An Information-Theoretic Approach to Cognitive Dimension Reduction

Maya Leshkowitz *
Department of Cognitive and Brain Sciences
The Hebrew University of Jerusalem
mleshkowitz@gmail.com

Abstract

We introduce Cognitive Dimension Reduction (CDR), a framework that sheds light on how individuals simplify the multidimensional world to guide decision-making and comprehension. Our proposal posits that cognitive limitations prompt the adoption of simplified models, reducing the environment to a subset of dimensions. Within these limitations, we propose that individuals exploit both environment structure and goal relevance. Employing information theory, we formalize these principles and develop a model that explains how environmental and cognitive factors influence dimension reduction. Furthermore, we present an experimental method for CDR assessment and initial findings that support it.

1 Introduction

Various decisions we make, such as when to invest in the stock market, entail a great amount of cognitive processing. We constantly make decisions based on streams of dynamic, high-dimensional information with limited cognitive resources (Bach and Dolan (2012)). The last decades have seen the emergence of influential theories, according to which we make such decisions using various simplifications. For example, schema theory hypothesizes cognitive structures that define relations between relatively few dimensions or categories (Gilboa and Marlatte (2017), Rumelhart (1980)). In reinforcement learning, it has been suggested that people select a small subset of all dimensions to learn about in a process known as representation learning (Gershman and Niv (2010), Wilson and Niv (2012)). A central tenet of these theories is that behavior and comprehension are based on a subset of prominent dimensions, which partition the world based on continuous and categorical features.

In computational terms, behavior is shaped by a dimension-reduction process. For example, an investor might sell a stock based on its past performance while ignoring other dimensions, such as the market's trend. This paper lays out principles of cognitive dimension reduction: finding a subset of dimensions that exploit the environment's structure and are goal-relevant. Crucially, this process is performed under the constraints of limited resources (Section 2). We formulate these principles using information theory and propose a quantitative model called Cognitive Dimension Reduction ($\mathcal{CDR}$) (Section 3). We conclude by proposing an experimental method for testing the $\mathcal{CDR}$ model and initial findings that support its validity (Section 4).

2 The principles of cognitive dimension reduction

People are often motivated to comprehend the environment in order to achieve their goals. However, limitations such as cognitive capabilities and time constraints allow processing only some dimensions,

*I would like to thank Ran R. Hassin for his support and helpful discussions and suggestions. I am also grateful to Uri Zak and Rasha Kardosh for their encouragement and many helpful comments.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

where dimensions correspond to attributes or aspects of the environment. The best feasible solution in this case, which is the one we propose in the model, is to use a subset of dimensions that approximates the environment's structure and are goal-relevant.

2.1 The structure of the environment

The scholarly consensus is that the world is, and perceived to be, structured (e.g., Rosch (1975), Rosch and Mervis (1975)). Rather than consisting of orthogonal dimensions with uniform distributions, the world consists of correlated dimensions (Berlin et al. (1973)). Humans and animals take advantage of structure to enhance comprehension and learning (Gershman and Niv (2010), Kemp and Tenenbaum (2009)). This is also evident at the neuronal level (Barlow et al. (1961), Simoncelli (2003)).

One way to take advantage of structure is through abstraction. By abstraction, we refer to the belief that two or more subjectively distinguishable objects have the same value along some dimension (Gilead et al. (2020)). For example, referring to companies such as Pfizer and Johnson&Johnson as pharmaceutical companies is an abstraction. Abstraction is thus a dimension reduction process that highlights some dimensions (such as involvement in pharmaceuticals) while ignoring others (such as involvement in consumer products).

2.2 Relevance for goals

People are sensitive to the extent that dimensions are relevant to their tasks and goals (Barsalou (1991), Eitam and Higgins (2010), Solomon et al. (1999)). To illustrate, different dimensions might become prominent when buying a stock, depending on whether the goal is short-term or long-term revenues. When learning, people can consider relevance alongside environment structure (Bates et al. (2019)).

