Seeing and Speaking: How Verbal 'Description Length' Encodes Visual Complexity

Zekun Sun & Chaz Firestone

Johns Hopkins University

Version: 3/4/21 — in press, *JEP:General*

Word Count: 8644

Address for correspondence

Zekun Sun (*zekun@jhu.edu*) or Chaz Firestone (*chaz@jhu.edu*)

Department of Psychological & Brain Sciences

Johns Hopkins University

3400 N Charles St.

Baltimore, MD 21218

Abstract

What is the relationship between complexity in the world and complexity in the mind? Intuitively, increasingly complex objects and events should give rise to increasingly complex mental representations (or perhaps a plateau in complexity after a certain point). However, a counterintuitive possibility with roots in information theory is an inverted U-shaped relationship between the "objective" complexity of some stimulus and the complexity of its mental representation, because excessively complex patterns might be characterized by surprisingly short computational descriptions (e.g., if they are represented as having been generated "randomly"). Here, we demonstrate that this is the case, using a novel approach that takes the notion of "description" *literally.* Subjects saw static and dynamic visual stimuli whose objective complexity could be carefully manipulated, and they described these stimuli in their own words by giving freeform spoken descriptions of them. Across three experiments totaling over 10,000 speech clips, spoken descriptions of shapes (Experiment 1), dot-arrays (Experiment 2), and dynamic motion-paths (Experiment 3) revealed a striking quadratic relationship between the raw complexity of these stimuli and the length of their spoken descriptions. In other words, the simplest and most complex stimuli received the shortest descriptions, while those stimuli with a "medium" degree of complexity received the longest descriptions. Follow-up analyses explored the particular words used by subjects, allowing us to further explore how such stimuli were represented. We suggest that the mind engages in a kind of lossy compression for overly complex stimuli, and we discuss the utility of such freeform responses for exploring foundational questions about mental representation.

Keywords: complexity; representation; language; information theory

1 Introduction

Objects and events frequently strike us as being *simple* or *complex*. A painting, for example, may be plain and minimalistic, or it may be elaborate and embellished; a highway system may be sparse and direct, or it may be dense and interconnected; even a story may be linear and straightforward, or it may twist and branch.

Although complexity can be found across the natural, artificial, and social world, the most pervasive and striking impressions of complexity are surely those that arise from *visual images*. For example, consider the pairs of images in Figure 1. Visual stimuli as diverse as flowers, abstract shapes, digital arrays, and patterns of motion all give rise to powerful impressions of complexity: For each pair, the image on the right is immediately perceived as more complex than the image on the left.

What is the nature of this experience, and how can we investigate it scientifically? Though mental representations of complexity have been of great interest both historically and recently (Attneave, 1957; Berlyne, 1958, 1970; Berlyne & Peckham, 1966; Biederman, 1987; Chater, 1996; Chipman & Mendelson, 1979; Cutting & Garvin, 1987; Feldman, 2016; Forsythe et al., 2011; Lewis & Frank, 2016; Madan et al., 2018; Oliva et al., 2004; Rosenholtz et al., 2007; Snodgrass & Vanderwart, 1980; Spehar et al., 2003; van der Helm, 2000; Watson, 2011; Wilder et al., 2016), all attempts to study psychological representations of complexity confront at least two major challenges: (1) the extreme variability in the kinds of stimuli that give rise to experiences of complexity (as shown by the diversity of images in Figure 1), and (2) the near-ineffable nature of the experiences themselves. Indeed, because of how difficult it can be to craft the right kinds of questions or measures to probe experiences of complexity, researchers have either supplied subjects with long and specific definitions of complexity to be used in rating scales (Berlyne & Peckham, 1966; Madan et al., 2018; Oliva et al., 2004; Snodgrass & Vanderwart, 1980), or they have developed measures that get at such experiences only very indirectly, such as asking for the name someone would give to an object, probing related notions such as its aesthetic appeal, or measuring the subject's ability to detect the object in noise (Alvarez & Cavanagh,

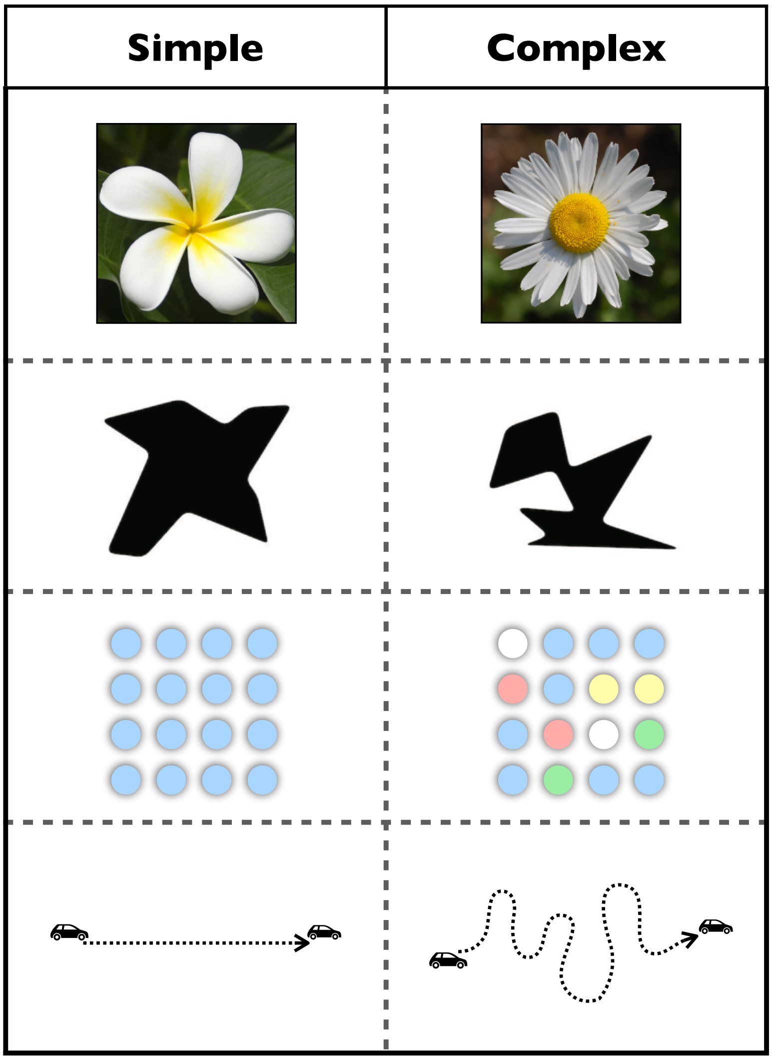

Figure 1: Examples of images that evoke impressions of *complexity*. The experience of complexity can span an extremely wide range of stimuli, including images that are natural, artificial, digital, and dynamic.

2004; Biederman, 1987; Lewis & Frank, 2016; Spehar et al., 2003; Wilder et al., 2016). Some researchers have even thought that such experiences are impossible to study in any unified way, and that instead the best we can hope to do is determine complexity according to the particular sociological and psychological context that a subject finds herself in (Simon, 1972).

Here, we explore a new and different experimental approach to both of these challenges, and use this approach to tackle a longstanding question about the psychological experience of complexity: What is the relationship between "objective" complexity in the world and the complexity of one's "subjective" internal representations?

1.1 Complexity as "description length"

A long tradition in computer science, information theory, and psychology conceives the complexity of a stimulus (e.g., an image, a sequence of numbers, a coded message, etc.) in terms of the length of its shortest description (Chaitin, 1969, 1975; Feldman, 2016; Feldman & Singh, 2006; Kolmogorov, 1965; Leeuwenberg, 1969; Shannon, 1948; van der Helm et al., 1992; Wallace, 2005). A natural and popular way to think of such descriptions is to imagine the computer program that one would write to reproduce some stimulus (as in the case of Kolmogorov complexity; Kolmogorov, 1965). For example, suppose you were asked to infer the computer program that produced the outputs shown in Figure 2. How would you encode an output string such as the 12 digits in "`1,1,1,1,1,1,1,1,1,1,1,1`"? One program that might produce this string would be approximately as long as the string itself — for example, a command like "`print (1,1,1,1,1,1,1,1,1,1,1,1)`". But a shorter program would "compress" this output — perhaps something more like "`print [1]*12`" (which could scale up to much longer strings; for example, a program to produce *a thousand* 1's needn't be much longer: "`print [1]*1000`"). A key feature of this notion of complexity is that increasingly complex strings are described by increasingly long programs: For example, the slightly more complex 12-character string "`0,1,0,1,0,1,0,1,0,1,0,1`" might be produced by the slightly longer program "`print [0,1]*6`"; and the even more complex 12-character string "`0,1,0,0,1,1,0,0,0,1,1,1`" would require an even longer program — perhaps "`print [([0]*N,[1]*N) for N in range(3)]`". The minimal-description-length conception of complexity thus offers a precise and rigorous way to capture an intuition we may have about the nature of complexity, which is that more complex stimuli are "less compressible" and so harder to capture in short descriptions.

