Machine learning-based measure of cognitive complexity explains variance in rank-ordered preferences

Shabnam Hakimi, Yan-Ying Chen, Monica Van, Scott Carter, Emily Sumner, Nayeli Bravo, Kalani Murakami, Yanxia Zhang, Charlene Wu, & Matthew Klenk
{shabnam.hakimi, yan-ying.chen, monica.van.ctr, scott.carter, emily.sumner, nayeli.bravo.ctr, kalani.murakami.ctr, yanxia.zhang, charlene.wu, & matt.klenk}@tri.global
Toyota Research Institute
Los Altos, CA 94022 USA

Abstract

Cognitive complexity can provide insight into how people make decisions, ranging from the most minor to the most impactful. Here, we present a novel approach to inferring the complexity of processes associated with preference and decision making. We measured the complexity of participant-generated descriptive features of consumer products and the relationship to preference rankings. In order to measure cognitive complexity over a sparse set of features, we developed a natural language processing approach that compared the descriptive words generated by participants to those generated by a machine learning model; words that were more distinct from those generated by the model were rated more complex. We show preliminary evidence that cognitive complexity is related to preference for products, explaining unique variance in rankings and also capturing a new facet of the process through which preference is revealed through choice. We also show the value of participant-generated features for understanding choice processes.

Keywords: cognitive complexity; decision making; preference; mental models; natural language processing

Introduction

Water or cola? Cola or water? Even simple choices made while standing in a grocery store aisle are deceptively simple. The moment the bottle of water and not the cola ends up in your hand is realized through a process that involves evaluating the space of possible actions, assigning value to them, choosing an action, and evaluating its outcome for use in future decisions (Rangel, Camerer, & Montague, 2008). Here, you must represent the value of choosing the water bottle or the cola bottle, considering the attributes of each beverage. Preferences over these features are a critical part of the valuation process. The mechanism through which preference emerges and influences valuation is much debated (e.g., constructed v. discovered preference; (Slovic, 1995) v. (Plott, 1996)), reflecting epistemic differences across and within disciplines as broad as economics, psychology, and philosophy. The way preference is defined has meaningful implications, affecting not only how we understand an individual's choice of cola over water, but also in the modeling and prediction of group decisions. On top of this, people's stated preferences are not always indicative of their true preferences or desires. People are prone to framing effects (Tversky & Kahneman, 1985; Chang, 2008), serial position effects (Murdock Jr, 1962; Deese & Kaufman, 1957; Bar-Hillel, Peer, & Acquisti, 2014; Sumner, DeAngelis, Hyatt, Goodman, & Kidd, 2015), and peer pressure. Such phenomena can make it difficult to infer what is important to people using standard preference elicitation methods.

Here, we take a parsimonious view of preference, defining it as *motive bias*, or more concretely, bias over a space of stimulus features that has the potential for action. But how should we define the feature space of a given stimulus? We argue that this space is subjective, and while there may be overlap in the features that an individual considers when expressing preferences, these features and their mental representations can vary meaningfully across people. The representation of these features can involve varying levels of processing, ranging from simpler "visceral" reactions that rely on observable/perceptual features (e.g., water droplets on the cola bottle) to more complex "behavioral" reactions to perceived utility or functional features (e.g., cold cola is refreshing) to even more complex and idiosyncratic "reflective" reactions that embed the stimulus in the user's stories (e.g., getting a cold cola bottle with friends on a hot summer day) (Norman, 2002; Craik & Lockhart, 1972). Moreover, such idiosyncratic features will introduce variance, affecting population-level inference over any measure of preference.

How might a memory of a cold cola with friends on a hot summer day influence the features that are important for decision making? In addition to the individual features of the cola stimulus, which can vary in their complexity (as above), the features can interact, leading to highly complex, emergent features that drive individual preferences (Craik, 2002). Understanding these idiosyncrasies and their complexity can give us insight into the mechanisms of decision making, including any properties that generalize across choice contexts.

