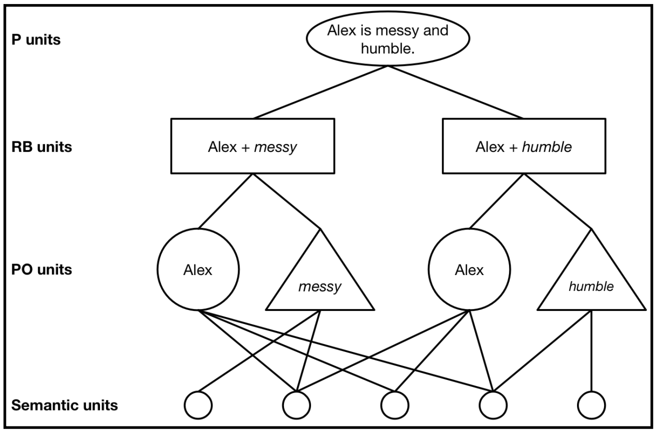

Figure 2. DORA representation of the proposition "Alex is messy and humble". PO units include roles (triangles) and objects (large circles) and are represented as patterns of simultaneous activation over semantic units (small circles) that encode their semantic content. RB units (rectangles) are represented as patterns of sequential activation of their constituent PO units (e.g., Alex followed by *messy*). P units (ovals) are represented as sequential activation of their constituent RB units (e.g., Alex + *messy* followed by Alex + *humble*). Simple propositions such as "Alex is messy and humble" can be bound to PO units to form higher-order relations.

These representations support structured thinking and when combined with a few simple processes (e.g., the ability to form mappings between co-active units in the driver and recipient, below) allow DORA and LISA to account for over 50 phenomena from the adult analogy making and relational reasoning literature (e.g., Hummel & Holyoak, 1997, 2003; Morrison, Doumas, & Richland, 2011; Viskontas, et al., 2004). Additionally, DORA provides an account of how these representations can be built from initially flat (i.e., unstructured) feature lists, and successfully simulates more than 40 phenomena surrounding the development of relational thinking (e.g., Doumas & Hummel, 2010; Doumas, Bassok, Guthormson, & Hummel, 2006; Doumas et al., 2008; Sandhofer & Doumas, 2008). Recent work explores how the model might account for linguistic behavior in a variety of domains (e.g., word segmentation, developing number and quantification (Hamer & Doumas, 2013), development of sociolinguistic markers, indicators, and stereotypes, etc). We now describe some of the core mechanisms in the DORA model. We focus on mechanisms crucial for simulating the findings of Tversky & Kahneman (1973) described above.

Retrieval During retrieval, units currently in the driver fire sequentially until every unit has been active (i.e., one phase set). Units in long-term memory (LTM) become active in response to the patterns imposed on the semantic units by the units in the driver. After all units in the driver have fired once, DORA places units from LTM into the recipient using Luce's (1959) choice axiom. For a full description of the retrieval algorithm used by LISA and DORA, see Hummel & Holyoak (2003).

Concept Development DORA uses comparison to bootstrap learning structured representations of new concepts. When DORA compares (via co-activation) items in the driver and the recipient, features common to both items are highlighted. For example, when DORA compares Emily Graslie and Alton Brown, the representation in the driver activates a set of semantic units while the representation in the recipient activates a different (overlapping) set of semantic units. Emily Graslie and Alton Brown share some traits (e.g., 'kind' and 'honest'), but also some differences (e.g., 'female' and 'scientist' vs. 'male' and 'chef'). When DORA compares these representations (i.e., when they are coactive) features shared by both people will receive roughly twice as much input and therefore become roughly twice as active as unshared features (see Figure 3). DORA exploits this differential feature activation and recruits a unit that learns connections to active semantic units in proportion to their activation. The new unit learns strong connections to any overlapping features and weaker connections to non-overlapping features. The result is a crude representation of the idea that 'kind' and 'honest' are related traits (along with several extraneous features, such as 'internet stardom'). Importantly, DORA can bind this new representation to objects in the future (via asynchrony-based dynamic binding, above), so this new representation functions as a single-place predicate.

The same process is used to refine previously learned predicates; one might later meet other people who are kind and honest and compare this experience to the existing representation of traits clustered around kindness and honesty. Again, some features will be shared, such as 'kind' and 'honest', while others, such as 'baker', will not. Features that are shared across many instances remain fundamental parts of the representation while features that occur rarely become more weakly connected to the concept. Over time, extraneous features "fall" out of the representation, leaving only the features that are invariant across instances of the concept. DORA uses this process to learn representations of invariance and refine these representations to learn concepts that are never experienced without contextual baggage, such as meeting a particular person. The ability to learn the invariants of concepts that are never encountered in isolation is a crucial element of human cognition. Currently, DORA provides the only account for how this learning might occur.