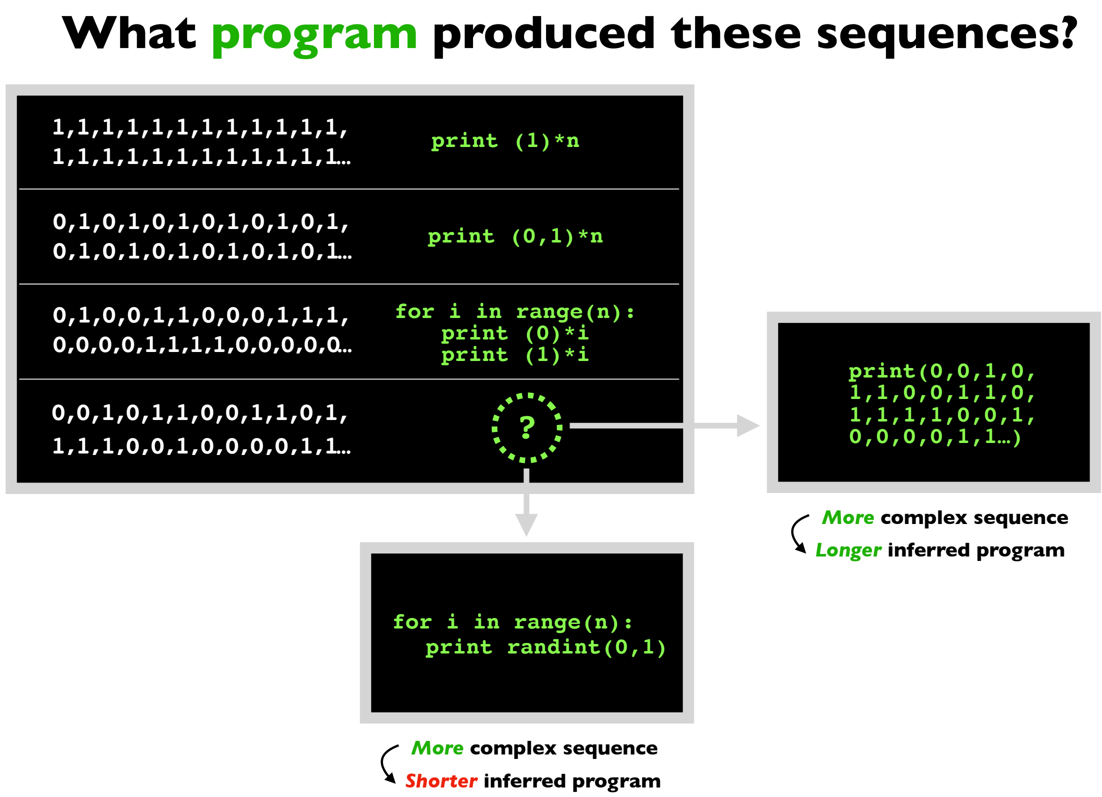

Figure 2: The relationship between complexity and description length might be best understood by analogy. Consider each of the binary strings appearing above (white), and then ask yourself what sort of computer program might have produced them (green). At first, increasingly complex outputs seem likely to have been generated by increasingly long programs. But when the output becomes so complex as to seem *random* (last row), it might best be summarized by a *shorter* program (such as one using a random number generator), instead of printing digits individually. Might a similar pattern hold for mental representations of complexity? (Programs here are written in Python pseudocode.)

Crucially, however, this same framework faces a puzzle when it comes to extremely complex stimuli. For example, consider the final output string in Figure 2; what is the minimal program for producing *that* string? That string seems — and is — highly *in*compressible, and so the minimal program for producing it may genuinely be the one that simply reiterates the string itself (i.e., a very long "`print`" statement). But another possibility is that this string is best described as being *randomly* generated; perhaps the program that created it simply chose random 1's and 0's, and the particular sequence that resulted — i.e., *this* sequence of random 1's and 0's rather than some other, equally random, sequence — isn't what's most essential. On one hand, this approach might be described as a kind of "lossy" compression, in that information about the string is inevitably lost by a program that simply treats the output as random. But on the other hand, this may well be the most fitting description of the sequence. In other words, it may actually be reasonable to infer that the program that produced this output really did just pick 1's and 0's at random.

This pattern raises an intriguing psychological question: If *we* represent complexity (as exemplified in Figure 1 and as explored in the relevant literatures; Attneave, 1957; Berlyne, 1958, 1970; Chater, 1996; Chipman and Mendelson, 1979; Cutting and Garvin, 1987; Feldman, 2016; Forsythe et al., 2011; Lewis and Frank, 2016; Madan et al., 2018; Rosenholtz et al., 2007; van der Helm, 2000; Watson, 2011), how do our minds confront this representational challenge? Do increasingly complex stimuli (as characterized by some objective metric) give rise to increasingly long internal descriptions? Or might there be an inflection point, such that complexity beyond a certain threshold is somehow represented using *fewer* resources or *shorter* internal descriptions?

Indeed, it has long been speculated that the "objective" complexity of a visual stimulus might not predict its corresponding subjective complexity in a monotonic fashion (Donderi, 2006; Wolfram, 2002); but this has been difficult to demonstrate experimentally, in part because of the challenges mentioned earlier. For example, in the most comprehensive review of recent work on visual complexity, Donderi considered the extreme case of *random* images, and concluded as follows:

> "We cannot detect the structure in a random image. This suggests that the subjective complexity of an image will be at a maximum somewhere between simple order and complete randomness. [...] From this it follows that subjective complexity might be an inverted U-function, rather than a monotonic function, of algorithmic complexity. There is at present no objective evidence about this." (Donderi, 2006)

Recent work does offer clues that this might be the case. For example, it has recently been shown that children preferentially allocate attention to visual and auditory events that are neither too predictable nor too surprising, with a preference instead for "Goldilocks" stimuli with moderate complexity (Kidd et al., 2012, 2014). Other work suggests that Prägnanz or perceptual "goodness" should be most likely to arise neither for uniform patterns (with zero entropy) nor for maximally noisy patterns (with very high entropy), but instead again for moderately entropic patterns (Koenderink et al., 2018). Still, this work focuses only on very specific types of stimuli, and does not necessarily aim to study the encoding of complexity *per se*, in the manner of the minimal program example explored above. Could a completely different kind of evidence make progress on this question?

1.2 The present experiments: Taking "description length" *literally*

Here, we take a different approach, by introducing a new measure of subjective complexity based in natural language. In addition to its communicative function, language is often conceived as mirroring human mental processes and reflecting the character of thought, in part because of its status as a generative process that uses a combinatorial system of rules to generate an unbounded range of expressions (Chomsky, 1957, 1968). In light of this, our approach is to explore internal representations of complexity by taking the notion of "description length" *literally*. We ask subjects simply to describe, in their own words, a variety of stimuli whose complexity — and in particular, with "compressibility" could be systematically varied. We then examine the length of the "verbal programs" that subjects freely generate, taking the length of such descriptions as proxies for the complexity of the relevant internal representations.*

To achieve this, we created a pipeline to collect auditory recordings of subjects verbally describing static and dynamic visual stimuli using the microphones of their home computers, and then transcribed these recordings into text. We hypothesized that the length of such descriptions (e.g., the number of words appearing in a transcript of a given audio description) would be related to the complexity of the stimulus being described (as determined by an independent measure) — but not linearly or monotonically. Instead, we asked whether subjects' free descriptions might confirm the speculation that *moderately* complex stimuli have the longest description lengths, resulting in a quadratic relationship between the objective complexity of a stimulus and the length of its corresponding verbal description. Across three experiments spanning several types of visual stimuli, we collected over 10,000 independent audio recordings, and discovered that this is indeed the case — with an inverted-U-shaped function relating objective complexity to verbal description length.

2 Audio collection and processing pipeline (for all experiments)

All of the experiments reported here involved the collection of freely-spoken descriptions of visual stimuli. We adopted a standardized procedure to establish a pipeline for gathering and processing these descriptions. All subjects were recruited online via Amazon Mechanical Turk. (For a discussion about the reliability of this subject pool, see Crump et al., 2013.) Prior to the experiment, all participants consented to have their voice recorded. (Such voice recording, as well as every other experimental procedure described below, was approved by the Institutional Review Board of Johns

*See Kahneman and Tversky, 1972, for an early articulation of this general hypothesis and approach, though with a prediction that is in some ways similar to ours and in other ways opposite: "Random-appearing sequences are those whose verbal description is longest" (p.436).

Hopkins University.) To ensure that the recordings would be easily processed and transcribed, we required that all participants be native speakers of English, and we instructed them to take any necessary steps to minimize external noise from interfering with the audio quality (e.g., background music, loud street sounds, conversations by other people, etc.). Subjects then completed a short test to ensure that their home setup could indeed record and upload audio; only if they successfully uploaded a test audio recording to our server were participants able to access the next stage of the experiment.

An experimental trial proceeded as follows. A stimulus (e.g., a shape, an array of dots, or a moving object) was shown on the display, and subjects were told to "just describe what you see on the screen". When subjects felt ready, they pressed a "Record" button and spoke a description into their microphone. The interface allowed subjects to start, end, replay, re-record and upload their audio descriptions however they pleased. Once they were satisfied with the description they gave, they pressed an "Upload" button to transfer their audio file to an internal server. The interface allowed subjects to record as long a description as they liked, but subjects were also given some very gentle (but unenforced) guidelines: "Your description can be as long or as short as you'd like, from just a few seconds up to a whole minute if you feel that is necessary. (No need to go overboard though: If you're spending more than a minute on one shape, that's probably longer than you need to.)"