In the present study, we aimed to capture the idiosyncrasies in the mental representation of preference by decomposing a commonly used preference elicitation method, rank-ordered preference, into two parts. This approach allowed us to move away from simplistic choice tasks (e.g., binary choice with limited, experimenter-defined features) and toward a more ecologically-valid, realistic choice scenario. Using a novel preference elicitation task, we demonstrate that individuals generate idiosyncratic feature spaces when considering stimuli (here, consumer goods). We expected that the complexity of the underlying mental representation associated with feature generation would be meaningful for predicting individual choices, so we developed a novel machine-learning approach

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

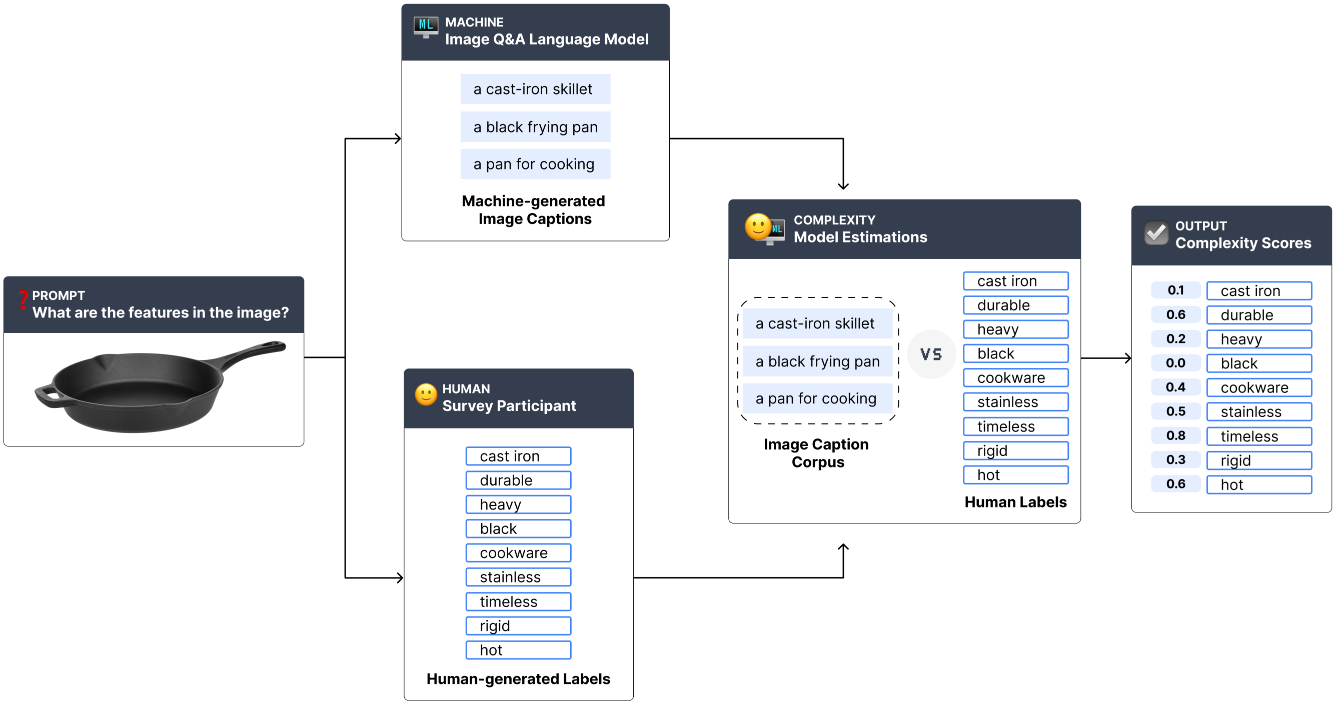

Figure 1: System for computing cognitive complexity of human-generated text using machine learning.

to compute complexity. We demonstrate that our measure of feature cognitive complexity is able to capture unique variance when predicting subsequent preference rankings. We close by discussing future work and potential applications of our approach.

Operationalization of Cognitive Complexity Using Machine Learning Models

We acknowledge the many facets of cognitive complexity and take a broad view of the construct. Here, we define cognitive complexity as the additional context necessary to understand human-generated language in response to a visual stimulus. But how do we define context? In this paper, we consider the context as the information about a visual stimulus that is not retrievable by an observer from the stimulus itself; for example, when seeing a product image, the word "couch" (an observable visual feature) requires less context than the phrase "watching videos" (a remembered or imagined state that is not depicted). Such text can be analyzed by psychological models and language analysis tools such as LIWC (Pennebaker, Booth, Boyd, & Francis, 2015), but these approaches lack an understanding of non-linguistic, visual context. A new measure built jointly on vision and language understanding is therefore needed to more fully realize the context of the language.

