

1           **Relating Objective Complexity, Subjective Complexity and Beauty**  
2           **using Binary Pixel Patterns**

3  
4           Surabhi S Nath<sup>1,2</sup>, Franziska Brändle<sup>1</sup>, Eric Schulz<sup>1</sup>, Peter Dayan<sup>\*,1,3</sup>, Aenne Brielmann<sup>\*1</sup>

5  
6           <sup>1</sup>Max Planck Institute for Biological Cybernetics, Tübingen, Germany

7           <sup>2</sup>Max Planck School of Cognition, Stephanstrasse 1a, Leipzig, Germany

8           <sup>3</sup>University of Tübingen, Tübingen, Germany

9           \*Joint last authors

10  
11  
12  
13  
14  
15  
16           **Author Note**

17           Surabhi S Nath  <https://orcid.org/0000-0002-0569-4908>

18           Franziska Brändle  <https://orcid.org/0000-0003-1031-337X>

19           Eric Schulz  <https://orcid.org/0000-0003-3088-0371>

20           Peter Dayan  <https://orcid.org/0000-0003-3476-1839>

21           Aenne Brielmann  <https://orcid.org/0000-0002-6742-2836>

22           We have no conflict of interest to disclose.

23           This work was supported by the Max Planck Society, the Alexander von Humboldt

24           Foundation, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—461354985.

25           Part of this paper is available online as part of the corresponding author's Masters thesis.

26           This work has been updated and extended for inclusion in the current article.

27           Correspondence concerning this article should be addressed to Surabhi S Nath, AGPD, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany.

28           E-mail: [surabhi.nath@tuebingen.mpg.de](mailto:surabhi.nath@tuebingen.mpg.de)

34

## Abstract

35 The complexity of images critically influences our assessment of their beauty. However,  
36 studies relating assessments of complexity and beauty to potential objective measures are  
37 hampered by the use of hand-crafted stimuli which are hard to reproduce and manipulate. To  
38 tackle this, we developed a systematic method for generating 2D black-and-white pixel patterns  
39 using cellular automata, and collected ratings of complexity and beauty from 80 participants.  
40 We developed various computational measures of pattern quantification such as density,  
41 entropies, local spatial complexity, Kolmogorov complexity, and asymmetries. We also  
42 introduced an “intricacy” measure quantifying the number of components in a pattern using a  
43 graph-based approach. We related these objective measures with participant judgements of  
44 complexity and beauty to find that a weighted combination of local spatial complexity and  
45 intricacy was an effective predictor ( $R^2_{\text{test}} = 0.47$ ) of complexity. This implies that people’s  
46 complexity ratings depended on the local arrangement of pixels along with the global number  
47 of components in the pattern. Furthermore, we found a positive linear influence of complexity  
48 ratings on beauty, with a negative linear influence of disorder (asymmetry and entropy), and a  
49 negative interaction between the two quantities ( $R^2_{\text{test}} = 0.65$ ). This implies that there is beauty  
50 in complexity as long as there is sufficient order. Lastly, a moderated mediation analysis  
51 showed that subjective complexity mediates the influence of objective complexity on beauty,  
52 implying that subjective complexity provides useful information over and above objective  
53 complexity. Our data and scripts are available on Github.

54 *Keywords:* empirical aesthetics, objective complexity, subjective complexity, beauty, cellular  
55 automata

56

57      **1. Background and Introduction**

58

59      What makes some geometric patterns such as tile designs more beautiful than others, such as  
60      chess boards or QR-codes? Multiple factors may influence the beauty of objects, for example  
61      colour (Palmer, Schloss & Sammartino, 2013) or curvature (Bertamini et al., 2016).  
62      Complexity is another factor that influences beauty, and is the focus of this work. Complexity  
63      and beauty greatly impact our sensory experiences. Researchers in the empirical aesthetics  
64      community have therefore tried to measure and quantify beauty and complexity. Understanding  
65      the relationship between beauty, complexity and objective image features, and the relationship  
66      between the beauty and complexity assessments themselves has been a topic of great interest  
67      over the last several decades (Chamberlain, 2022; Jacobsen, 2010; Machotka, 1980;  
68      McWhinnie, 1971; Nadal & Vartanian, 2021; Van Geert & Wagemans, 2020). While progress  
69      has been made, there is a general lack of consensus in the relationships found.

70

71      1.1 Relationship between Subjective and Objective Complexity

72

73      The complexity literature includes several attempts to arrive at an understanding of the basis  
74      of human ratings of complexity. Although complete consensus has not been reached, two  
75      prominent streams of work emerge – one based on explicit measures which focusses on  
76      identifying separable object features that contribute to experienced complexity, and the other  
77      based on statistical measures, which focuses on the objects' statistical properties.

78

79      We will first discuss the first stream of work, comprising of a line of studies that has attempted  
80      to delineate relevant object features that contribute to complexity. These features include the  
81      density, number and variety of elements (such as vertices, lines, turns or sides), colours and  
82      variety of colours in the image (Chikhman et al., 2012; Friedenberg and Liby, 2016; Munsinger  
83      and Kessen, 1964; Tinio and Leder, 2009). The numbers of vertices, sides and independent  
84      elements in regular geometric stimuli have been shown to be good predictors of their  
85      complexity (Arnoult, 1960; Attneave, 1957; Berlyne et al., 1968; Hall, 1969). The presence of  
86      symmetry has been noted to reduce the perceived complexity (Arnoult, 1960; Attneave, 1957;  
87      Day, 1967, 1968; Eisenman and Gellens, 1968), whereas broken symmetries increased  
88      perceived complexity (Gartus and Leder, 2013).

89

90      In an integrative contribution, Nadal (2007) proposed seven measures of complexity that  
91      related to subjective complexity judgements: unintelligibility of elements, disorganization,  
92      number of elements, variety of elements, asymmetry, variety of colours and degree of three-  
93      dimensional appearance. Subjective complexity was reliably predicted using one or two of  
94      these seven dimensions, however, they varied according to the type of stimulus. While the  
95      number of elements was the most informative predictor, variety of colours and three-  
96      dimensional appearance had little influence on complexity judgements. These seven measures  
97      were further decomposed into three factors: (1) an elements factor comprising of number and  
98      variety of elements, (2) an organization factor, comprising of unintelligibility of elements and  
99      disorganisation, and (3) asymmetry, with each factor relating to different perceptual and  
100     cognitive processes.

101 Parallely, the works of Chipman (1977) used binary pixel patterns to study the determinants of  
102 subjective complexity. Complexity was predicted by a linear combination of number of turns,  
103 horizontal-vertical and diagonal symmetry. Chipman and Mendelson (1979) extended the  
104 study to find that a measure of contour and double symmetry were significant predictors of  
105 complexity across age groups (aside from group-specific factors). Together, these works  
106 proposed that subjective complexity was determined by two components – a quantity-based  
107 component, for example number of elements, turns or amount of contour in the stimuli and a  
108 structure-based component based on the structural aspects like symmetry, repetitions etc. They  
109 also indicated that these two components might involve different cognitive mechanisms. This  
110 dual processing theory was supported by Ichikawa (1985), who found a quantitative number  
111 of turns factor and a structural subsymmetry factor together predicted the subjective complexity  
112 of binary pixel patterns. The work attributed different levels of processing to these two  
113 components – a fast lower level/primary processing for quantity-based features and a slow  
114 higher-level processing for the discovery of structure. More recently, additional evidence for  
115 this theory came from Gartus and Leder (2017), where a quantitative contrast factor and  
116 structural mirror symmetry factor explained complexity ratings for two sets of abstract binary  
117 patterns.