To transcribe the audio recordings, we used the Google Speech Recognition API from the Python library "SpeechRecognition" (Version 3.8) (Zhang, 2017, https://github.com/Uberi/speech_recognition). To validate the accuracy of this tool, we first ran a short pilot study using the procedure described above; the audio clips from this pilot were then transcribed both by the API and an experimenter (author Z.S.). A comparison of these two transcriptions revealed a high degree of overlap: The experimenter's transcription and the API's transcription agreed about the number of words in a description to a degree of 95% (i.e., the average *disagreement* in number of words for a given file was only 4.7%), and they agreed about the identity of the words themselves over 90% of the time. (The discrepancies that arose often involved

Figure 3: Audio collection and transcription pipeline for all of the present experiments. Participants viewed stimuli on their home computer and described them in their own words. They then uploaded the audio file of their description to a Web server, and the file was transcribed by speech-to-text software.

homophones, such as a subject saying "U shape" and the API transcribing this as "you shape".) Given that our primary interest was in the *length* of these descriptions — and, in particular, their *relative* length — this degree of overlap encouraged us that the transcription tool was sufficiently reliable for analyzing thousands of such speech clips.

Pre-registrations: For each experiment, we pre-registered the sample size, experimental design (including the general nature of the stimuli, as well as the number of trials, and the experimental procedure), as well as the main analysis (and occasionally some secondary analyses); all experiments below follow these pre-registered details. Importantly, however, we intentionally chose *not* to pre-register a prediction about the particular way in which complexity and description length (i.e., number of words) would be related; with little precedent for this kind of experiment, we wanted to be open to multiple possibilities, and so our pre-registered hypothesis was simply that complexity would be related to description length (without assuming the specific nature of that relationship; for example, in Experiment 1, we pre-registered both a linear and a quadratic model). For data analyses, we pre-registered the primary variables of interest (e.g., in Experiment 1, skeletal complexity of a shape as an independent variable, and number of words as a dependent variable), exclusion criteria, and a regression examining the relationship between stimulus complexity and the average number of words that stimulus would receive. We report all of these pre-registered analyses below; however, we also report several variations, including additional analyses that occurred to us after examining the data, as well as improvements to our analyses suggested by reviewers (such as using mixed-effects models that include random-effects terms for subjects and items). Where relevant, we indicate below which results reflect pre-registered analyses and which are exploratory.

An archive of the data, experiment code, stimuli, experiment pre-registrations, analysis code, and other relevant materials is available at https://osf.io/a57gy/.

3 Experiment 1: Shape

Are increasingly complex stimuli mirrored by increasingly complex internal representations? Or might such factors be related non-monotonically, or even quadratically? Our first experiment explored this question using freely spoken verbal descriptions of shapes. Geometric shapes have a long history as a stimulus class in studies of perceived complexity (Attneave, 1957), and they are appropriate here too for several reasons. First, it is striking just how readily such shapes can evoke powerful (and variable) impressions of complexity: Simply drawing one kind of blob on a page rather than another kind of blob causes one to have a strong impression of its complexity that is easily compared across stimuli; for example, you can easily see that the three shapes in the top row of Figure 4 are changing in their complexity. Second, recent advances in computational geometry allow for a succinct and standardized measure of a shape's "objective" complexity, based on their internal "skeletal" structure (see below), allowing us to vary the objective complexity of a shape in a systematic way. Third, such shapes tend to have few (if any) previous associations: Whereas familiar objects such as tools, animals, or paintings may also vary in their complexity, it can be difficult to disentangle the different factors that contribute to their complexity (which may encompass not only how they *look* but also knowledge of their function, origin, social significance, etc.). These three factors make geometric shapes ideal candidates for isolating, measuring, and manipulating complexity per se, and so we used them here in our first experiment.

3.1 Method

3.1.1 Subjects

As stated in our pre-registration, we aimed for 240 usable subjects for this experiment; to reach this target sample, we recruited 260 subjects total (since we expected a small number of subjects to drop out of the study or upload untranscribable audio files).

3.1.2 Stimuli

We adopted an algorithmic procedure to generate a variety of novel shapes that varied in their complexity. For each shape, we first defined the number of sides that the object would ultimately have, and then created a set of randomly located points that would serve as the vertices of the shape's edges. We then connected these points using the method of Delaunay Triangulation, which maximizes the minimum angles formed by the overall structure of the connected lines. Next, facets along the boundary of this triangle mesh were removed until the resulting polygon had the predefined number of sides. Finally, for each shape, the edges of the resulting polygon were smoothed in order appear more natural.

160 objects were generated using the procedure mentioned above, ranging from 3-sided objects to 34-sided objects, with 5 instances of each N-sided object (Figure 4). Though the nature of online experiments prevents us from knowing the exact visual properties of our stimuli as they appeared to subjects, each shape was approximately 500 pixels wide.

To calculate the "objective" complexity of each shape, we first derived its medial axis skeleton, classically defined as the set of points having two or more closest points on the shape's perimeter (Blum, 1973). The skeleton of a shape might be thought of as its "blueprint": the internal structure that explains why a shape has the external features it does. Shape skeletons capture many aspects of a shape's large- and small-scale organization, including not only the number of "parts" a shape has (Siddiqi et al., 1996) but also how these parts are configured with respect to one another, and even to some extent the complexity *of* each part itself (since the shape of a skeletal branch captures the shape of its corresponding part). Shape skeletons are also psychologically plausible, with growing empirical evidence that they are computed and represented by human vision (Ayzenberg et al., 2019; Ayzenberg & Lourenco, 2019; Firestone & Scholl, 2014; Lowet et al., 2018; Sun & Firestone, 2021; Wilder et al., 2011). To turn this representation into a measure of complexity, we followed previous work (including Sun and Firestone, 2021) in computing the integral of the turning angle along each skeletal branch, summed over the total number of branches (using ShapeToolBox1.0; Feldman and Singh, 2006). An intuitive way to capture this measure might be to imagine a person walking along the skeleton of a shape; the more often this person changes direction (such that their next step was not easily predictable from their previous step), the greater the complexity of their "walk", and so the greater the complexity of the shape itself. What results is a measure of the "amount of information" required to describe the shape in terms of its basic underlying structure.

3.1.3 Procedure

Each experimental session consisted of 20 trials, following the data collection pipeline described above. On each trial, participants saw a novel shape and were asked to describe it. (For this experiment, we added the suggestion that the subjects' description should be sufficient to allow someone to *draw* this shape.) To ensure that subjects saw a wide range of complex shapes, the 20 stimuli for a given subject were selected from the broader pool of 160 shapes pseudorandomly, with equal spacing throughout the range of complex shapes as ranked by our objective measure of shape complexity. The shapes were shown in a newly generated random order for each subject.

As stated in our pre-registration, we excluded any subject whose average spoken description length was less than 5 words long (since we suspected that such subjects were not very engaged in the task), or whose entire set of audio files could not be recognized by the speech recognition software. We also excluded any subjects who did not complete the HIT. At the level of individual trials, we excluded any audio file whose transcribed description was only 1 word long, or any files that couldn't be