Relational Mapping DORA attempts to find correspondences between representations in the driver and representations in the recipient by activating items in the driver. Items in the driver impose a pattern of activation on semantic units, and items in the recipient compete via lateral inhibition to respond to this pattern of semantic activation. A set of mapping nodes build connections to the coactive items in the driver and recipient. This process differs from the learning algorithm described above by connecting to the *tokens* in the driver and recipient rather than the semantic units these tokens are connected to. This allows the system to build analogies without fundamentally altering the properties associated with the objects involved.

Where the representation in the driver has no structure corresponding to the representation in the recipient new units will be recruited in the recipient to fill the structural gaps (e.g., the corresponding trait for an item in a highly-related pair) using previously-learned corresponding items (e.g., 'kind' and 'honest').

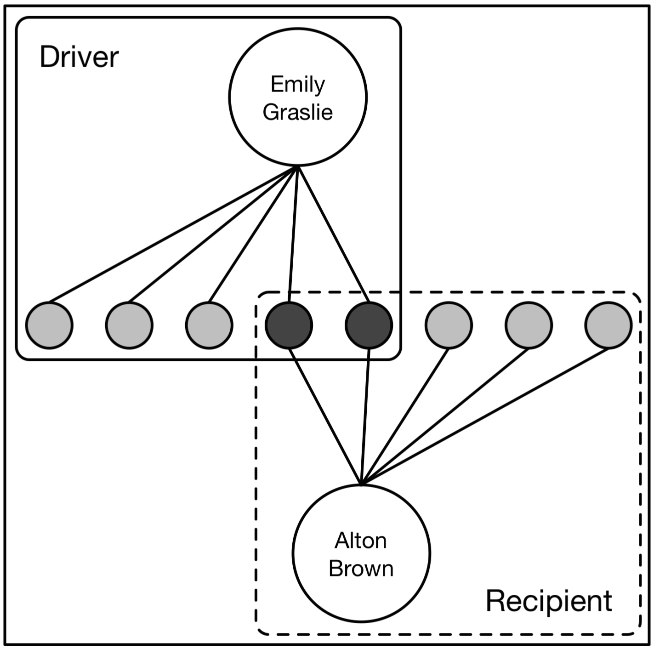

Figure 3. Concept development in DORA. Emily Graslie is in the driver while Alton Brown is in the recipient. A comparison is performed between them (i.e., they are coactivated) and the shared features (darker circles) receive approximately twice as much input as the unshared features. A new node is recruited to store this pattern of activation, the concept that 'kind' and 'honest' are related traits.

This process and its sensitivity to both structure and semantic features allow DORA to rule out inferences that are unreasonable; while the propositions "Aaron is honest" and "Aaron is messy" may exist in the recipient, the system will not infer that "Aaron is humble" because of the lack of correspondence between messiness and humility. This sensitivity to both structure and semantics allows DORA to exploit correspondences at the structural level (through shared role semantics) as well as at the feature level (through shared object semantics), and provides a comprehensive account of relational reasoning, including analogy discovery, analogical inference, schema induction, and (with representations of quantities) scalar implicature.

Simulations

Study 5 We simulated Study 5 in two steps. In the first step, we placed representations of 10 objects (i.e., the potential committee members) in the recipient. In the second step, we allowed DORA to run for three phase sets (a rough approximation of the length of time people focus on a task before serializing to something else) with four placeholder members in the driver, performing mappings between placeholder objects in the driver and committee members in the recipient (analogous to sampling with replacement). We constructed sets of distinct committees by examining the mappings incrementally, moving the starting point for committee composition forward by one place whenever we encountered a member already in the current set. We recorded both the median and mean number of distinct groups recalled of each size, reported in Table 1.

Group Size	DORA Mean	DORA Median	Tversky & Kahneman Median
2	5.5	5	68
3	4	4	50
4	3	3	40
5	2	2	28
6	2	2	28
7	1.5	2	32
8	1	1	18

Table 1: Results from simulating Tversky & Kahneman's (1973) Study 5

As can be observed in Table 1, DORA's performance in this simulated task accounts for much of the variance reported by Tversky & Kahneman (R^2 = .96 mean; .98 median). These results depend on the working memory constraints of the model that arise from the interaction between two factors in the neural system: its temporal resolution (i.e., the minimum amount of time needed to activate and deactivate a unit) and phase length (i.e., the maximum amount of time a unit can be inhibited before losing its exited state).

This is not a new explanation for sampling probability distribution at the computational level; however, we posit a mechanism by which such sampling processes might occur. Furthermore, this mechanism already accounts for many other phenomena within the availability literature and beyond.