118 By contrast, the second stream of work has focused more on statistical correlates of subjective  
119 complexity judgements based on statistical properties of the stimuli. Fred Attneave (1919-  
120 1991) was the first to apply information theory to quantify stimulus properties in the context  
121 of aesthetics. Shannon's information theory (Shannon, 1948) conceptualised "entropy" as a  
122 measure of the quantity of information potentially contained in a signal. This was later adopted  
123 as a measure of order and complexity in aesthetics by several authors (Arnheim, 1956, 1966;  
124 Bense, 1960, 1969; Moles, 1958; Schmidhuber, 2009). Low entropy implies low uncertainty,  
125 high predictability, high order, and less complexity, and vice versa. Shannon entropy also found  
126 its way into the creative fields of art (used to measure pattern complexity in the Kolam artform  
127 Tran et al., 2021) and music (used to measure complexity in sequences of tones (Delplanque  
128 et al., 2019)). The compressibility of an image could be used as a measure of its complexity  
129 with less compressibility implying more complexity and vice-versa (Birkin, 2010; Donderi,  
130 2006). Kolmogorov complexity is a similar quantification of data compression determined by  
131 the length of the shortest computer program that produces a desired output in any standard  
132 universal computer programming language, and is one of the most direct applications of  
133 algorithmic information theory to stimulus description (eg. Chikhman et al., 2012; Singh and  
134 Shukla, 2017). Güçlütürk et al. (2016) assessed subjective complexity of digitally generated  
135 abstract grayscale images. File size was used as an approximation of Kolmogorov complexity  
136 and was found to correlate highly with subjective complexity. Snodgrass (1971) formulated  
137 two information-theory measures along with symmetry and grouping measures to predict  
138 subjective complexity of algorithmically-generated binary pixel patterns. One of the  
139 information theory measures, based on the entropy of the distribution of submatrices, was  
140 found to be superior in predicting subjective complexity across a series of experiments.  
141 Javaheri Javid recently introduced a measure of spatial complexity based on the probabilistic  
142 spatial distribution of pairs of pixels and also examined the applicability of Lempel and Ziv  
143 compression as Kolmogorov complexity to evaluate the complexity of pixel patterns created

144 using aesthetic automata (Javaheri Javid, 2021). Corchs et al. (2016) used computational  
145 measures quantifying spatial, frequency and colour properties to predict complexity of real-  
146 world stimuli. Other modern studies have used computational image properties such as  
147 histogram of oriented gradients (HOG), Fourier slope and fractal dimension to quantify  
148 complexity (refer to Van Geert and Wagemans, 2020 for a detailed review).

149

## 150 1.2 Relationship between Beauty and Complexity

151

152 In contrast to work focussing directly on complexity, beauty, and aesthetic evaluations more  
153 generally (aesthetic preference/liking, pleasantness, pleasure), have been the subject of a rather  
154 large body of studies, including many attempts to model the processes leading to aesthetic  
155 evaluations. Nevertheless, these two streams of work are closely coupled, as complexity has  
156 been considered an important contributor to aesthetic evaluations since the early days of  
157 empirical aesthetics (although most modern models of aesthetic value consider the role of  
158 complexity only implicitly (Brielmann and Dayan, 2022; Iigaya et al., 2020)).

159

160 Fechner's principle of "unitary connection" suggested that pleasant stimuli express a balance  
161 of complexity and order (Cupchik, 1986), and work by Birkhoff mathematically formulated an  
162 aesthetic measure ( $M$ ) that varied positively with order ( $O$ ) and negatively with complexity  $I$ ,  
163 or  $M = O / C$  for polygonal figures, vases, poetry and music (Birkhoff, 1933). However, Davis  
164 (1936) criticized the measure  $M$  as being inappropriate for empirical test. Several studies aimed  
165 at testing the applicability of  $M$  have yielded a high variance in correlations between actual  
166 rankings and those given by the formula (Eysenck, 1941; Harsh and Beebe-Center, 1939). Later  
167 experiments by Eysenck suggested an empirical formula for  $M$  which yielded much higher  
168 correlations with subject rankings. He suggested an approximation where aesthetic preference  
169 varied positively with *both* order and complexity, altering the equation to  $M = O \times C$  (Eysenck,  
170 1942, 1968).

171

172 Berlyne also formulated a relationship between complexity and aesthetic preference, via a more  
173 general inverted-U relationship between hedonic value and arousal potential (the  
174 "psychological strength" or the extent to which a stimulus is capable of raising arousal  
175 (Berlyne, 1967)) of a stimulus (Berlyne, 1960). Berlyne proposed three classes of variables that  
176 determined arousal potential, namely, psychophysical variables, ecological variables and  
177 collative variables (Berlyne, 1971). Collative variables include subjective novelty, complexity  
178 and surprise. This theory was further applied to the context of art, suggesting that the collative  
179 variable of subjective complexity is one of the most significant determinants of aesthetic  
180 preference. The implication was that there is a general preference for stimuli of intermediate  
181 complexity, as described by the inverted-U shape.

182

183 Several subsequent works have attempted to reproduce this inverted-U relationship between  
184 stimulus complexity and aesthetic preference. However, there is little evidence in support of  
185 this theory, but rather contradictions (Nadal et al., 2010). While several studies supported the  
186 importance of complexity in shaping aesthetic preferences (e.g., Jacobsen and Höfel (2002);  
187 Jacobsen et al. (2006); Tinio and Leder (2009)), the type and directionality of the relationship

188 has not been conclusively settled to date (Nadal et al., 2010). Some studies concurred with the  
189 original inverted U-relationship (Chmiel and Schubert, 2017; Eisenman, 1967; Farley and  
190 Weinstock, 1980; Gordon and Gridley, 2013; Hekkert and Van Wieringen, 1990; Lakhal et al.,  
191 2020; Madison and Schiölde, 2017; Marin et al., 2016; Munsinger and Kessen, 1964; Nasar,  
192 2002; Nicki, 1972; Vitz, 1966), while others suggested a linear relationship (Eysenck, 1941;  
193 Day, 1967; Heath et al., 2000; Javaheri Javid, 2019; Nicki and Moss, 1975; Nicki and Gale,  
194 1977; Osborne and Farley, 1970; Stamps III, 2002; Taylor and Eisenman, 1964). Surprisingly,  
195 a few other studies reported either a non-inverted U (Adkins and Norman, 2016; Norman et al.,  
196 2010), or no relationship at all (Messinger, 1998).

197  
198 The reasons for these discrepancies include theoretical and empirical challenges similar to  
199 those for complexity studies, along with other, unique factors. First, there are contrasting  
200 methods of defining, measuring and manipulating objective complexity (Nadal et al., 2010).