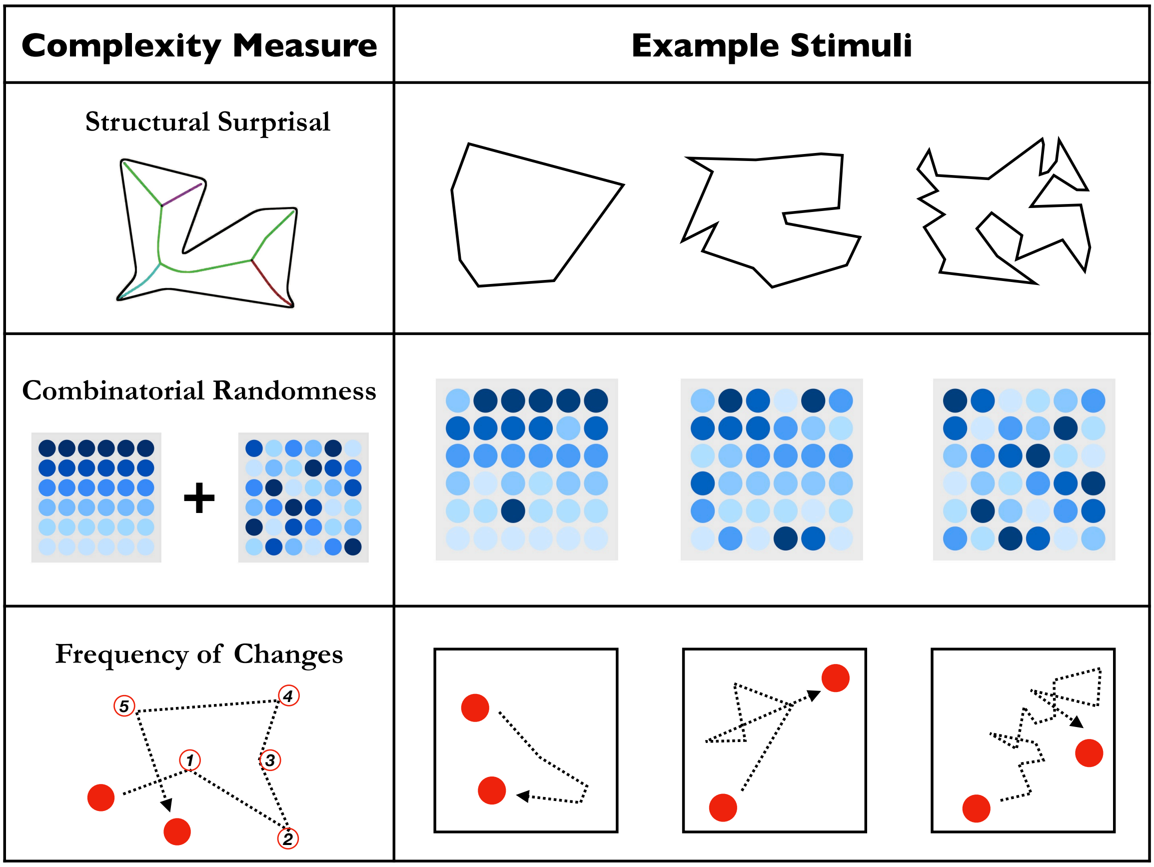

Figure 4: Sample stimuli and complexity measures from the three experiments reported here. Experimet 1: A sample shape with its medial-axis skeleton inset. Experiment 2: Matrices of dots, ranked according to how "organized" or "disorganized" they appear. Experiment 3: Paths of a moving dot, ranked according to the number of changes in the dot's trajectory.

read or transcribed by the speech recognizer. Very occasionally, we failed to obtain a complete dataset from a subject who otherwise completed the entire experimental session (e.g., they completed all 20 trials, but successfully uploaded only 18 audio files to our server); when that happened, we chose to keep their data for the purposes of our analyses.

3.1.4 Analysis

Our primary interest here is in *the number of words* subjects used to describe the shapes (though see Section 6 below for follow-up analyses that examine the *content* of such descriptions). To this end, we averaged the "verbal description length" of all the descriptions provided for each of the 160 shapes. We then modeled the relationship between average verbal description length and the "objective" description length of the shape's skeletons (as described above), using a 2nd-degree polynomial regression.

Given that we were interested in *relative* differences in this quantity, we normalized the average length of each participant's descriptions to the grand mean of all descriptions; in other words, the particular tendency of a given subject to give very long or very short descriptions (relative to the group) was factored out from any subsequent modeling, by rescaling that subject's average description length to the average description length of the cohort as a whole. (However, no result we report below depends in any way on this normalization procedure; in other words, all trends and effects remain statistically significant, in the same direction, even without normalizing description lengths.)

3.2 Results and discussion

9 subjects were excluded for giving extremely short descriptions or uploading low-quality audio files, leaving 251 subjects with analyzable data. From these subjects, 4897 audio clips were successfully uploaded to the server (indicating a rate of successful uploading of 97.5%). Of these files, 546 (11%) could not be processed as audio input by the speech recognition software, and 400 (8.2%) could be processed

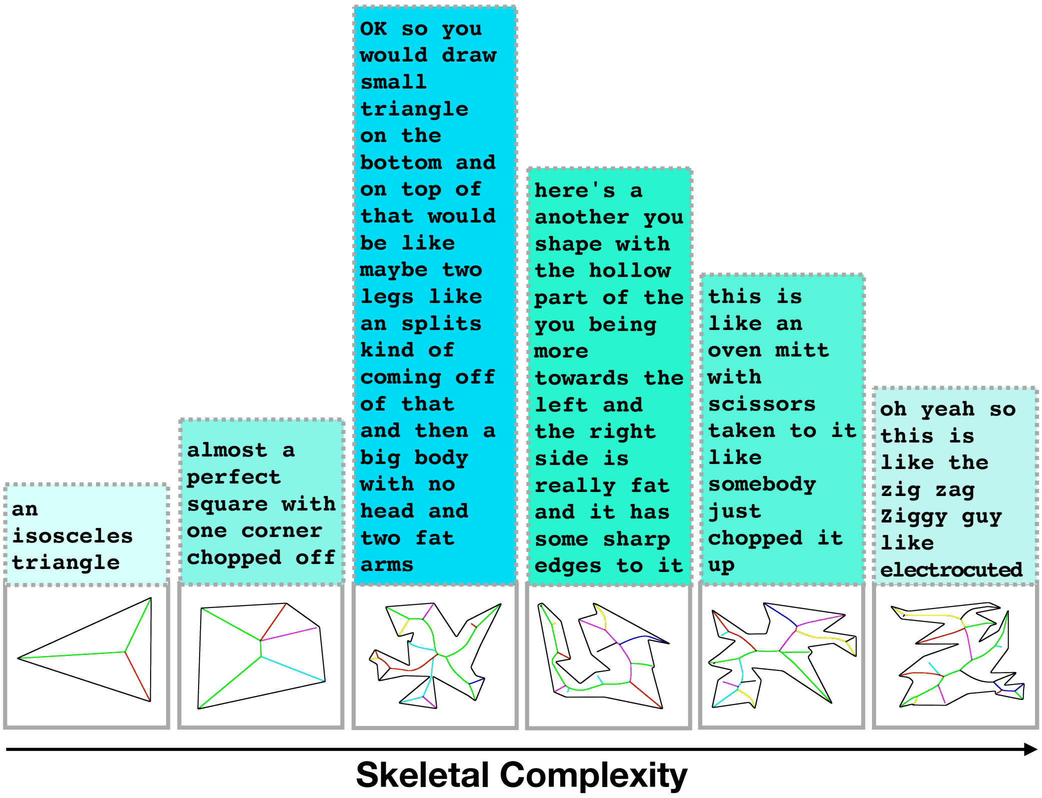

Figure 5: Sample descriptions from a participant in Experiment 1. Six objects from the stimulus pool are shown here, in order of skeletal complexity, with their medial-axis skeletons inset (in the actual experiment, these objects appeared as outlined shapes, without their skeletons, and in a random order). Each bar show the transcript of the description that the participant spoke. (The few transcribing errors shown here — e.g., "you shape" instead of "U shape" — are representative of the broader sample, and so are left intact here for reference.) You can hear these spoken descriptions for yourself at https://perceptionresearch.org/speaking.

as input but not transcribed; finally, another 132 (2.7%) were excluded for being only one word long. This left 3818 audio files included in our final analysis (78% of all uploaded files), resulting in approximately 24 unique descriptions for each of the 160 shapes. The average duration of these included speech files was 23 seconds (MIN=2, MAX=62, SD=14), and the average description length was 41 words (MIN=2, MAX=215, SD=31).

Figure 5 shows sample descriptions from six trials of one participant. As you can see, this subject gives short descriptions for simple shapes, and then increases the length of his descriptions as shape complexity increases. Crucially, however, this pattern breaks past a certain point: Once the shapes become excessively complex, the participant's descriptions *shorten* again, producing a distinct inverted-U relating description length to objective complexity.

Indeed, this pattern was evident across the subject pool as a whole. Figure 6 (Experiment 1) shows the average description length of all 160 shapes and all 3818 audio files. As may be clear from visual inspection alone, description length does not increase monotonically with objective complexity; instead, it plateaus at a moderate-to-high degree of complexity, before dipping down again for the most complex objects. Indeed, when the data are fit to a 2nd-degree polynomial, this analysis reveals the expected inverted-U-shaped relationship between objective complexity and the length of verbal descriptions ($R^2 = 0.41$, $F(2, 157) = 55.36$, $p = 2.20 \times 10^{-16}$)[†]. Crucially for our purposes, the coefficient of the squared term was significantly *negative* (and the addition of the quadratic term substantially improved variance explained over and above the simple linear regression model [$b = -33.59$, $95\%CI = [-45.27, -21.91]$, $t = -5.68$, $p = 6.36 \times 10^{-8}$]) — whereas it might have been (undetectably different from) zero if the pattern had been linear, or even positive if there were some other pattern. This pattern suggests a "Goldilocks" relationship between an object's complexity and the length of its corresponding description: moderately complex objects yielded the longest verbal descriptions, while less or more complex objects yielded shorter descriptions.

Due to there being multiple observations per subject, we also fit the data to a mixed-effects model that included subject and shape as random effects; this exploratory analysis also revealed a significant quadratic term for the relationship between complexity and number of words ($b = -141.34$, $95\%CI = [-190.29, -92.19]$, $t = -5.61$, $p < 1.13 \times 10^{-7}$))[‡].

This broader pattern may also be seen in a "binned" version of the data. To further examine how description length changes as the complexity enhances, we grouped the

[†]We also pre-registered a simple linear model in addition to the quadratic model. Here, the linear model explained a significant proportion of variance ($R^2 = 0.29$, $F(1, 158) = 65.5$, $p = 1.46 \times 10^{-13}$).