Following this definition, we use a vision and language model, OFA (Wang et al., 2022), to generate a corpus as a reference of information that can be observed in a given image. OFA is a neural network model pre-trained by a set of vision and language datasets; it serves a variety of downstream tasks such as text-to-image grounding, visual question answering, and image captioning, as used in this paper. The vision and language datasets that power OFA include RefCOCO (Kazemzadeh, Ordonez, Matten, & Berg, 2014; Yu, Poirson, Yang, Berg, & Berg, 2016), VQAv2 (Goyal, Khot, Summers-Stay, Batra, & Parikh, 2017), MSCOCO (Chen et al., 2015), etc.; the language data are generated and evaluated through crowdsourcing and are directly relevant to perceptible image content. For example, the human annotators for MSCOCO dataset were instructed to describe all the important parts of the scene and not to describe things that might have happened in the future or past. Given the nature of these human annotations in the training data, OFA can generate visual descriptions of what can be observed in a given image. Figure 1 (machine-generated image captions) includes sample visual descriptions generated by OFA for the consumer product images used in our experiments. These machine-generated descriptions become a reference corpus of normative descriptions for images and are therefore a reasonable comparison set for features generated by our participants. If the machine generates a word, it is more likely to be commonly used and, by our definition, is less complex.

Experimental Data

Methods

Participants 1266 U.S.-based adults (ages 18 and older) were recruited via Amazon Mechanical Turk (Amazon Mechanical Turk, n.d.) in February - March 2022 to participate

Age	%	Gender	%	Race	%
18-25	5.63	Woman	41.94	American Indian or Alaskan Native	2.59
26-35	38.35	Man	56.70	Asian	6.94
36-45	29.42	Genderqueer or Non-binary	0.49	Black or African American	8.33
45-55	16.21	Agender	0	Middle Eastern or North African	0.83
56-65	6.80	Prefer to self-describe	0.10	Native Hawaiian or Other Pacific Islander	0.56
66+	2.91	Prefer not to answer	0.78	White	78.89
Prefer not to answer	0.68			Prefer to self-describe	1.30
				Prefer not to answer	0.56

Table 1: Participant demographics (N=1030 adult, U.S. participants).

in the study. In order to participate, individuals were required to have a HIT approval rate of >95% and >5000 approved HITs. 236 participants were excluded from the dataset because they did not complete all study tasks. Thus, the final sample size was 1030 participants. (See Table 1 for demographics.)

Procedure

Our research procedure was reviewed and considered exempt by the WCG IRB. Upon agreeing to participate in the study, all participants provided informed consent. Participants then reviewed instructions for the Preference Elicitation Task (described below). After completing the Preference Elicitation Task participants completed a demographic assessment. Median participation time was 31 min 55 sec and each participant was compensated $7.50 USD.

Preference Elicitation Task All participants completed a Preference Elicitation Task (Figure 2).

Because we were interested in resolving the mechanism of rank-ordered preference elicitation across individuals, we designed the task to interfere as little as possible in the underlying psychological processes associated with decision making. We prioritized several factors in the design of the task:

1. Minimized experimenter demand. We allowed participants to generate as many—or as few—words as they liked, over a relatively long period. (A Description Challenge trial would advance automatically after two minutes if a minimum of five words had been entered.) We measured task behaviors through reaction times, etc., but allowed the process to proceed in as uninterrupted a way as possible. Given that engaging with cognitive and affective processes can change their their trajectory, we prioritized avoiding such effects (e.g., (Torre & Lieberman, 2018)).

2. Minimized effort cost. Relatedly, we designed the interface such that interactions were simple and intuitive (e.g., an empty text box would appear onscreen only after text had been entered into the previous one), allowing participants to focus only on the preference elicitation task.

3. Diversity of participant-generated features. We accepted any text, even if grammatically inappropriate given the instructions (e.g., not an adjective). This allowed us to capture a wider array of responses reflecting individual variability in preference representation.