2.3 Information processing constraints

When confronted with high-dimensional information, time constraints and cognitive limitations prevent people from making optimal decisions (Gigerenzer and Selten (2002), Simon (1955)). In such situations, decisions are often formed after reducing the environment to a few prominent dimensions (Brewer and Treyens (1981), Kleider et al. (2008), Sims (2010)). One advantage of using only a subset of the dimensions is that it requires less memory capacity (Brady et al. (2009)). In addition, focusing on fewer dimensions reduces the attention load, which may facilitate learning (Bhui and Jiao (2023), Leong et al. (2017)).

In the next section, we propose an information-theoretic model of cognitive dimension reduction that ties together the aforementioned principles. The model offers a quantitative method for determining the dimensions to which the environment is reduced.

3 CDR: an entropy-based dimension reduction model

Consider an investor who thinks that the value of Pfizer's stocks depends on two dimensions: the general trend of the stock market and Pfizer's achievements. The value of the stock and the two dimensions used to explain its value can be formulated as random variables. In information theory, the Shannon entropy of a random variable is a measure of the average information inherent in the variable's outcomes (Shannon and Weaver (1949)). The Shannon entropy is suited for measuring the joint information of these two dimensions since it considers the redundancies between them. The amount of information regarding the stock value obtained from observing the other dimensions can be measured using their mutual information (see also Cover and Thomas (2012) for detailed explanations on information theory).

In the model, we use the Shannon entropy not only as a measure of informativeness but also as a proxy for the cognitive complexity of attending, memorizing, and using dimensions. This hypothesis builds on previous work that applied information theory across a range of cognitive processes (for a review, see Sayood (2018)). For example, the time it takes to process and recognize elements is linearly related to their entropy (Hick (1952), Hyman (1953)). Recently, links between entropy and cognitive neuroscience were established in the predictive brain framework (Clark (2013)). Within this framework, the free energy principle postulates that the brain copes with the overload of high-dimensional information by striving to minimize the entropy of its prediction errors (e.g., Friston (2010)).

3.1 The model

We assume that the set of dimensions in the environment $D = \{d_1, .., d_k\}$ and their distributions are known. Cognitive Dimension Reduction outputs a subset of these dimensions $D' \subseteq D$ in the context of comprehending or predicting a target dimension V.

$$\mathcal{CDR}\,(D, V) = \underset{D' \subseteq D}{\arg\max}\,(I\,(D'; V)) \tag{1}$$

subject to

$$H\,(D') \leq C \tag{2}$$

Equation 1 represents the incentive to accurately learn the dimensions D' most informative of the target dimension V. The mutual information I measures the amount of information from V that can be learned by observing a subset of dimensions D'. Equation 2 represents the information processing constraint on the dimensions that can be used. The Shannon entropy, H, measures the expected information in dimensions D'. The cost parameter, $C > 0$, may be affected by situational factors such as time constraints and individual abilities such as working memory and attention capacities.

Put together, the dimensions $\mathcal{CDR}\,(D, V)$ are maximally informative of dimension V, out of all subsets of dimensions whose entropy is upper bounded by C.

4 Experimental evidence

Next, we introduce an experiment that demonstrates an application of the $\mathcal{CDR}$ model and initial evidence supporting it. This experiment examines the dimensions used for decisions and evaluations, tapping into the downstream consequence of cognitive dimension reduction. We stress that the experiment was not run to test the $\mathcal{CDR}$ model, but rather, it inspired the model. Therefore, we present the experiment as an example of the model's application rather than a verification of its validity. Moreover, the experiment only examines one set of values for the model's variables, and additional work should test the model's predictions with other values.

4.1 Grouping and averaging

Researchers have demonstrated that people evaluate aggregate options by averaging across values in various domains (Anderson (1965), Brusovansky et al. (2019)), including stock market evaluations (Betsch et al. (2006)). The grouping and averaging approach (Shah and Oppenheimer (2011)) extends this observation by showing that people first group information and then evaluate each group by averaging the values associated with the group; finally, an overall evaluation is formed by averaging groups' evaluations. Our experimental results are consistent with the possibility that the dimension that $\mathcal{CDR}$ outputs is the one according to which grouping and averaging are performed.