Study 10 We simulated Study 10 in three steps. In the first step, we used DORA's relation learning algorithm to learn representations of the word pairs. We presented DORA with 240 instances of feature sets (corresponding to the trait pairs from Tversky & Kahneman, 1973). We used 120 instances to train the highly-related pairs and 120 instances to train the unrelated pairs. In the highly-related training set, 60 instances consisted of a highly-related pair of features as the only elements, while the remaining 60 instances consisted of sets of three traits which did not occur as a pair (roughly approximating the mean standardization reported for the highly related pairs). In the unrelated training set, 20 instances consisted of an unrelated pair of features as the only elements, while the remaining 100 instances consisted of sets of five traits, none of which occurred as a pair (roughly approximating the mean standardization reported for the unrelated pairs).

In the second step, we allowed DORA to learn new predicates for 600 learning iterations. During each iteration, a random object was retrieved from LTM and placed in the driver. This object was used to retrieve other objects from LTM via DORA's retrieval algorithm. DORA then compared the item in the driver with the items in the recipient using the concept development algorithm (described above) to create (or refine) predicates.

In the final step, we presented DORA with objects consisting of a single trait and used these objects to retrieve a predicate from LTM. In the highly-related condition, DORA retrieved the corresponding trait 38% of the time, while in the unrelated condition, DORA retrieved the corresponding trait 13% of the time. These results closely mirror those reported by Kahneman & Tversky, 41% in the highly-related condition and 19% in the unrelated condition.

Discussion

Dual-process accounts of the availability heuristic provide a computational level description of how resulting cognitive biases might arise. However, these accounts fall short of full explanatory power by assuming the existence of heuristics that drive cognitive biases (Chaiken (1980) and Evans (1989) go so far as to label their System 1 analogs *heuristic*). Additionally, dual-process accounts fail to explain how such heuristic processing might be implemented in the brain.

DORA moves beyond a computational level description by providing an account of how the availability heuristic might arise as a consequence of simple operations already in place for concept development and relational reasoning. These operations are implemented in a neurally plausible symbolic-connectionist architecture and already account for more than 40 developmental findings surrounding the development of relational concepts and reasoning as well as over 50 findings from the adult relational reasoning and analogy-making literature. In DORA, the availability heuristic emerges from the interaction of working memory constraints, relational mapping algorithms (Study 5), concept development, and retrieval algorithms (Study 10) originally designed to provide an account for the development of relational thinking. We believe that other heuristics characteristic of cognitive biases may arise through similar interactions in the DORA model.

Acknowledgements

The first author would like to thank Kaskade for the Atmosphere at The Shrine set.

References

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personal and Social Psychology*, 39, 752-766.

Cowan, N. (2001). The magical number four in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, 24, 87–114.

Doumas, L. A., Bassok, M., Guthormson, A., & Hummel, J. E. (2006). A theory of reflexive relational generalization. In R. Sun (Ed.), *Proceedings of the Twenty Fourth Annual Conference of the Cognitive Science Society* (pp. 1245-1250). Matwah, NJ: Lawrence Erlbaum.

Doumas, L. A. A. & Hummel, J. E. (2005). Approaches to modeling human mental representations: What works, what doesn't, and why. In K. J. Holyoak & B. G. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 73-91). Edinburgh, UK: Cambridge University Press.

Doumas, L. A., & Hummel, J. E. (2010). A computational account of the development of the generalization of shape information. *Cognitive Science*, 34(4), 698-712.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, 115(1), 1-43.

Evans, J. S. B. (1989). *Bias in Human Reasoning: Causes and Consequences*. Brighton, UK: Erlbaum.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.

Hamer, A. J. & Doumas, L. A. A. (2013). Discovering quantification and number in a role filler model. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the Thirty Fifth Annual Meeting of the Cognitive Science Society* (pp. 2276-2281). Matwah, NJ: Lawrence Erlbaum.

Holyoak, K. J. (2005). Analogy. In K. J. Holyoak & B. G. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 117-142). Edinburgh, UK: Cambridge University Press.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 234-259). New York, NY: Oxford University Press.

Hummel, J. E., & Choplin, J. M. (2000). Toward an integrated account of reflexive and reflective reasoning. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society* (pp. 232-237). Mahwah, NJ: Erlbaum.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104(3), 427-466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110(3), 306-340.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199-249.

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, 119, 617-648.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York, NY: Courier Dover Publications.

Morrison, R.G., Doumas, L.A.A., & Richland, L.E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, 14(3), 516-529

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1). Cambridge, MA: MIT Press.

Sandhofer, C. M., & Doumas, L. A. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development*, 9(2), 194-221.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1-66.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences*, 23(5), 645-665.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging*, 19(4), 581.