The task had two main phases: the Description Challenge and the Ranking Challenge (Figure 2). These sections were followed by a brief assessment of an individual's knowledge of and experience with each item. First, in the Description Challenge, participants viewed 20 randomly ordered images of consumer products selected from the Amazon-Berkeley Objects Dataset, which was accessed on January 5, 2022 from https://registry.opendata.aws/amazon-berkeley-objects. Images represented a variety of consumer products (Figure 2a). Participants were instructed to use the text boxes to the right of each item to list any words that describe the image. Participants could use as much time as they liked to enter up to 20 words, one per box; a minimum of five descriptive features was required to advance to the next trial. The descriptive features generated in this phase of the task served as a representation of each participant's feature space or mental model of the stimuli; this individual feature space was then used in preference elicitation.

Next, in the Ranking Challenge (Figure 2b), participants were shown the features that they themselves had generated in the Description Challenge. Participants were instructed to use their cursor to order the text boxes by dragging them to the position that reflected each word's importance to them, with the top box (numbered '1') being the most important. Features were presented in random order and participants were not required to reorder the text boxes, although most chose to do so; randomization aimed to minimize the effect of feature generation order on the ranking process. Participants were also asked to indicate the words that were most important to their assessment of the item in the image using a horizontal blue line. They were instructed to place the line to indicate where in the list they became indifferent to the rank ordering.

Cognitive Complexity Model

To quantify the cognitive complexity of the features generated by the participants, we developed a machine learning approach capable of more effectively capturing the visual context of the Preference Elicitation Task. Simply put, we computed cognitive complexity by calculating the overlap between a feature generated by a human participant and a reference corpus generated by a machine for the same image. Be-

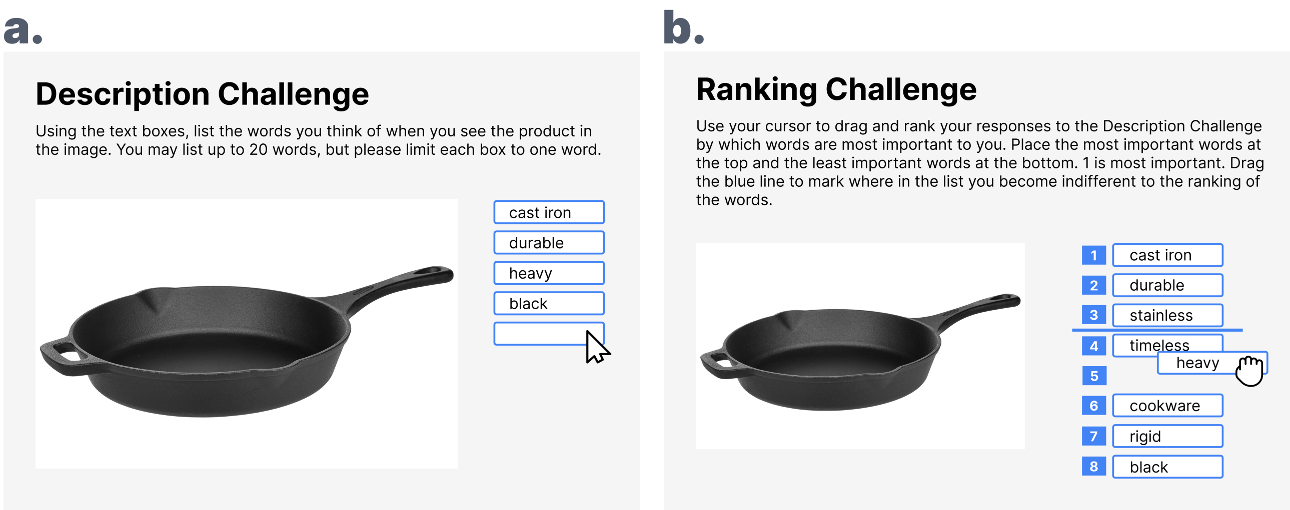

Figure 2: Preference Elicitation Task. a) Description Challenge. In the first phase of the task, participants were asked to enter up to 20 words describing each of 20 consumer product images. b) Ranking Challenge. In the second phase of the task, participants were asked to rank order their own self-generated features, again for all 20 consumer products. Features for each image were presented in random order. Participants also used a horizontal line to indicate the point in the list where they were indifferent to the rankings.

cause the vision and language model used to generate the reference corpus will necessarily be biased toward descriptions of the image's visual characteristics, lower overlap between a feature and reference corpus suggests that more context beyond the given image is required to understand the feature; the human-generated descriptive feature will therefore have higher complexity. The flow-chart of computing complexity scores is presented in Figure 1.