201  
202 Second, most of the above-mentioned works use handcrafted stimuli. Human-generated stimuli  
203 are more likely to suffer from implicit biases of the designer. Moreover, such stimuli cannot be  
204 reproduced without access to the original stimuli set, nor can be manipulated easily, making it  
205 difficult to falsify findings or identify factors causing the failure of generalisation. One  
206 remarkable recent deviation from this pattern is the OCTA toolbox (Van Geert et al., 2022)  
207 which allows the user to generate stimuli comprised of multiple elements based on a series of  
208 settings that aim to manipulate order and/or complexity. This toolbox also provides complexity  
209 and order measures. However, these measures are derived directly from the settings used to  
210 create the stimulus and cannot be applied to an image which it did not itself generate.

211  
212 There is therefore the need for a systematically-generated class of experimental stimuli along  
213 with well-defined programmatic measures that are not linked to the generative process and can  
214 apply generally across instances of the class of stimuli. In this work, we relinquish generality  
215 and naturalness in favour of reproducibility and rigour, and develop an algorithmically defined  
216 stimulus generator which is capable of systematically producing a class of simple yet diverse  
217 abstract patterns that span a range of subjective assessments of complexity and beauty. We  
218 develop various programmatic pattern quantification measures defined for our pattern class  
219 including measures of density, entropy, asymmetry, information in the pattern, along with a  
220 novel “intricacy” measure quantifying the number of separable visual components in the  
221 pattern.

222  
223 Based on this, our work has two contributions:

224  
225 Firstly, we model the subjective complexity of our patterns as a function of the programmatic  
226 measures to obtain an objective measure of complexity.

227

228 Secondly, we inspect the relationship between ratings of the beauty<sup>1</sup> and both the subjective  
229 and objective complexity of our patterns.

230

## 231 2. Methods

232

### 233 2.1 Cellular Automata for Pattern Generation

234

235 We first require a reproducible way of creating visual stimuli spanning a suitable range  
236 of subjective complexities. For this, we require an algorithm that provides us a principled  
237 method for generating a diverse set of stimuli. Furthermore, we want the generated stimuli to  
238 be exactly reproducible by providing the algorithm a set of generation parameters. To satisfy  
239 the aforementioned criteria, we use Cellular Automata (CA) to generate pattern stimuli for our  
240 tasks. CA are iterative algorithms where cells on 1D, 2D or 3D spaces are assigned states as a  
241 function of the states of their neighbouring cells, based (conventionally) on deterministic rules.  
242 They have previously been used to generate many forms of computer art (Adamatzky and  
243 Martínez, 2016; Wolfram et al., 2002).

244

#### 245 **Definition:** *Cellular Automata*

246 A 2D CA, is specified by a quadruple  $\langle L, S, N, f \rangle$  where:

- 247 i.  $L$  is a  $P \times Q$  grid with cells  $(i, j)$ ,  $1 \leq i \leq P$ ,  $1 \leq j \leq Q$
- 248 ii.  $S = \{1, 2, \dots, k\}$  are the potential states of each cell  $(i, j) \in L$ . Therefore, each cell  $(i, j)$   
249 has a state at time  $t$  denoted by  $s_{(i,j)}^t \in S$
- 250 iii.  $N_{(i,j)}$  is the neighbourhood of cell  $(i, j)$  which can either be von Neumann/5-cell ( $N = 5$ )  
251 or Moore/9-cell ( $N = 9$ ) neighbourhoods, where  $N$  is the neighbourhood size. The  
252 neighbourhood is specified by a set of vectors  $\{e_a\}$ ,  $a = 1, 2, \dots, N$  such that  $N_{(i,j)}$  is given  
253 by  $\{(i, j) + e_1, (i, j) + e_2, \dots, (i, j) + e_N\}$  and can be denoted as  $\{(i_1, j_1), (i_2, j_2), \dots, (i_N,$   
254  $j_N)\}$ . A cell is always considered its own neighbour hence one of  $\{e_a\}$  is the zero vector  
255  $(0,0)$ . Periodic boundary conditions are applied at the edges of the lattice so that  
256 complete neighbourhoods exist for every cell in  $L$ .
- 257 iv.  $f$  is the state-transition function which computes the state of cell  $(i, j)$  at the  $t+1 = s_{(i,j)}^{t+1}$   
258 as a function of the states of the cells in its neighbourhood. Hence,  $s_{(i,j)}^{t+1} =$   
259  $f(s_{(i_1,j_1)}^t, s_{(i_2,j_2)}^t, \dots, s_{(i_N,j_N)}^t)$ , where  $(i_1, j_1), (i_2, j_2), \dots, (i_N, j_N) \in N_{(i,j)}$ . A quiescent state  $s_q$   
260 satisfies  $f(s_q, s_q, \dots, s_q) = s_q$ .

261 Starting from an initial configuration of cells at time  $t = 0$ .

262

263 We use binary 2D CA,  $S = \{0, 1\}$  (corresponding to the colours white and black). Taking  
264 inspiration from pixel-art sketches, we set our grid size to  $15 \times 15$  ( $P = Q = 15$ ). We consider  
265 simple rules and initial configurations. We use two classes of rules, conventionally called  
266 “totalistic” (Tot) or “outer-totalistic” (Otot) (Refer to Appendix I for specific details of the  
267 generation algorithm). For our initial configurations (ICs), based on work by Javaheri Javid

---

<sup>1</sup> While we only record beauty evaluations, we consider preference, liking and pleasantness to refer to closely related constructs and expect our results to apply more generally across various measures of aesthetic evaluation (also supported by previous findings, see Marin et al., 2016).

268 (2019), we use either a single central cell (IC\_1), a small disordered central region (IC\_2, also  
269 known as the glider shape in Game of Life CA (Gardner, 1970)) or a fully random starting  
270 grid configuration (IC\_3) (Figure 1A). We limit the number of iterations to  $t = 40$ , and add  
271 every 5<sup>th</sup> pattern to the stimuli set. Therefore, each distinct set of algorithm parameter values  
272 produces 8 patterns.

273

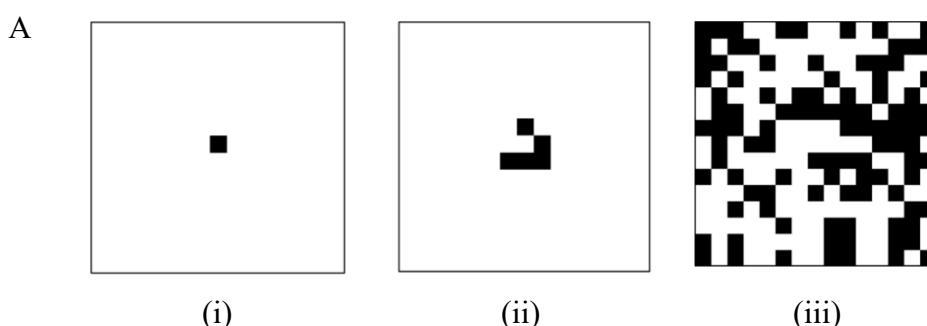
274 We use 51 rules in total with differing combinations of neighbourhood size (5-cell/9-cell  
275 neighbourhood), rule code, rule type (totalistic/outer-totalistic) and initial configuration (IC\_1,  
276 IC\_2 or IC\_3) resulting in a total of nearly 400 patterns. Since pattern evolutions may enter  
277 oscillating configurations, some of the generated patterns are identical. We remove such  
278 patterns from the stimuli set. Figure 1B shows an example pattern evolution based on a  
279 specified rule and Figure 2 shows some of the produced patterns. We also attempted to avoid  
280 any recognizable semantic content in our patterns, since complexity and beauty perception is  
281 largely influenced by familiarity (Tinio and Leder, 2009; Forsythe et al., 2008; Fechner, 1876)  
282 which could confound our findings.