[‡]We thank a reviewer for suggesting this analysis.

stimuli into five levels based on skeletal complexity (Figure 6, second row), with each level containing 32 objects. An one-way analysis of variance (ANOVA) reveals a significant effect of objective complexity ($F(1, 158) = 53.3$, $p = 1.32 \times 10^{-16}$, $\eta^2 = 0.25$), and the follow-up trend analysis showed that the quadratic component is highly significant ($b = -34.79$, $95\% CI = [-45.27, -21.91]$, $t = -4.83$, $p = 3.22 \times 10^{-6}$).

These results provide initial evidence of an inverted-U-shaped relationship between the complexity of a stimulus and the length of its corresponding description. In other words, instead of generating longer and longer "verbal programs" to encode more and more complex stimuli, participants generated the longest expressions for moderately complex stimuli, apparently willing to "lose" information about such stimuli (perhaps as in the bottom panel of Figure 2).

4 Experiment 2: Organization

The previous experiment found that the length of verbal descriptions that subjects generated for simple and complex novel shapes was captured by a non-linear (and perhaps even quadratic) function of the shapes' objective complexity. However, it remains unclear whether this pattern truly shows a *decline* in description length past a certain complexity level (rather than, e.g., a plateau or leveling off). Moreover, Experiment 1 explored only one type of stimulus, which varied in only one type of way — in particular, the addition of information along the skeleton of a shape. By contrast, the experience of complexity spans a strikingly wide range of categories (as in Figure 1); how robust, then, is this pattern? And how far does it generalize?

Experiment 2 repeated the design of Experiment 1 with a brand new set (and kind) of stimuli — arrays of colored dots. Unlike novel shapes whose complexity varies as a function of their internal structure, the stimuli we used in this experiment were composed of nearly identical elements (36 blueish dots), such that the impression of their complexity derived from their *grouping* or *organization*. Grouping describes the impression that certain elements "go together", and is a pervasive experience that, at least in principle, holds over any kind of segmented stimulus. This process is often thought to exist for precisely the sorts of reasons we explore here: to reduce the amount of information required to encode visual patterns (Leeuwenberg, 1967; van der Helm et al., 1992; Wagemans et al., 2012) — since a highly organized or regular pattern is more easily compressed than a disorganized, irregular, or random pattern.

Experiment 2 thus used stimuli that varied in exactly this way: patterns of dots whose arrangement varied from highly organized to highly random. Though complexity here was determined by a rather different process (i.e., the addition of randomness, as opposed to the addition of structure), what unites both of these classes of stimuli is their varying "compressibility"; shapes with more branches, and arrays with more randomness, are both less compressible, which is the notion of complexity we explore here. This approach thus allowed us to ask whether the pattern from Experiment 1 would generalize to a new kind of stimulus, and also provided an opportunity to discover a more robust inverted-U-shaped trend.

4.1 Method

4.1.1 Subjects

As stated in our pre-registration, we aimed for 300 usable subjects for this experiment; to reach this target sample, we recruited 320 subjects total.

4.1.2 Stimuli and Procedure

To generate the stimuli for this experiment, we adapted a procedure from earlier work for manipulating the degree of perceptual organization in a matrix of dots (Barbot et al., 2018); previous work using similar grid patterns also shows that greater randomness in such matters correlates with greater perceived complexity (Falk & Konold, 1997). We generated a library of 6×6 matrices of dots, appearing in six different shades of blue (Figure 4, second row). Each matrix was created by combining some proportion of a "maximally grouped" matrix (in particular, a matrix where each row [or column] had a uniform color, and each adjacent column [or row] progressed smoothly from one shade to another) and a "maximally ungrouped" matrix (in particular, a matrix where the positions of all the same dots were shuffled randomly, with the additional constraint that no adjacent dots had the same shade). The degree of organization was then determined by biased sampling from the two matrices. For example, a matrix with a "disorganization level" of 80% would have 20% of its dots drawn from the maximally grouped parent matrix, and 80% of its dots drawn from the maximally ungrouped parent matrix, whereas a matrix with a "disorganization level" of 10% would have 90% of its dots drawn from the maximally grouped parent matrix, and 10% of its dots drawn from the maximally ungrouped parent matrix. Using this procedure, we generated 140 matrices total: 20 each from 7 different levels of disorganization or randomness (0%, 10%, 20%, 30%, 50%, 80%, 100%)[†]. Figure 4 shows three examples of these images, at disorganization levels of 10%, 50%, and 100%.

The design was otherwise the same as in Experiment 1, with only a few small changes. Each experimental session consisted of 14 trials (2 matrices from each of the 7 levels, in a random order). Additionally, subjects' instructions were even simpler than in Experiment 1: they were told only to "describe the array" (not to do so in a way that could allow someone else to draw it), giving them maximum freedom to generate their description in whatever way felt most natural and intuitive.

4.2 Results and discussion

6 subjects were excluded for giving extremely short descriptions or uploading low-quality audio files, leaving 314 subjects with analyzable data. From these subjects, 4320 audio clips were successfully uploaded to the server (indicating a rate of successful uploading of 98.3%). Of these files, 51 (1.2%) could not be processed as audio input by the speech recognition software, and 222 (5.1%) could be processed as input but not transcribed; finally, another 56 (1.3%) were excluded for being

[†]Careful readers will note that our description here essentially reverses the meaning of "disorganization level" as described in our pre-registration (where 0% meant highly *random* and 100% meant highly organized). On reflection, we found our pre-registered naming convention to be less intuitive, but in fact the procedure we follow here is identical to the one specified in the pre-registration.

only one word long. This left 3991 audio files included in our final analysis (92.4% of all uploaded files), resulting in approximately 29 unique descriptions for each of the 140 arrays. The average duration of these included speech files was 19 seconds (MIN=3, MAX=62, SD=11), and the average description length was 35 words (MIN=2, MAX=148, SD=22).

As can be seen in Figure 6, a similar inverted-U-shaped pattern emerged again: Highly grouped and highly random matrices were tersely described, while moderately grouped arrays garnered the longest descriptions. As before, a 2nd-degree polynomial model revealed this inverted-U relationship between matrix complexity and description length ($R^2 = 0.58$, $F(2, 137) = 93.06$, $p < 2.20 \times 10^{-16}$), again with a significantly negative squared-complexity term ($b = -13.85$, $95\%CI = [-19.87, -7.83]$, $t = -4.55$, $p < 1.18 \times 10^{-5}$)[§]. Similarly, a mixed-effects model with subject and matrix as random effects revealed this significant quadratic complexity effect ($b = -81.41$, $95\%CI = [-117.77, -45.05]$, $t = -4.83$, $p < 2.28 \times 10^{-5}$).[¶]

This experiment thus generalizes the "Goldilocks" pattern from Experiment 1 to an entirely new kind of stimulus — one whose complexity derives from the organization of otherwise-identical perceptual elements — and shows this inverted-U-shaped pattern even more conclusively (as you may see in Figure 6). [‖]

[§] We pre-registered both a linear model and an exponential model in addition to the quadratic model, to explore whether there would be a reliable decline in description length at high levels of complexity. However, as inspection of the data reveals (Figure 6), the data were not patterned according to a growth exponential model. Instead, we used a least square regression to fit the data to a decay exponential function $y = b_1 e^{b_2 x} + b_3$; none of the coefficients were significant ($b_1 : p = 0.92$; $b_2 : p = 0.92$; $b_3 : p = 0.94$). A linear model ($R^2 = 0.51$, $F(1, 158) = 144.8$, $p < 2.20 \times 10^{-16}$) explained a significant proportion of variance.

[¶] Note that the plot in Figure 6 for this experiment shows a local polynomial regression (also known as a LOESS regression), which is often more appropriate when one variable has a limited range — which was true here in Experiment 2 but not in Experiments 1 or 3 (since no matrix can be less organized than 0% or more organized than 100%, whereas shapes can in principle increase in complexity indefinitely). However, the statistics we report here are for the standard polynomial regression, as in Experiment 1 and as described in our pre-registration.

[‖] Simonsohn 2018 proposes a "two-lines" test to demonstrate U-shaped relationships, to avoid possible false positives that can arise from quadratic regression. This approach sets a 'break point' of the independent variable and then uses two separate linear regressions to test for both an upward trend and a downward trend, requiring two slopes with opposite signs that are both significant. Though our experiments were not designed for this analysis (and we did not pre-register it), we note that the current data showed a fairly consistent pattern even under this new test. For Experiment 1, we set the break point at x = 0.52, and found a significant positive line before 0.52 ($F(1, 94) = 39.12$, $p < 0.0001$) and the significant negative line after 0.25 ($F(1, 64) = 39.12$, $p < 0.05$). A similar pattern