As described in the Introduction, OFA, a statistical vision and language model for a range of tasks such as image captioning and image question answering, is used to generate the reference corpus for each image. OFA uses transformers as the backbone architecture. The model is pretrained by using vision and language benchmark datasets. By feeding a text prompt (i.e., the survey question "what are the features in the image?" used in the preference elicitation paradigm) and an image (i.e., one of the product images shown to the human participants) to OFA as the input, the pretrained OFA generates a caption output. To increase the coverage of potentially generated captions, we implement two data augmentation strategies. First, we paraphrase the original question to M prompts. Second, for each prompt we set the beam size in the beam search of the caption generation process to N to generate N variants of captions. The data augmentation creates $M \times N$ captions as a reference corpus for a given image, where $M = 33$ and $N = 100$ in our implementation.

Once the reference corpus for each image has been generated, it can be compared to the features generated by the participant in the Description Challenge (as shown in Figure 1). Unlike the keyword-like descriptors generated by human participants, the machine-generated reference text is caption-like. To address this, we match the human-generated labels in the corpus of machine-generated text and count the occurrences. The complexity of a text label generated by a human participant is formulated as $1 - (overlap/max(overlap, cap))$, where $overlap$ is defined as the number of times the human label appears in the machine-generated reference corpus and cap is between 1 and the number of total words in the reference corpus for regularization that prevents the value from vanishing in a large corpus. Note that, unlike humans that generate a set of features over time, the machine generates multiple captions based on beam search to expand the most possible words and explore variability. The nature of different mechanisms between human and machine is not addressed in this paper and is an interesting topic for the future work.

The proposed operationalization and modeling approach attempt to quantitatively measure cognitive complexity; however, many challenges remain. For one, this paper only addresses complexity that may be captured in sparse human language about static visual images; complexity that arises outside of this scope may need alternative operationalizations to estimate. Another challenge is the coverage of complexity estimation. We start from an intuitive approach by creating a reference corpus, which keeps the interpretibility of the data for better sense-making. However, the coverage of reference corpus is limited to what the vision and language model can generate, which is highly dependent on the training data that was used to train the model. It is possible that a human-generated label requires no additional context and yet still not overlap with the reference corpus because the generated captions has a very limited vocabulary. We will discuss how to address these challenges in section .

Behavioral Data Analysis

Analysis was conducted using `R` (R Core Team, 2022), RStudio (Posit team, 2022), and the packages `plyr` (Wickham, 2011b), `tidyverse` (Wickham et al., 2019), and `GLMMadaptive` (Rizopoulos, 2022). Data were visualized using `ggplot2` (Wickham, 2011a), `viridis` (Garnier et al., 2021), and `sjPlot` (Lüdecke, 2022).

Hypotheses

We tested three main hypotheses. First, we tested whether we could indeed decompose a common rank-ordered preference elicitation and capture individual differences in the feature space over which people express preferences. Given established individual differences in perception and making meaning of visual stimuli (Partos, Cropper, & Rawlings, 2016), we expected that individuals would differ in the size and quality of self-generated features for consumer products. Second, we expected that the way in which individuals generated features (i.e., the order of descriptive feature generation) would not be the sole determinant of subsequent ranking order. In other words, features that were generated first would not necessarily be ranked as most important. (Important features related to episodic memory processes, for example, may take longer to generate than solely perceptual characteristics.) Third, we tested whether the complexity of self-generated descriptive features would predict subsequent choice, as measured by ranking. We expected that this novel measure would improve prediction of rankings because of a higher likelihood of capturing idiosyncratic drivers of preference such as memories (Weber & Johnson, 2006).

Results

Individuals have different feature spaces We first tested the hypothesis that individuals would vary in the number and characteristics of the features they generated in the Description Challenge. While the modal number of features generated was the required minimum of 5, many participants generated additional features, up to the maximum of 20 (mean number of features generated = 8.36, s.d. = 3.41; median = 5, IQR = 1).

Individuals also generated qualitatively different features from one another (Table 2). For example, when describing a cast-iron skillet, participants generated a range of responses. Some descriptions were of the perceptual characteristics of the skillet (e.g., “black”), while others were more evaluative (e.g., “beautiful”, “durable”). Others seemed to draw on complex semantic representations. One participant, ID14, generated a sequence of words that suggested a specific episode, evoking “camping” and “delicious” “eggs” and “bacon.”