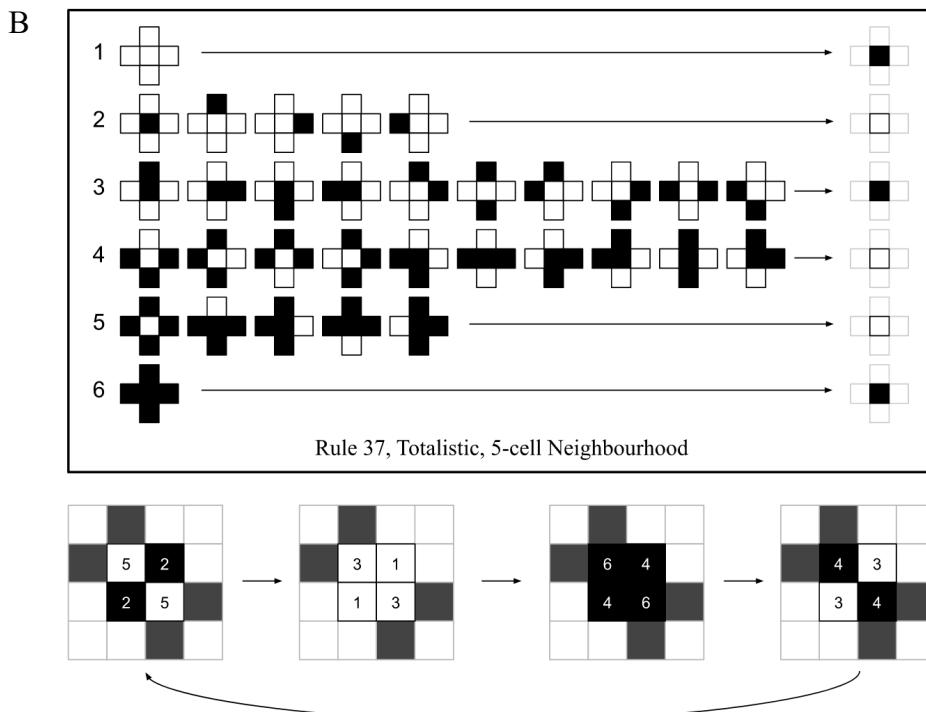
283

#### 284 **Figure 1**

285 *Cellular automata (CA) with (A) 3 initial configurations (ICs): (i) IC\_1, with a single central  
286 cell (ii) IC\_2, with a small disordered region (also called the glider shape in Game of Life CA  
287 (Gardner, 1970)) (iii) IC\_3, with a fully random starting configuration - the figure shown is  
288 an example random starting configuration but the pattern was different for each CA using  
289 IC\_3 (B) example rule (shown in box) along with example pattern updates. The rule shown is  
290 an example of a 5-neighbourhood totalistic rule. The  $i^{\text{th}}$  row in the rule is labelled  $i$  and  
291 describes the update for a cell with exactly  $i-1$  black cells in its neighbourhood, irrespective of  
292 the state of the cell at the time. In this example, row 2 defines the update rule for a cell with  
293 exactly 1 black cell in its neighbourhood to white irrespective of the state of the cell itself at  
294 the time. The example pattern trajectory shows the result of the repeated application of the rule  
295 to a starting pattern. Note that the grid size and IC used here are different from what we use  
296 for our pattern generations and are only used here for illustration purposes. In the example  
297 pattern trajectory, only the central 4 cells are updated and the outer rim of cells is kept  
298 constant. The number on the cell defines the update rule that is being performed on the cell  
299 based on which row in the rule matches the neighbourhood of the cell. In this example, a cell  
300 will be marked 5 if it has exactly 4 black cells in its neighbourhood and would follow the update  
301 rule in row 5 of the rule hence will change-to or stay white.*

302



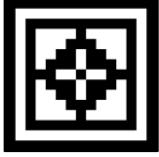
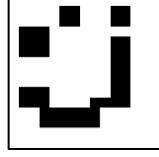
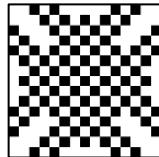
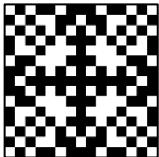


303 **Figure 2**  
304 Examples of cellular automata generated patterns with (rule code, IC, rule type, N, iteration).  
305 Some (manually identified) visual feature categories of diversity are exemplified, each with 2  
306 example patterns – (1) the proportion of black pixels which can be low or high, (2) degree of  
307 symmetry including full symmetry (horizontal, vertical, diagonal and rotational symmetry),  
308 unidirectional symmetry (either horizontal or vertical symmetry), partial asymmetry or full  
309 asymmetry, (3) shape of the pattern, for example, diamond or square, (4) style of the pattern,  
310 such as, maze-like, block-like or framed, and (5) number of components in a pattern where a  
311 component is defined as a connected set of black or white cells where diagonal neighbours of  
312 the same colour are not considered part of the same component.

313

Category	Sub-category	Example 1	Example 2
Proportion of black pixels	Low		
	(224, 3, Otot, 9, 35)      (452, 1, Otot, 5, 5)		
	High		
	(493, 1, Otot, 5, 15)      (191044, 2, Otot, 9, 40)		

Symmetry	Full symmetry (horizontal, vertical, diagonal and rotational symmetry)		
		(467, 1, Otot, 5, 15)	(510, 1, Otot, 5, 10)
	Unidirectional symmetry (either horizontal or vertical symmetry)		
		(736, 2, Otot, 9, 40)	(256746, 2, Otot, 9, 10)
	Partial asymmetry		
		(699054, 2, Otot, 9, 8)	(93737, 2, Otot, 9, 5)
	Full asymmetry		
		(93737, 2, Otot, 9, 15)	(3276, 3, Otot, 9, 35)
Shape	Diamond		
		(462, 1, Otot, 5, 10)	(750, 2, Otot, 5, 8)
	Square		
		(451, 1, Otot, 9, 5)	(374, 1, Otot, 5, 15)
Style	Maze-like		
		(736, 3, Otot, 9, 30)	(736, 3, Otot, 9, 20)
	Block-like		
		(196623, 3, Otot, 9, 40)	(196623, 3, Otot, 9, 30)

Framed		
	(483, 1, Otot, 5, 35)	(481, 1, Otot, 5, 35)
Number of Components	Low	
		(24, 3, Tot, 5, 35)
	High	
		(452, 1, Otot, 5, 25)
		
		(469, 1, Otot, 5, 25)

314  
315       CA provide a principled and structured method for pattern generation since it is fully  
316 specified by  $\langle L, S, N, f \rangle$ , we can deterministically reproduce all our patterns with algorithm  
317 parameters of state space, grid size, neighbourhood size, state-transition function, along with  
318 initial configuration and timestep. Moreover, from Figure 2, we see that the produced patterns  
319 are diverse and vary in multiple features, such as proportion of black pixels, symmetry, shape  
320 or number of components (while this diversity is hard to predict at the outset, some of it is  
321 pre-determinable – for example, IC\_1 results in fully symmetric patterns).