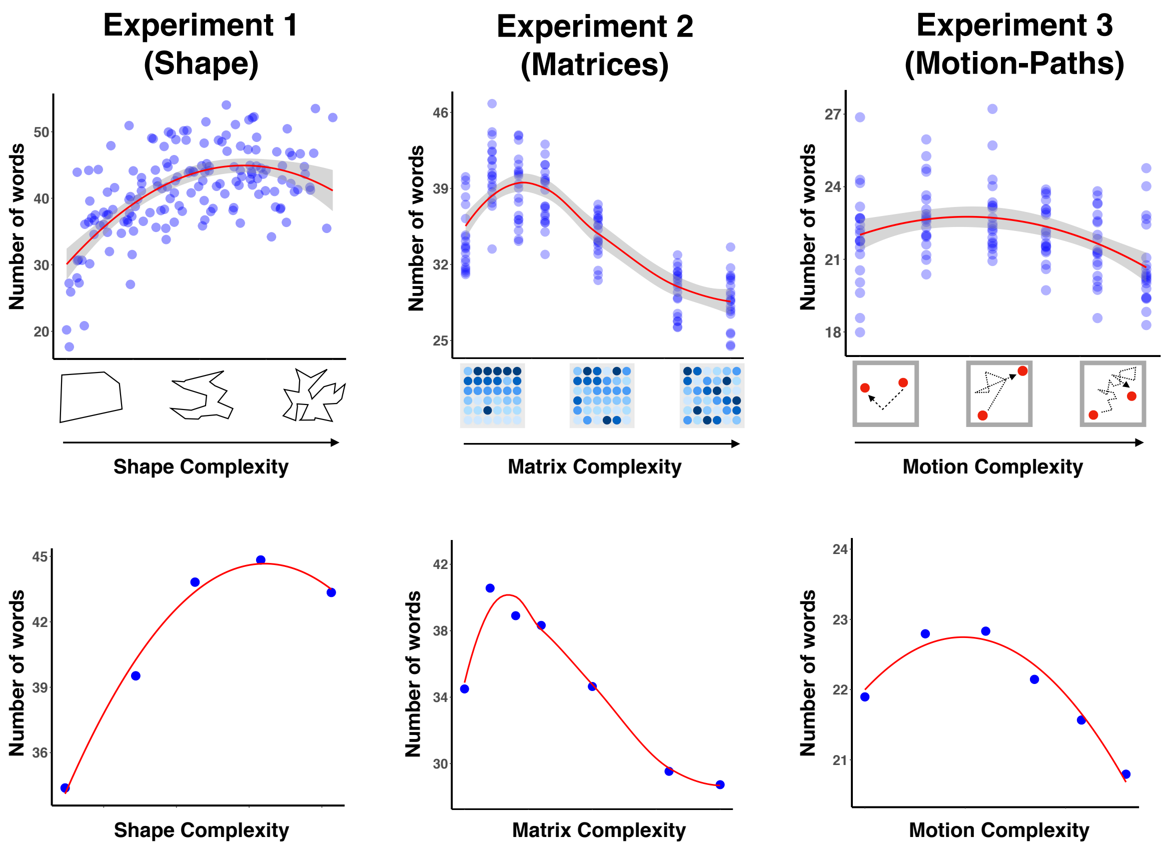

Figure 6: Results from Experiments 1–3, plotting data from over 10,000 freely generated verbal descriptions. Across shapes, matrices and motion-paths, we observed a consistent pattern relating objective complexity and linguistic descriptions: Participants spoke relatively short descriptions for both very simple and very complex stimuli, using the greatest number of words for describing moderately complex stimuli. In the top row, each dot represents a unique stimulus (e.g., a shape, matrix, or motion-path). The bottom row plots the same data in 'bins', where each dot collapses over multiple stimuli from a given complexity range, which may help visualize patterns that are not as clearly evident in the top row.

5 Experiment 3: Dynamic Stimuli

The first two experiments explored very different kinds of visual stimuli, and found a "Goldilocks" pattern in verbal descriptions. However, all such stimuli were static visual images, whereas experiences of complexity in the world often involve more dynamic cues. For example, we can appreciate the complexity of a dancer's movement, a bird's flight path, or even the trajectory of a tricky baseball pitch. Such experiences are interesting not only because they involve an additional property (i.e., motion), but also because they are temporally extended, and so necessarily involve not only perception but also (at least some form of) memory. Does the pattern we have been exploring extend to such cases?

Experiment 3 tested this by asking subjects to observe a disk moving around the display, and to give a spoken description of its motion sequence. We manipulated motion complexity by varying how many times the disk changed its direction throughout the sequence, and asked whether the number of direction changes might be related to spoken descriptions by the same characteristic inverted-U-shaped pattern.

5.1 Method

5.1.1 Subjects

As stated in our pre-registration, we aimed for 300 usable subjects for this experiment; to reach this target sample, we recruited 320 subjects total.

5.1.2 Stimuli and Procedure

We generated 120 moving paths, involving 6 "levels" of direction changes (1, 2, 4, 7, 12, and 20). Each path was 630 pixels long in total. To ensure that any changes in direction would be clear enough for subjects to observe, the shortest segment of

arose in Experiment 2 (positive line: $F(1, 38) = 39.12$, $p < 0.0001$; negative line: $F(1, 118) = 143.7$, $p < 0.0001$). For Experiment 3, we set the break point at 0.3 and found a significantly negative line ($F(1, 98) = 143.7$, $p < 0.05$), but a non-significant positive line (though the relevant effect was in the predicted direction: $F(1, 38) = 1.38$, $p = 0.24$). This overall pattern encouraged us that the inverted U-shaped relationship we explore here is relatively robust to different analytical approaches, though of course future work could employ experimental designs that are better suited to this test.

any motion sequence (i.e., the minimum distance between two turning points) was at least 30 pixels long, and the smallest turning angle was 10 degrees. The disk moved at a speed of 200 pixels per second, resulting in a full motion-animation lasting 3.15 seconds.

The experiment otherwise proceeded in the same way as Experiments 1 and 2, with the following changes. Subjects saw 2 instances of each # of direction changes, randomly chosen from the 20 paths that were generated for each level, resulting in $2 \times 6 = 12$ trials shown in total. On each trial, a red disk first appeared in a random position on the screen, and then subjects pressed a button labeled "Start motion" to begin the sequences. After the sequence was over, the disk disappeared and subjects were instructed simply to "describe what the ball did". (They were not permitted to replay the animation.)

5.2 Results and discussion

17 subjects were excluded for giving extremely short descriptions or uploading low-quality audio files, leaving 303 subjects with analyzable data. From these subjects, 3599 audio clips were successfully uploaded to the server (indicating a rate of successful uploading of 99.0%). Of these files, 185 (5.1%) could be processed as input but not transcribed; finally, another 169 (4.7%) of files were excluded for being only one word long. This left 3245 audio files included in our final analysis (90.2% of all uploaded files), resulting in approximately 27 unique description clips for each of the 120 animations. The average duration of these included speech files was 11 seconds (MIN=2, MAX=58, SD=6), and the average description length was 22 words (MIN=2, MAX=97, SD=13).

Given the demands on memory, one might have expected descriptions of these stimuli to simply plateau (e.g., once a subject's memory capacity was reached), rather than dip *downward* past a certain point. However, the same striking inverted-U-shaped relationship emerged again here in Experiment 3, with a clear and gradual drop-off in description length from moderate complexity to high complexity. Once again, a 2nd-degree-polynomial model ($R^2 = 0.17$, $F(2, 117) = 11.76$, $p <$

2.22×10^{-5}) revealed an inverted-U-shaped relationship between the complexity of moving paths and the number of words used to describe them, with the squared-complexity term being significantly negative ($b = -3.13$, $95\%CI = [-4.79, -1.46]$, $t = -3.70$, $p < 3.27 \times 10^{-4}$). As in previous experiments, this term was also significant in a mixed-effects effect model ($b = -28.47$, $95\%CI = [-45.73, -11.22]$, $t = -3.23$, $p = 0.0016$).[**] Thus, even with dynamic stimuli, subjects' verbal descriptions showed a "Goldilocks"-like pattern.

6 Semantic Analyses

The experiments reported here revealed an inverted-U-shaped relationship between the objective complexity of a stimulus and the length of the verbal description that subjects generated to represent this stimulus. We have interpreted this pattern along the lines of the "programming" example from Figure 2 — a sense that overly complex objects are represented in the mind as having a more "random" or "patternless" underlying explanation. Is there any evidence for this interpretation, beyond the number of words themselves? To further understand the relationship between verbal descriptions and stimulus complexity, we analyzed the content of the words spoken, in two follow-up exploratory analyses: One analysis (reported below) explores the number of "random" words that subjects used. The other analysis (reported in Supplementary Material S1) examines the frequencies of the words used in subjects' descriptions.

6.1 Do subjects use more "random" words?

To explore the possibility that subjects interpret more complex stimuli as being more "random", we identified a small number of "keywords" that indicate randomness and patternlessness. We entered the words "random" and "patternless" into a thesaurus and identified the words "random(ly)", "irregular(ly)", "odd(ly)", "disorganized", "crazy", "strange", and "weird" as keywords that were both (a) aligned with the

[**] As in Experiment 2, we pre-registered a linear model and an exponential model in addition to a quadratic model. Here, the linear model reached statistical significance ($R^2 = 0.070$, $F(1, 158) = 8.85$, $p = 0.0035$), while the exponential model failed to converge.

concepts we had in mind, and also (b) spoken with some meaningfully non-zero frequency in our experiments. (For example, the word "arbitrary" also appeared as a result of this keyword search, but not even a single subject in any experiment used this word in any of their descriptions.) The use of such "random" words would indicate exactly the sort of "lossy" compression we discussed earlier, since describing a shape as "random" or "crazy" leaves out the particular *way* or *reason* it has that property (just as describing a long string of 1s and 0s in terms of a program that randomly chooses 1s and 0s summarizes the pattern but loses the particular sequence that was generated). We then asked whether these words were more likely to appear in the verbal descriptions of complex stimuli.