4.2 Experimental method

Participants were told they would be presented with two stock portfolios, each consisting of equal stock shares. Participants were then shown two sequences of 19 stocks, one for each portfolio (the order portfolios were presented was counterbalanced between participants). For each stock, they first saw what industry this stock belonged to, and then, if the value of the stock rose or fell over the previous week (Figure 1). After the stocks' presentation, participants were asked to choose the portfolio that performed better and only then to evaluate the performance of the industries. (The method of this experiment was adapted from Woiczyk and Le Mens (2021)).

The two portfolios in the experiment had the same sequence of rising and falling stocks and, hence the same overall performance (Figure 2). The stocks' assignment to industries differed between the two portfolios. When grouping and averaging based on industries (see 4.1), one portfolio, i.e., the better grouping by industry portfolio, was more favorable. However, if participants grouped and

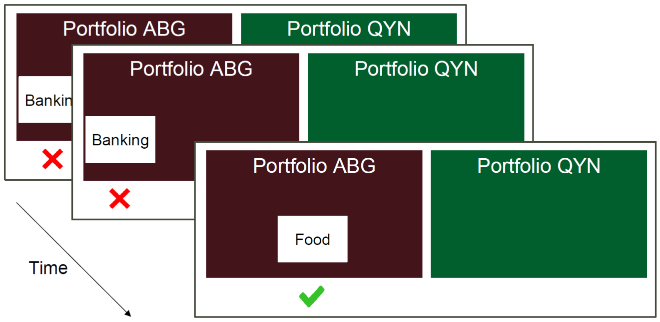

Figure 1: For each portfolio, participants saw a sequence of stocks. For each stock, they saw what industry it belonged to (e.g., Banking, Food) and if its value rose or fell.

averaged based on individual stocks (or did not group at all), they would be equally likely to choose either portfolio.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Better by industry portfolio:	V	V	X	V	V	X	V	X	X	X	V	X	V	X	V	V	X	X	V
Worse by industry portfolio:	V	V	X	V	V	X	V	X	X	X	V	X	V	X	V	V	X	X	V

Figure 2: A summary of the portfolio information participants saw in the experiment. The sequence of stocks was presented according to the above order. The cells' colors represent the industries (e.g., Food, Utilities, Products) which were randomly assigned in the experiment.

4.3 Relation to the CDR model

As we describe next, conditions in this experiment were such that the model predicted most participants would reduce the information to the industry dimension and use it for grouping and evaluating a portfolio's performance. We model each stock as a random sample from a three-dimensional space (IN, S, V). The industry dimension IN has three possible values (known to participants in advance), S is an identifying dimension with a unique value for each stock, and V is the binary change in stock value (rise or fall), which was the target dimension in this experiment. A priori, participants can perceive a portfolio's performance by reducing the information to the industry dimension IN, the stock dimension S, both, or neither of these dimensions.

Since there were 19 stocks in each portfolio, the stock dimension S had relatively high entropy ($H(S) = \log 19 = 4.25$). The industry dimension had relatively low entropy since there were only three industries (i.e., $H(IN) \leq \log 3 = 1.59$). Following Miller (1956), we expected most participants' parameter C (in Eq. 2) to be in the range that the information constraint would be satisfied for the industry but not for the stock dimension.

In addition to being a sufficiently simple dimension, the industry dimension is highly informative in this experiment. For every industry, either the values of all its stocks rose or they all fell (i.e., $I(IN; V) = H(V) > 0$). Participants saw several stocks in each industry, allowing them to learn the association between industries and values throughout the task. The remaining alternative of ignoring all the dimensions would not reveal any information about a stock's value before it was presented ($I(\emptyset; V) = 0$). It follows that out of the options that satisfy the information constraint (Eq. 2), the industry dimension attains $\arg\max_{D' \subseteq D} \left(I\left(D'; V\right)\right)$.

To conclude, using $\mathcal{CDR}$, we predicted that the information would be reduced to the industry dimension in this experiment. Thus, even though the two portfolios had the same performance, we expected participants would prefer the portfolio with better grouping by industry performance.