322  
323 2.2 Computational Measures for Pattern Quantification

324 We defined six measures that could be potential influencers of subjective ratings: density,  
325 entropy, local spatial complexity, Kolmogorov complexity (approximate), intricacy, and  
326 symmetry; these are described below. We chose these six measures since they are commonly  
327 studied in the literature as potential determinants of complexity and beauty judgements  
328 (Arnheim, 1956, 1966; Attneave, 1957; Bense, 1960, 1969; Chikhman et al., 2012; Damiano  
329 et al., 2021; Fan et al., 2022; Friedenberg and Liby, 2016; Gartus and Leder, 2017; Javaheri  
330 Javid, 2016; Moles, 1958; Nadal, 2007; Rigau, 2008; Schmidhuber, 2009; Singh and Shukla,  
331 2017; Silva, 2021; Snodgrass, 1971). We do not consider other popular measures such as  
332 number of vertices or edges since as remarked in the previous section, they are hard to define  
333 for our CA patterns.

- 334  
335       a. Density: Density is defined as the proportion of black pixels in the pattern.  
336  
337       b. We define two different measures of entropy:  
338

- 339 (i) Entropy: Entropy assesses the Shannon entropy of the density of black/white  
 340 pixels at a single scale (Eq. 1).  
 341

$$342 -(P(b) \log_2 P(b) + P(w) \log_2 P(w)) \quad (1)$$

343 where  $P(b)$  and  $P(w)$  are the proportions of black and white pixels in the  
 344 pattern respectively such that  $P(w) = 1 - P(b)$ . The range of this quantity is  
 345 from 0 to 1 and the unit of measurement is bits.

- 346 (ii) Multiscale Entropy: Since the entropy measure does not take spatial  
 347 arrangement into consideration, we compute density entropies at all scales and  
 348 average them (Eq. 2).

$$349 -\frac{1}{ns} \sum_{s=1}^{ns} \frac{1}{nw_s} \sum_{w=1}^{nw_s} P(b)_{s,w} \log_2 P(b)_{s,w} + P(w)_{s,w} \log_2 P(w)_{s,w} \quad (2)$$

350 where  $ns$  is the number of scales (in our case,  $ns = P = Q = 15$ ),  $nw_s$  is the  
 351 number of sliding windows at scale  $s$  (defined with overlap  $nw_s = (15 - s +$   
 352  $1)^2$ ),  $P(b)_{s,w}$  and  $P(w)_{s,w}$  are the proportions of black and white pixels in the  
 353 sliding window  $w$  at scale  $s$  respectively.  $P(w)_{s,w} = 1 - P(b)_{s,w}$ . The range  
 354 of this quantity is from 0 to 1 and the unit of measurement is bits.

- 355 c. Local Spatial Complexity: Local Spatial Complexity (LSC) is defined as the mean  
 356 information gain of pixels having homogeneous (same colour) or heterogeneous  
 357 (different colour) neighbouring pixels (Jawaheri Javid, 2019). This measure takes the  
 358 *local* probabilistic spatial distribution of pixels into consideration. The average spatial  
 359 complexity across 8 directions of pixel-neighbour pairs is evaluated (Eq. 3). However,  
 360 this measure is implemented only across one scale giving it the name *local* spatial  
 361 complexity.

$$362 LSC = \frac{1}{8} \sum_{d=1}^8 \bar{G}_d = -\frac{1}{8} \sum_{d=1}^8 \sum_{s_1=1}^2 \sum_{s_2=1}^2 P(s_1, s_2)_d \log_2 P(s_1|s_2)_d \quad (3)$$

363 Here,  $d$  denotes the direction. The direction specifies the positional relationship between  
 364  $s_1$  and  $s_2$ . Assuming  $s_1$  refers to the cell  $(i, j)$ , when  $d = 1$ ,  $s_2$  refers to  $(i + 1, j)$ ; when  
 365  $d = 2$ ,  $s_2$  refers to  $(i - 1, j)$ ; when  $d = 3$ ,  $s_2$  refers to  $(i, j + 1)$ ; when  $d = 4$ ,  $s_2$  refers to  
 366  $(i, j - 1)$ ; when  $d = 5$ ,  $s_2$  refers to  $(i + 1, j + 1)$ ; when  $d = 6$ ,  $s_2$  refers to  $(i + 1, j - 1)$ ;  
 367 when  $d = 7$ ,  $s_2$  refers to  $(i - 1, j + 1)$ ; and when  $d = 8$ ,  $s_2$  refers to  $(i - 1, j - 1)$ .  $s_1, s_2$   
 368 denote the state combination (black-black, black-white, white-black or white-white)  
 369 under consideration.  $P(s_1|s_2)_d$  is the joint probability that a pixel pair (along direction  
 370  $d$ ) has states  $(s_1, s_2)$ .  $P(s_1|s_2)_d$  is the conditional probability that a pixel has state  $s_1$   
 371 given its neighbouring pixel (along direction  $d$ ) has state  $s_2$ .  $\bar{G}_d$  is the mean information  
 372 gain across all state combinations for direction  $d$ .  $LSC$  is the mean of the 8 directional  
 373

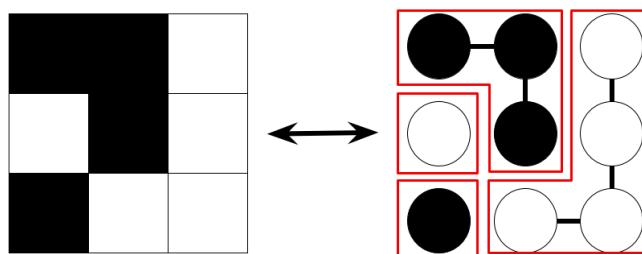
377  $\bar{G}_d$ 's, or the mean of  $\{\bar{G}_1, \bar{G}_2, \dots, \bar{G}_8\}$  measured in bits.