Indeed, there was a strong correlation between objective stimulus complexity and the frequency of "random" words, in all three experiments (Shape: $r(158) = .42$, $p = 4.22 \times 10^{-8}$; Matrices: $r(138) = .76$, $p = 2.2 \times 10^{-16}$; Motion-Path: $r(118) = .59$, $p = 1.08 \times 10^{-12}$) (Figure 7). In other words, subjects were much more likely to describe extremely complex shapes, matrices, and motion-paths as "random" than they were for simple or moderately complex stimuli, just as would be predicted by an account of "lossy compression" in the representations of such stimuli.[††]

This analysis also suggests that shorter descriptions for overly complex stimuli do not merely result from subjects "giving up" in describing the stimuli. For example, it might have been that subjects don not believe they will be able to capture the detail they would like to in their description, and so simply stop their descriptions early. But while that may of course be part of the explanation here — and it would not be an entirely uninteresting one (after all, knowing when to give up must rely in part on a representation of complexity in the first place) — the present analysis shows that it couldn't be the *whole* explanation: Subjects are not just 'interrupting' their descriptions and stopping; they are actively attributing randomness and patternlessness to the stimuli, and for the complex stimuli in particular.

[††]Note that this was only an exploratory analysis; we mentioned in our pre-registration that we would examine the content of the spoken words, but not at the level of the specific keywords used in this analysis. However, this pattern was not limited to the particular keywords we chose. For example, the pattern remained robust and highly significant even with only three keywords — "random", "irregular", and "crazy".

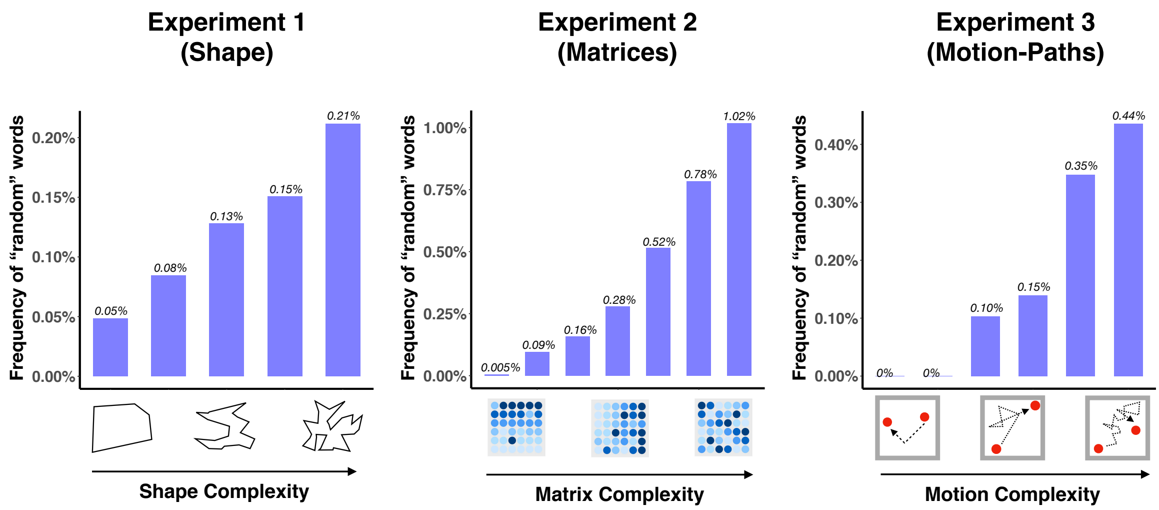

Figure 7: Results from an analysis exploring the use of words indicating "randomness". Across all three experiments, subjects consistently and reliably used more "random" descriptors for increasingly complex stimuli.

7 Discussion

What is the relationship between complexity in the world and complexity in the mind? Whereas it may seem intuitive that more complex stimuli would generate more complex representations of such stimuli, the experiments reported here suggest an alternative (or additional) possibility: that the complexity of a stimulus' representation may vary as a quadratic function of the stimulus' objective complexity. By asking subjects to freely describe stimuli whose "objective" complexity was systematically manipulated from very simple (e.g., triangles and rectangles, highly regular dot-arrays, and a moving disc making one brief turn) to very complex (e.g., extremely irregular "zigzag" shapes, maximally disorganized dot-arrays, and a moving disc turning 20 times), we found that subjects initially generated longer descriptions to represent more complex stimuli, but then passed a "sweet spot" after which their descriptions *shortened* (rather than continued to increase in length or remain at a plateau). Across these distinct categories of stimuli, the results showed an inverted-U-shaped relationship between the length of verbal descriptions and the underlying complexity of the stimuli themselves — a "Goldilocks" effect relating these two notions of complexity.

7.1 "Description length" in language and mind

Exploring the length of freely generated verbal descriptions is, to our knowledge, a new approach for studying the nature of internal object representations. But why think that language — especially the spoken descriptions analyzed here — offers the right kind of access to such representations? Of course, linguistic descriptions are only one way to represent the world, and there are many intervening cognitive steps between seeing a stimulus and verbally describing it. Indeed, the effects we observe here could arise from perception (how complex the objects *look*, as a function of their objective complexity), conceptual representation (roughly, how complex we interpret them to be upon seeing them), memory (e.g., if the motion paths in Experiment 3 are remembered as being differently complex than they really are), and of course the translation of any of these representations into speech.

Nevertheless, our use of linguistic descriptions here doesn't imply that they are some kind of "pure" or "direct" window into one's underlying mental representations; instead, as is reflected in our analyses themselves, our interest in linguistic descriptions lies in their utility for *comparing* representations across different stimuli (whatever the nature of those representations might be). Whatever cognitive constraints apply to transforming a perceptual representation into a linguistic one presumably hold for *any* stimulus one observes (at least in the context of a single experiment above); and so it is the *pattern* of such descriptions — how they *vary* as a function of stimulus complexity — that we believe provides insight here. Moreover, linguistic descriptions have certain advantages over other, more conventional measures. For example, unlike more typical measures from vision science (e.g., subtle differences in reaction time that require sensitive measurements and emerge only after hundreds of trials), linguistic descriptions are easily recorded, and even a single trial can be interpreted. At the same time, unlike other more explicit measures (e.g., "ratings" of complexity on a numerical scale), linguistic descriptions derive from processes that are far more natural and universal, requiring little to no special instruction (e.g., explanations of how to use a certain scale, etc.); subjects in our experiments simply spoke in whatever way felt best.

In other words, even if some other measures of complexity (e.g., ratings) might reveal a linear (rather than quadratic) relationship between objective and subjective complexity, we suggest that there is a special and even unique value inherent to the linguistic description we collect here. Indeed, the value we see in such descriptions is precisely that they are *descriptions*. Many information-theoretic approaches to perception and cognition hold that the mind represents stimuli in the world by transforming them into a format akin to the computer programs appearing in Figure 2 — e.g., as *instructions* for generating such stimuli (Feldman & Singh, 2006). And so studying such representations using measures that are themselves formatted as descriptions may offer a new kind of insight (as it did in the present studies), in ways that make it a valuable addition (though certainly not a replacement) to cognitive psychology's methodological toolkit. (Similarly, recent and creative work has explored connections between visual processing and subjects' *drawings* of visual stimuli [Davis et al., 2019; Fan et al., 2018]. We think those studies carry a similar kind of value, owing in part to the generative nature of such "visual production" tasks.)

This work also joins a small but influential number of recent studies that have made connections between visual complexity and linguistic production. For example, subjects choose longer words to label objects that have more parts (Lewis & Frank, 2016), in ways that suggest a connection between complexity and word-choice. Additionally, subjects who freely produce instructions to identify a target within a grid of distractors (e.g., to pick out a certain green square appearing within a field of red, green, and blue squares and circles) show sensitivity to the complexity of the grid, with delayed speech onset for more complex displays (Elsner et al., 2018). Our work adds to these contributions data that is perhaps even richer (involving full sentences and paragraphs rather than just a single word; Lewis and Frank, 2016) and less constrained (involving a freely spoken description using whatever criteria the subject prefers, rather than a more specific task involving picking out a target; Elsner et al., 2018). All such studies, however, testify to the usefulness of linguistic descriptions as tools for studying other processes in the mind (Perry et al., 2018; Schulz et al., 2020; Zettersten & Lupyan, 2020), in ways that have yielded previously unknown insights about those processes.

One outstanding question is to what degree the quadratic relationship revealed here by verbal descriptions reflects a more general mental representation of complexity. Though we explored three fairly diverse kinds of stimuli within the domain of visual perception, complexity is an attribute that can be assigned to almost any stimulus we encounter, in ways that go far beyond the present work. Within the domain of visual stimuli, one possibility could be to explore images with a high degree of symmetry or self-similarity; a highly symmetric shape, for example, may have a high degree of skeletal entropy but a simpler mental representation that is not captured by this measure. And of course many other kinds of stimuli could be explored as well, including stimuli in other sense modalities and even more abstract concepts that go beyond mere sensory experiences (see *Complexity and Beyond* below).