4.4 Results

One hundred and twelve participants recruited via Prolific completed the experiment; we excluded 11 participants based on low accuracy on the industry evaluation questions, leaving 101 participants.

Even though both portfolios had the same performance, when asked which portfolio performed better, a significant majority of the participants chose the portfolio with better grouping by industry performance (68%, i.e., 73 of the 101 participants, chi-square: $\chi^2 = 19.17$, $p < 0.0001$).

5 Discussion

According to the $\mathcal{CDR}$ model, people prioritize a subset of the possible dimensions in the environment that allow them to achieve high values without incurring a high informational cost. In computational terms, people perform a lossy dimension reduction, which is optimal once accounting for cognitive and environmental limitations.

$\mathcal{CDR}$ is a static model that assumes dimensions and distributions are fixed and known. As a result, the model is less suited for predicting behavior when there are misconceptions regarding the distributions of the dimensions or their informativeness, which hinders revealing the best dimensions. Such misconceptions may occur when uninformative dimensions are salient. $\mathcal{CDR}$ should fare better when predicting the behavior of experienced individuals or when the environment is relatively stable.

References

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of experimental psychology*, 70(4):394.

Bach, D. and Dolan, R. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature reviews neuroscience*, 13(8):572–586.

Barlow, H. B. et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01):217–233.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In *Psychology of learning and motivation*, volume 27, pages 1–64. Elsevier.

Bates, C. J., Lerch, R. A., Sims, C. R., and Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of vision*, 19(2):11–11.

Berlin, B., Breedlove, D. E., and Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American anthropologist*, 75(1):214–242.

Betsch, T., Kaufmann, M., Lindow, F., Plessner, H., and Hoffmann, K. (2006). Different principles of information integration in implicit and explicit attitude formation. *European Journal of Social Psychology*, 36(6):887–905.

Bhui, R. and Jiao, P. (2023). Attention constraints and learning in categories. *Management Science*.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4):487.

Brewer, W. F. and Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology*, 13(2):207–230.

Brusovansky, M., Vanunu, Y., and Usher, M. (2019). Why we should quit while we're ahead: When do averages matter more than sums? *Decision*, 6(1):1.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204.

Cover, T. and Thomas, J. (2012). *Elements of Information Theory*. Wiley.

Eitam, B. and Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a representation (roar) as a new framework. *Social and personality psychology compass*, 4(10):951–967.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

Gershman, S. J. and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current opinion in neurobiology*, 20(2):251–256.

Gigerenzer, G. and Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gilboa, A. and Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in cognitive sciences*, 21(8):618–631.

Gilead, M., Trope, Y., and Liberman, N. (2020). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, 43:e121.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology*, 4(1):11–26.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of experimental psychology*, 45(3):188.

Kemp, C. and Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological review*, 116(1):20.

Kleider, H. M., Pezdek, K., Goldinger, S. D., and Kirk, A. (2008). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(1):1–20.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., and Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2):451–463.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, 104(3):192.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605.

Rumelhart, D. E. (1980). *Schemata: The Building Blocks of Cognition*. Erlbaum Associates, Mahway.

Sayood, K. (2018). Information theory and cognition: a review. *Entropy*, 20(9):706.

Shah, A. K. and Oppenheimer, D. M. (2011). Grouping information for judgments. *Journal of Experimental Psychology: General*, 140(1):1.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press, Mahway.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current opinion in neurobiology*, 13(2):144–149.

Sims, C. A. (2010). Rational inattention and monetary economics. In *Handbook of monetary economics*, volume 3, pages 155–181. Elsevier.

Solomon, K. O., Medin, D. L., and Lynch, E. (1999). Concepts do more than categorize. *Trends in cognitive sciences*, 3(3):99–105.

Wilson, R. C. and Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in human neuroscience*, 5:189.

Woiczyk, T. K. and Le Mens, G. (2021). Evaluating categories from experience: The simple averaging heuristic. *Journal of Personality and Social Psychology*, 121(4):747.