- 378
- 379 d. Kolmogorov Complexity: Kolmogorov complexity (KC) is based on algorithmic  
380 information theory. It is defined as the length of the shortest computer program that  
381 produces the desired pattern and can be used as a measure of pattern complexity.  
382 Kolmogorov complexity is uncomputable, and methods to compute it only estimate an  
383 upper bound. Some such methods are the LZ78 universal compression algorithm (Ziv  
384 and Lempel, 1978) and the Block Decomposition Method (Zenil et al., 2018). In this  
385 work, we use the Block Decomposition Method.
- 386
- 387 e. Intricacy: To quantify the number of elements in a pattern, we introduce an intricacy  
388 measure using a graph-based approach (Authors, 2022). The pattern is encoded as a  
389 graph with each pixel as a node. Edges are added between neighbouring pixels of the  
390 same colour. We considered up, down, right and left (non-diagonal) adjacent pixels as  
391 valid neighbours. Depth first search is performed to count the number of connected  
392 components which is used as our intricacy measure. This value is the sum of black  
393 components and white components in the pattern. This procedure is shown in Figure 3  
394 for an example  $3 \times 3$  grid. For  $15 \times 15$  stimuli, the range of intricacy values is 1 (for  
395 an all-white or all-black pattern) to 225 (for a checkerboard).
- 396

397 **Figure 3**

398 *Intricacy computation for an  $3 \times 3$  example pattern. The pattern is encoded as a graph*  
399 *- the graph on right is constructed from the pattern on left. The number of components*  
400 *(same coloured groups of connected pixels) are counted. The red boxes in the left*  
401 *indicate connected components. Here, intricacy = 4*

402



403 Note. Diagonally adjacent cells of the same colour are not considered part of the same  
404 connected component.

405

- 406
- 407 f. Asymmetry: We use three measures to capture the global and local asymmetry in the  
408 patterns. For global asymmetry, we restrict ourselves to the horizontal and vertical  
409 directions as they are most readily perceived by humans (Giannouli, 2013). Horizontal  
410 asymmetry (Hasymm) and vertical asymmetry (Vasymm) were computed as the  
411 percentages of mismatches in the horizontal and vertical directions respectively (refer  
412 to appendix AII for a mathematical description and illustrated example of these  
413 symmetries). Our third asymmetry measure, local asymmetry was computed as the  
414 average difference in mean information gains as specified by the LSC, along 4  
415 directions.

416

417 Along with these computed metrics, we added the generation algorithm parameters to the set  
 418 of potential predictors of subjective ratings including neighbourhood size ( $N$ ), rule type  
 419 (Tot/Otot), iteration number and IC. Furthermore, based on evidence for  $\lambda$  (defined as the  
 420 percentage of all the entries in a rule table which map to non-zero states – in our case, 1s)  
 421 corresponding with the behaviour of 1D CA (Langton, 1986; Li et al., 1990), we added it as a  
 422 potential predictor as well.  $\lambda$  is calculated as the number of 1s in the binary representation of  
 423 the rule code. Refer to Appendix II for two additional measures we implemented along with  
 424 some examples of the computed measures on various patterns.

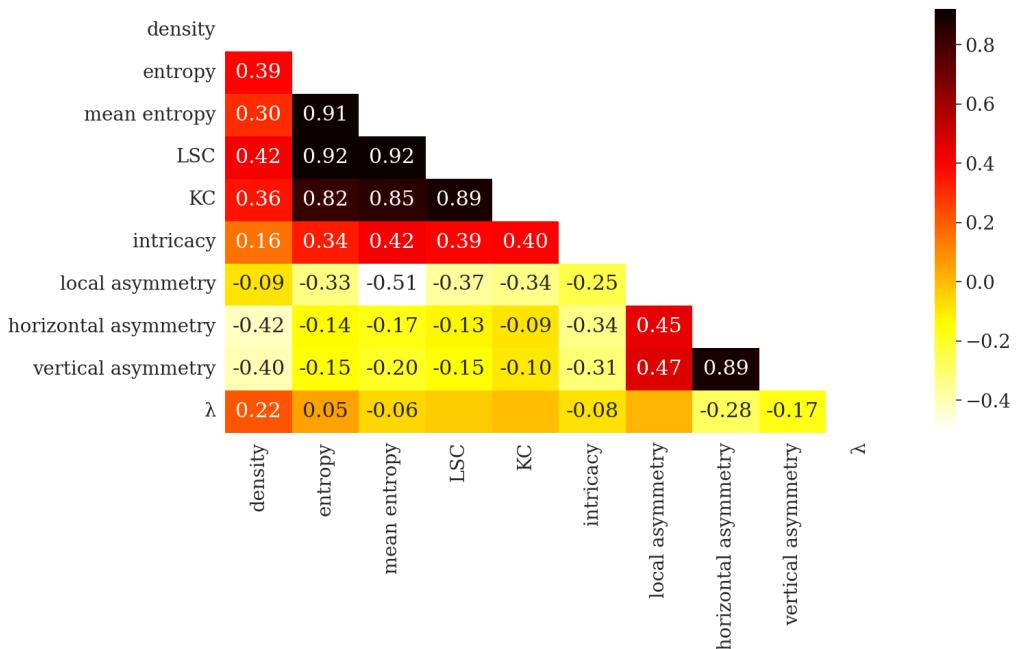
425

426 Figure 4 shows the Pearson correlations between the programmed measures. We see a high  
 427 positive correlation between LSC and KC ( $r(4678) = 0.89, p < 0.01, 95\% \text{ CI} = [0.88, 0.89]$ ).  
 428 This is consistent with Javaheri Javid (2019). Intricacy is positively correlated with LSC  
 429 ( $r(4678) = 0.39, p < 0.01, \text{CI} = [0.37, 0.42]$ ), KC ( $r(4678) = 0.40, p < 0.01, \text{CI} = [0.38, 0.42]$ )  
 430 and (mean) entropy, and negatively correlated with asymmetry measures. Density is correlated  
 431 with LSC ( $r(4678) = 0.42, p < 0.01, \text{CI} = [0.4, 0.45]$ ), KC ( $r(4678) = 0.36, p < 0.01, \text{CI} = [0.33,$   
 432  $0.38]$ ). The asymmetry measures are positively correlated with each other. These correlation  
 433 values helped us exclude variables from the predictor set, for example, we did not use both  
 434 LSC and KC in the same model simultaneously, and combined horizontal and vertical  
 435 asymmetry measures into a single mean asymmetry measure.

436

#### 437 **Figure 4**

438 *Pearson correlations between the measures – density, entropy, multiscale entropy, local spatial  
 439 complexity (LSC), Kolmogorov complexity (KC), intricacy, local asymmetry, horizontal  
 440 asymmetry, vertical asymmetry and number of rules ( $\lambda$ ).*



441

442 *Note.* The colour scale on the right represents the strength of correlation ( $r(4678)$ ). Only  
 443 correlations with  $p < 0.01$  are shown with numbers.

444

445 2.3 Pattern Rating Experiment

446

447 For obtaining human ratings on the CA patterns, we programmed an online behavioural  
448 experiment where participants were recruited to view and rate the beauty and complexity of  
449 the patterns.