Complexity Explains Unique Variance in Rankings As expected, the first feature generated was often ranked first and the order of generation was associated with the rank order. However, the relationship between feature generation and rank order was not correlated one-to-one (Figure 3).

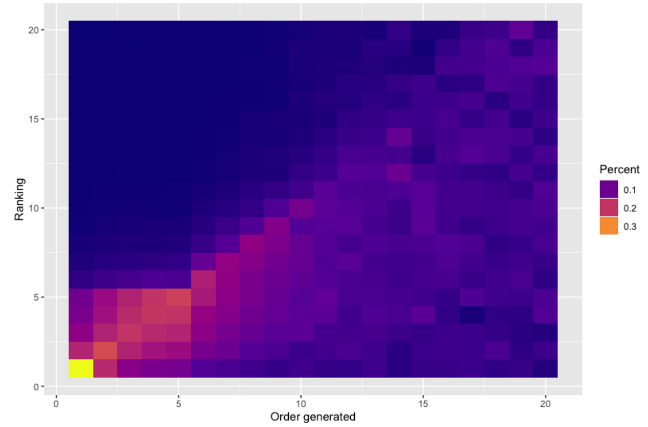

Figure 3: Heatmap depicting relationship between feature generation order and subsequent ranking. While the order of feature generation is correlated with the ranking, the relationship is not deterministic.

To understand this pattern and its relationship to cognitive complexity further, we fit a mixed effects ordinal regression to test the relationship between the order of feature generation and the ranking, while controlling for demographic characteristics and participant identity (random effect) (M1[1]).

We then fit the model again, now adding complexity and its interaction with generation order (M2[2]). The addition of complexity to the model improved model fit (Table 3). Additionally, the complexity term explains unique variance separate from of the interaction with feature generation order (Z = 28.07, p <0.0001).

Discussion

Here, we demonstrated a novel experimental approach to preference elicitation as well as a novel machine learning model for measuring the cognitive complexity of self-generated stimulus features; we used these new methods together to show that cognitive complexity explains variance in the relationship between feature representation and rank-ordered preference.

Individuals showed extensive variability in the way they approached the Description Challenge, generating both different numbers and types of features. These data suggest that individuals differed in the way that they constructed and represented the feature space for the product in each image. Moreover, individuals also seemed to vary in the depth of their features, with some “deeper” features evoking rich scenarios (Norman, 2002; Craik, 2002). The diversity of the features generated speaks to value of moving away from experimenter-defined features in preference elicitation and other tasks. In addition to the influence of their own personal biases, an experimenter might focus only on shallow features

[1] rank_order $\sim$ generation_order + age + gender + (1|participant)

[2] rank_order $\sim$ generation_order $*$ complexity + age + gender + (1|participant)

Participant	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Feature 9	Feature 10	Feature 11
ID7	beautiful	colour	useful	looking	valuable	easy	cook	better	super		
ID11	iron	heavy	durable	handles	pour						
ID14	heavy	solid	delicious	camping	eggs	bacon	outdoors	kitchen	hard		
ID27	cast iron	durable	heavy	black	cookware	stainless	timeless	rigid	hot		
ID30	cooking	iron	durable	lasting	food	meat	steam	sizzle	black	useful	dynamic

Table 2: Examples of features generated to describe a cast iron skillet.

	M1	M2
Marginal R^2	0.106	0.115
Conditional R^2	0.195	0.205

Table 3: Marginal and conditional variance in rank order explained by M1 and M2 terms.

(e.g., shape) or those with direct relevance to hypotheses (e.g., value for money). In either case, data may be more difficult to interpret, but less true to the underlying processes associated with preferences.