7.2 Complexity and descriptive language(s)

The preceding discussion repeatedly invokes the notion of "language"; but of course, our study explores only a single natural language (English). Does the pattern we discover here hinge on this choice? Consider the "programming" example from our Introduction (Figure 2): Many of the programs in this example only get to be "short" because they include certain predefined functions (e.g., list multiplication) that are already recognized by the relevant compiler. By analogy, some of the "short" verbal descriptions we obtained do something similar: A description like "almost a perfect square with one corner chopped off" (Figure 5) requires first knowing words like "square", "corner" and "chopped". Importantly, such terms may not be universal across languages: For example, cross-linguistic agreement in naming even simple shapes may be surprisingly low (Majid et al., 2018), such that short descriptions in one language needn't imply short descriptions in another.

However, one reason to suspect some cross-linguistic universality in the connection between complexity and language comes from other work. For example, Lewis

& Frank's (2016)'s study of conceptual complexity and word length (where subjects choose longer names for more complex concepts) included cross-linguistic data from 80 languages, finding surprising universality in their observed pattern. Though the pattern they found was linear rather than quadratic (with increasingly complex objects receiving increasingly complex names), we nevertheless take this earlier work to suggest that, given some mapping between objective complexity and linguistic complexity (whatever that mapping may be), that mapping may be *unlikely* to vary arbitrarily across languages. Of course, settling this question empirically (by exploring freeform verbal descriptions of complex stimuli in other natural languages) marks an opportunity for future work.

Another kind of "cross-linguistic" investigation might explore artificial or constructed communication and notation systems, in addition to natural languages. For many specialized activities, practitioners have developed codes for efficiently representing actions or states relevant to those activities. For example, the knitting community uses a set of abbreviations and conventions that compose into "formula"-like codes for producing a complex pattern; for example, the formula "*CO even # of sts. R1: * K1, P1. Repeat.*" specifies a complex sequence of knitting actions that are understandable to a knitter who can read the notation. (Similar examples exist for chess moves, origami folds, and so on.) One possible difference between these notation systems and natural languages is that the former is intended to losslessly or deterministically convey information. Future work, then, might ask whether an inverted-U pattern would emerge even in such a highly structured and efficient coding system — e.g., if a knitter were to describe an extremely complex pattern in a paradigm similar to the one we use here.

7.3 Complexity and beyond

A final contribution of the present work is the general nature of both (1) the data we have collected, and (2) the freely-spoken-description method itself. Though we have applied this method to the study of visual complexity, and drawn particular conclusions from the results yielded by this method, future studies and analyses may go far beyond the questions (and answers) we offer here.

First, beyond the theoretical conclusions that we draw from the spoken descriptions we have collected, other researchers may well find value in them for other purposes. For example, researchers interested in different but related visual-linguistic concepts — such as concreteness, nameability, or mappings between images and sounds (as in, e.g., the "kiki"/"bouba" effect; Köhler, 1929) — might explore this corpus for patterns relevant to those questions (e.g., "rounder" sounds spontaneously generated for "rounder" images). Second, the pipeline for collecting many thousands of free spoken descriptions may prove useful for other research questions as well, including other explorations of complexity. For example, a natural extension of the current work would be to explore spoken descriptions of simple and complex stimuli in other sensory modalities (e.g., simple and complex auditory stimuli), as well as instances of complexity that go beyond perception itself. For example, the length of freely spoken descriptions could be used to examine subjective representations of the complexity of a story (e.g., if a subject were asked to summarize that story; Chen et al., 2017), a scientific theory or concept (e.g., if a subject were asked to explain that theory or concept; Feldman, 2000), or even various social structures and networks (e.g., if a subject were asked to describe who is friends with whom, which judges or politicians vote together, and so on; Kemp and Tenenbaum, 2008).

Though such questions are beyond the present work, they represent new and promising opportunities for this method. Indeed, a persistent challenge in the study of complexity has long been to find a measure that can apply to the many different stimuli and domains that cause us to experience complexity. Incorporating notions of "description length" has long been seen as a promising step in this direction; here, we've explored the utility of taking this approach one step further, by pursuing a connection between computational and linguistic versions of this notion.

Acknowledgements

For helpful discussion and/or comments on earlier drafts, we thank Yuan Chang Leong, Gary Lupyan, Martin Zettersten and members of the JHU Perception and Mind Lab.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, *53*(4), 221.

Ayzenberg, V., Chen, Y., Yousif, S. R., & Lourenco, S. F. (2019). Skeletal representations of shape in human vision: Evidence for a pruned medial axis model. *Journal of Vision*, *19*, 6.

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, *9*, 9359.

Barbot, A., Liu, S., Kimchi, R., & Carrasco, M. (2018). Attention enhances apparent perceptual organization. *Psychonomic Bulletin & Review*, *25*(5), 1824–1832.

Berlyne, D. E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology*, *55*(3), 289.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286.

Berlyne, D. E., & Peckham, S. (1966). The semantic differential and other measures of reaction to visual complexity. *Canadian Journal of Experimental Psychology*, *20*, 125.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: Analyzing text with the natural language toolkit.* " O'Reilly Media, Inc."

Blum, H. (1973). Biological shape and visual science (part i). *Journal of Theoretical Biology*, *38*(2), 205–287.

Chaitin, G. J. (1969). On the length of programs for computing finite binary sequences: Statistical considerations. *Journal of the ACM (JACM)*, *16*(1), 145–159.

Chaitin, G. J. (1975). A theory of program size formally identical to information theory. *Journal of the ACM (JACM)*, *22*(3), 329–340.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*(3), 566.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115.

Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 365–378.

Chomsky, N. (1957). *Syntactic structures*. The Hague, Mouton.

Chomsky, N. (1968). *Language and mind*. Cambridge University Press.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410.

Cutting, J. E., & Garvin, J. J. (1987). Fractal curves and complexity. *Perception & Psychophysics*, *42*(4), 365–370.

Davis, C. P., Morrow, H. M., & Lupyan, G. (2019). What does a horgous look like? nonsense words elicit meaningful drawings. *Cognitive Science*, *43*, e12791.

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73.

Elsner, M., Clarke, A., & Rohde, H. (2018). Visual complexity and its effects on referring expression generation. *Cognitive Science*, *42*, 940–973.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*(2), 301.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*(8), 2670–2698.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*(6804), 630.

Feldman, J. (2016). The simplicity principle in perception and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 330–340.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Firestone, C., & Scholl, B. J. (2014). “Please tap the shape, anywhere you like”: Shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science*, *25*(2), 377–386.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Koenderink, J., Van Doorn, A., & Pinna, B. (2018). Measures of prägnanz? *Gestalt Theory*, *40*(1), 7–28.

Köhler, W. (1929). *Gestalt psychology*. Liveright.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition ofinformation'. *Problems of Information Transmission*, *1*(1), 1–7.

Leeuwenberg, E. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*(2), 216.

Leeuwenberg, E. (1967). *Structural information of visual patterns: An efficient coding system in perception* (Doctoral dissertation). Radboud University Nijmegen.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195.

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*(5), 1278–1289.

Madan, C. R., Bayer, J., Gamer, M., Lonsdorf, T. B., & Sommer, T. (2018). Visual complexity and affect: Ratings reflect more than meets the eye. *Frontiers in Psychology*, *8*, 2368.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'grady, L., Woll, B., LeLan, B., De Sousa, H., Cansler, B. L., Et al. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376.

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes, In *Proceedings of the annual meeting of the cognitive science society*.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), e12572.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 1–17.

Schulz, E., Quiroga, F., & Gershman, S. J. (2020). Communicating compositional patterns. *Open Mind*, *4*, 25–39.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Siddiqi, K., Tresness, K. J., & Kimia, B. B. (1996). Parts of visual form: Psychophysical aspects. *Perception*, *25*(4), 399–424.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*(5), 369.

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, *1*(4), 538–555.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.

Speer, R., Chin, J., Lin, A., Jewett, S., & Nathan, L. (2018). LuminosoInsight/wordfreq: v2.2. *journal*.

Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003). Universal aesthetic of fractals. *Computers & Graphics*, *27*(5), 813–820.

Sun, Z., & Firestone, C. (2020, October 15). Lossy compression of a complex world. Retrieved from osf.io/a57gy.

Sun, Z., & Firestone, C. (2021). Curious objects: How visual complexity guides attention and engagement. *Cognitive Science*, *45*(4), e12933.

van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*(5), 770.

van der Helm, P. A., van Lier, R. J., & Leeuwenberg, E. (1992). Serial pattern complexity: Irregularity and hierarchy. *Perception*, *21*(4), 517–544.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172.

Wallace, C. S. (2005). *Statistical and inductive inference by minimum message length.* Springer Science & Business Media.

Watson, A. B. (2011). Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity. *The Mathematica Journal*, *14*.

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*(3), 325–340.

Wilder, J., Feldman, J., & Singh, M. (2016). The role of shape complexity in the detection of closed contours. *Vision Research*, *126*, 220–231.

Wolfram, S. (2002). *A new kind of science* (Vol. 5). Wolfram media Champaign, IL.

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition*, *196*, 104135.

Zhang, A. (2017). Speech recognition (version 3.8).
https://pypi.org/project/speechrecognition/3.8.0/.