450

451 I. Design

452

453 We asked each participant in our experiment to rate the beauty and complexity of the patterns  
454 as they perceived them. We did not provide any definitions of complexity or beauty in order to  
455 elicit people's unbiased opinion. However, to set the prior over the possible types and variety  
456 of patterns, we showed 12 example patterns in 2 groups of 6 patterns, where each group  
457 comprised of sufficient visual diversity. We randomized the order of the example patterns to  
458 avoid biasing participant ratings. We recorded the complexity and beauty ratings using two  
459 slider bars ranging from 0 to 100. Both slider bars were shown below the pattern at the same  
460 time. They were labelled only at their ends as either "Low/High Complexity" for the  
461 complexity rating slider and "Low/High Beauty" for the beauty rating slider, with no  
462 intermediate marking. This was done to encourage participants to rate evenly over the whole  
463 range of possible ratings. We also did not set any default value on the sliders to avoid  
464 influencing the participant ratings and participants had to click on the slider to make the pointer  
465 appear. We programmed the experiment in JavaScript using jsPsych (De Leeuw, 2015).

466

467 II. Stimuli

468

469 We generated nearly 400 CA patterns using Python (Version 3.8.8). This set was split into 4  
470 sets of 54 patterns each (216 patterns in all). The patterns in each set were picked such that  
471 they span the range of complexity values as quantified by our LSC, density and intricacy  
472 measures. The sets were manually balanced to contain visually similar (but non-identical)  
473 patterns spanning over the features listed in Figure 2. Participants were assigned one of the 4  
474 sets in serial order (participant I received set  $(i\%4)+1$ ). For each participant, we randomly  
475 selected 6 patterns (out of the total 54) and showed them twice to get repeated measures. We  
476 ensured no two repeats occurred together in the pattern sequence. Following this design, each  
477 participant rated 60 patterns.

478

479 III. Procedure

480

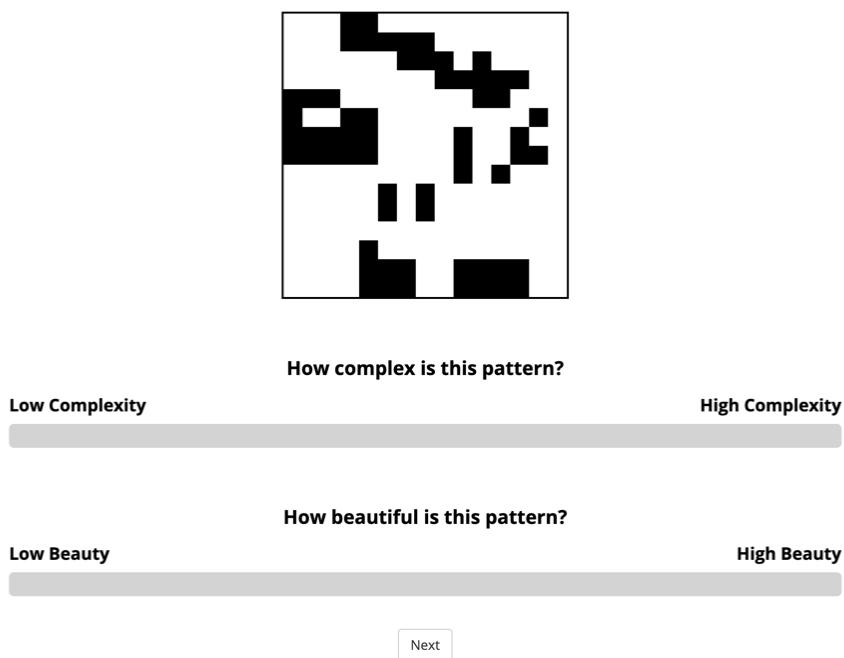
481 The experiment opened with a welcome screen, followed by a consent form and data protection  
482 form. An overview page displayed the task instructions. Participants then proceeded to the  
483 ratings where they were shown a pattern at the top of the screen with 2 sliders positioned below  
484 the pattern. The two sliders recorded responses to the two questions – "How complex is this  
485 pattern?" and "How beautiful is this pattern?" (Figure 5). The pattern was displayed in the size  
486 300 px  $\times$  300 px irrespective of screen size. People could use either laptops or desktops to take

487 part in the experiment. The order of questions remained the same across trials and across  
488 participants.

489

490 **Figure 5**

491 *Experimental interface as seen by participants. The pattern is displayed on the top centre in*  
492 *the size 300 px × 300 px. Below it are two sliders with the questions “How complex is this*  
493 *pattern?” and “How beautiful is this pattern?” respectively and with no initial slider marking*  
494 *on them. Below them is a next button which allows the participant to proceed to the next pattern*  
495 *only when both sliders are marked.*



496

497

498 In addition to the 60 pattern ratings, each participant also encountered 2 or 3 attention checks  
499 where the pattern contained an overlaid text “read the questions below” and the slider questions  
500 were modified to read “place the slider head at the extreme right”. The slider endpoint labels  
501 also changed from “Low/High Complexity”, “Low/High Beauty” to “Left/Right”. These two  
502 staged checks served as both a compliance and attention check. Following the ratings,  
503 participant demographic data on gender, age and nationality were recorded. We used the  
504 Vienna Art Interest and Art Knowledge questionnaire to record each participant’s level of art  
505 training (Specker et al., 2020). Finally, we included some open-ended questions about rating  
506 behaviour. These questions asked the participants to indicate the strategies they used to rate  
507 complexity and beauty, and the patterns they found most complex and beautiful. We required  
508 participants to answer all questions. In all, the study took less than 20 minutes to complete.

509

510 IV. Participants

511

512 80 participants from Prolific (50 female, 29 male, 1 other; mean age = 32.3, min age = 18, max  
513 age = 79; 20 participants per set) took part in the study. 2 participants failed all attention checks

514 and were removed from the analysis leaving us with 78 participants. All participants were based  
515 in the United States, were fluent English speakers and had not previously participated in the  
516 experiment. The average study completion time was 14.35 minutes. Each participant was paid  
517 £3.50. All experiments were approved by the ethics committee of the University of Tübingen.  
518

## 519 V. Analysis

520

521 Having acquired the complexity and beauty ratings, we (1) sought a combined computational  
522 measure that could suitably predict the subject-specific complexity ratings across the  
523 population, and (2) determine the relationship between beauty and complexity ratings.  
524

525 Complexity and beauty ratings were found to be balanced across sets, no significant sequential  
526 effects (trends, autocorrelation between participant ratings) were observed, and participant  
527 repeated responses were consistent (refer to Appendix II for details). In reporting their  
528 strategies, participants indicated that pattern intricacy (participants spontaneously used this  
529 word) or number of elements/blocks (~28 participants), their structure, arrangement and  
530 symmetry (~13 participants), and the ability to replicate a pattern (~23 participants) influenced  
531 their complexity ratings while symmetry (~22 participants) and “intuition” or liking (~42  
532 participants) influenced participants’ beauty ratings.  
533

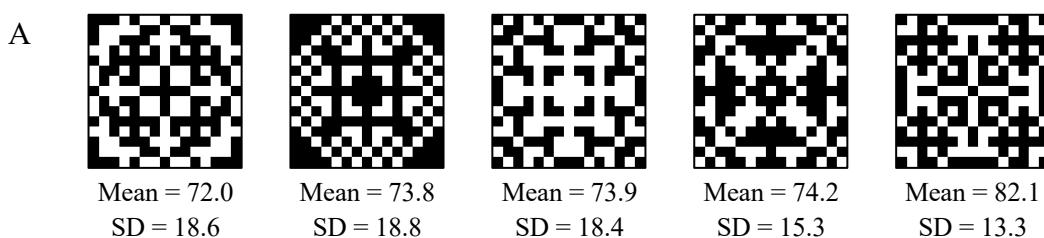
534 Figures 6A and 6B show the patterns with the highest and lowest average complexity ratings  
535 and Figures 6C and 6D show the patterns with the highest and lowest average beauty ratings.  
536 We also studied variance in ratings per pattern and categorized patterns into “high agreement”  
537 (low variance) or “low agreement” (high variance) (Figures 6E and 6F for complexity and 6G  
538 and 6H for beauty).