Individual differences were also manifest in the relationship between feature generation order and rank order. We expected that the order of feature generation would be important for rankings, since top-of-mind associations can drive preference (Keller, 1993; Szalay & Deese, 1978; Nelson, McEvoy, & Dennis, 2000). At the same time, however, features generated first are likely to be shallower (Craik & Tulving, 1975) and less semantically complex (Lohman, Tomanek, Ziegler, & Hahn, 2010). Our data suggest that the relationship between feature generation and rank order, while somewhat ordinal, is highly variable across individuals. This variability was better explained when the model included our measure of complexity and its interactions with feature generation order. Varying cognitive complexity of the feature space offers one possible mechanism for the observed variability in preference. While others have investigated personality trait-level cognitive complexity and economic decision making (Tan & Dolich, 1981; Rejikumar, Asokan-Ajitha, Dinesh, & Jose, 2022), this study is, to our knowledge, the first demonstration of how task-dependent cognitive complexity may impact the psychological underpinnings of preference and choice.

Our measure of complexity also offered another advantage: the ability to use automated, natural language processing methods to analyze a sparse dataset. The small number of words (both within-trial and within-participant) as well as its lack of syntactic and semantic structure pose significant challenges for measuring the semantic characteristics of the participant-generated features using commonly-used natural language processing approaches (e.g., LIWC (Pennebaker et al., 2015)). Our novel approach to cognitive complexity largely overcomes this barrier, providing a new tool for analyzing decision making and other types of data efficiently and at scale.

Future directions We are extending these initial results with further development of the cognitive complexity modeling approach. One challenge of our approach is in the coverage of the reference corpus, since it is necessarily limited to what the vision and language model can generate. The model's output is not only dependent on its architecture but also on the data used to train the model. Thus, our machine-generated reference corpus may not contain words that can be understood without additional context (i.e., are not complex), resulting in high complexity scores for our participant-generated features. Indeed, although our model generally performs well, more features have near ceiling complexity than we expected. Future work should address limitations in the reference corpus. We are addressing these limitations in several ways. First, we are testing ways to increase the scope of the reference corpus. Conversely, we are also testing a model architecture designed to be less reliant on the properties of the reference corpus that uses text-image joint embedding (e.g., CLIP; (Radford et al., 2021)) instead of the current vision and language model, OFA. Such models tend to be less reliant on the specific populations that provided training data and more generalizable across image types (Radford et al., 2021).

We are also testing other possible extensions of the cognitive complexity model, including psychological priors on image processing and mental representation. Attentional and perceptual biases such as visual saliency (Itti & Koch, 2001) may affect feature generation. Specific to our Preference Elicitation Task, familiarity with individual consumer products may drive the complexity of feature representation (Conover, 1982; Sakamoto, 1991).

Future work should also test the utility of our complexity modeling approach to other contexts. Here, we focused narrowly on preference for a small set of consumer goods, as realized through mental representation and semantic description of visual stimuli. (Our work-in-progress is dramatically expanding both the size of the stimulus space and the sample size.) However, our approach is agnostic to the content of the visual stimuli and can provide an additional analytical tool for investigating the mental representation of other types of images, in other contexts. Further, our approach may be useful in other settings whether humans produce sparse text responses that cannot be reliably analyzed by most natural language processing approaches. Such tools will be increasingly useful as the community of experimental behavioral scientists embraces more naturalistic research methods (e.g., ecological momentary assessment, observed social interactions online) that aim to interfere as little as possible with psychological processes (Kingstone, Smilek, & Eastwood, 2008; Osborne-Crowley, 2020; Shamay-Tsoory & Mendelsohn, 2019).

References

Amazon Mechanical Turk. (n.d.). Amazon mechanical turk [Computer software manual].

Bar-Hillel, M., Peer, E., & Acquisti, A. (2014). “heads or tails?”—a reachability bias in binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1656.

Chang, C.-T. (2008). To donate or not to donate? product characteristics and framing effects of cause-related marketing on consumer purchase behavior. *Psychology & Marketing*, *25*(12), 1089–1110.

Chen, X., Fang, H., Lin, T., Vedantam, R., Gupta, S., Dollar, P., & Zitnick, L. C. (2015). *Microsoft coco captions: Data collection and evaluation server.* arXiv. Retrieved from `https://arxiv.org/abs/1504.00325` doi: 10.48550/ARXIV.1504.00325

Conover, J. N. (1982). Familiarity and the structure of product knowledge. *ACR North American Advances*.

Craik, F. I. (2002). Levels of processing: Past, present... and future? *Memory*, *10*(5-6), 305–318.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671–684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, *54*(3), 180.

Garnier, Simon, Ross, Noam, Rudis, Robert, ... Cédric (2021). viridis - colorblind-friendly color maps for r [Computer software manual]. Retrieved from `https://sjmgarnier.github.io/viridis/` (R package version 0.6.2) doi: 10.5281/zenodo.4679424

Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., & Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering. In *Conference on computer vision and pattern recognition (cvpr).*

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), 194–203.

Kazemzadeh, S., Ordonez, V., Matten, M., & Berg, T. (2014). Referitgame: Referring to objects in photographs of natural scenes. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp).*

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of marketing*, *57*(1), 1–22.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340.

Lohman, S., Tomanek, K., Ziegler, J., & Hahn, U. (2010). Getting at the cognitive complexity of linguistic metadata annotation: A pilot study using eye-tracking. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 32).

Lüdecke, D. (2022). sjplot: Data visualization for statistics in social science [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=sjPlot` (R package version 2.8.12)

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, *64*(5), 482.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & cognition*, *28*(6), 887–899.

Norman, D. (2002, 07). Emotion & design: Attractive things work better. *Interactions Magazine*, *9*, 36-42. doi: 10.1145/543434.543435

Osborne-Crowley, K. (2020). Social cognition in the real world: reconnecting the study of social cognition with social reality. *Review of General Psychology*, *24*(2), 144–158.

Partos, T. R., Cropper, S. J., & Rawlings, D. (2016). You don’t see what i see: Individual differences in the perception of meaning from visual stimuli. *PloS one*, *11*(3), e0150615.

Pennebaker, J., Booth, R., Boyd, R., & Francis, M. (2015). Linguistic inquiry and word count : liwc2015. [Computer software manual]. Austin, TX. Retrieved from `www.liwc.net`

Plott, C. R. (1996). Rational individual behaviour in markets and social choice processes: The discovered preference hypothesis. In (p. 225 - 250).

Posit team. (2022). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from `http://www.posit.co/`

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021, 18–24 Jul). Learning transferable visual models from natural language supervision. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 8748–8763). PMLR. Retrieved from `https://proceedings.mlr.press/v139/radford21a.html`

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, *9*(7), 545–556.

Rejikumar, G., Asokan-Ajitha, A., Dinesh, S., & Jose, A. (2022). The role of cognitive complexity and risk aversion in online herd behavior. *Electronic Commerce Research*, *22*(2), 585–621.

Rizopoulos, D. (2022). Glmmadaptive: Generalized linear mixed models using adaptive gaussian quadrature [Computer software manual]. (https://drizopoulos.github.io/GLMMadaptive/, https://github.com/drizopoulos/GLMMadaptive)

Sakamoto, A. (1991). Familiarity, favorability, and cognitive complexity: integration of frequency of interaction hypothesis and vigilance hypothesis. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *61*(6), 392–399.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspectives on Psychological Science*, *14*(5), 841–859.

Slovic, P. (1995). The construction of preference. *American psychologist*, *50*(5), 364.

Sumner, E., DeAngelis, E., Hyatt, M., Goodman, N. D., & Kidd, C. (2015). Toddlers always get the last word: Recency biases in early verbal behavior. In *Cogsci.*

Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Lawrence Erlbaum Associates.

Tan, C. T., & Dolich, I. J. (1981). The moderating effects of cognitive complexity and prior product familiarity on the predictive ability of selected multi-attribute choice models for three consumer products. *ACR North American Advances*.

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review*, *10*(2), 116–124.

Tversky, A., & Kahneman, D. (1985). *The framing of decisions and the psychology of choice*. Springer.

Wang, P., Yang, A., Men, R., Lin, J., Bai, S., Li, Z., . . . Yang, H. (2022). Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. In *Proceedings of the 39th international conference on machine learning (icml).*

Weber, E. U., & Johnson, E. J. (2006). Constructing preferences from memory. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (p. 397–410). Cambridge University Press. doi: 10.1017/CBO9780511618031.022

Wickham, H. (2011a). ggplot2. *Wiley interdisciplinary reviews: computational statistics*, *3*(2), 180–185.

Wickham, H. (2011b). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from `https://www.jstatsoft.org/v40/i01/`

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi: 10.21105/joss.01686

Yu, L., Poirson, P., Yang, S., Berg, A. C., & Berg, T. L. (2016). Modeling context in referring expressions. In *Computer vision – eccv 2016.*