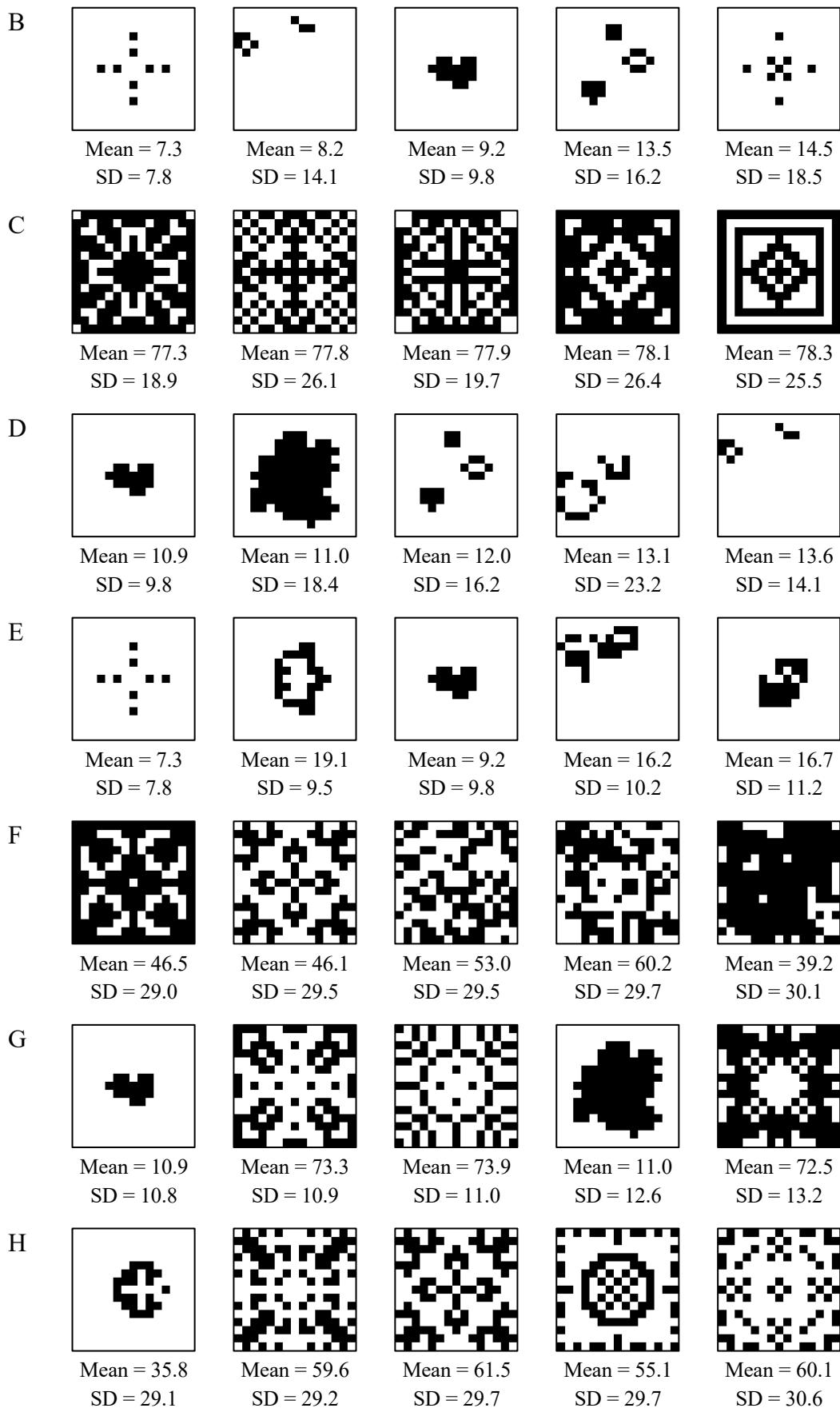
539

## 540 **Figure 6**

541 *Pattern visualizations with mean and variance in ratings. (A) Patterns with highest average  
542 complexity ratings, (B) Patterns with lowest average complexity ratings, (C) Patterns with  
543 highest average beauty ratings, (D) Patterns with lowest average beauty ratings, (E) Patterns  
544 with highest agreement in complexity ratings, (F) Patterns with lowest agreement in complexity  
545 ratings, (G) Patterns with highest agreement in beauty ratings, (H) Patterns with lowest  
546 agreement in beauty ratings*

547





549            *V.I Metric for Complexity*

550  
551 We performed regression analyses on participant complexity ratings in R (Version 4.1.1,  
552 library *lme4* (Bates et al., 2015), function *lmer()*). We used an incremental, bottom up approach  
553 (*i.e.*, starting from the predictors that had high linear correlations with our dependent variable  
554 and adding predictors one by one to our models (Table 1)), to arrive at the objective complexity  
555 metric that best predicted subjective complexity. High linear correlations between the ratings  
556 and the measures justify the use of linear models; hence we fit linear mixed effects models. We  
557 performed cross validation by splitting our data into 3 stratified folds. Each test fold had 20  
558 randomly sampled ratings from every participant and each training fold had the remaining 40  
559 ratings from every participant. The seed was set to 20 for all our experiments. We z-scored all  
560 the predictor variables. We also created additional variables by squaring each predictor to  
561 check for quadratic effects. We used the *bobyqa* optimizer in R with all other parameters same  
562 as default. Horizontal and vertical asymmetry measures were combined into one mean  
563 asymmetry (asymm) measure because of their high correlations. We evaluated our model using  
564 3-fold cross validation and reported the average Akaike information criterion (AIC), Bayesian  
565 information criterion (BIC) and the average  $R^2$  values across the three folds. We also evaluated  
566 Root Mean Square Error (RMSE) values on train and test sets. In addition, we checked for  
567 multicollinearity using Variance Inflation Factors (VIF). We experimented with random  
568 intercepts, random slopes, quadratic and interaction effects in our models. The best model was  
569 defined as the one that had significant predictors and resulted in a low BIC and high  $R^2$  while  
570 ensuring VIFs did not exceed 5. If a model achieved higher  $R^2$  at the expense of higher BIC,  
571 we prioritized model simplicity.

572  
573            *V.II Relationship between Beauty and Complexity Ratings*

574  
575 Finally, to investigate the relationship between beauty and complexity ratings, we used a  
576 similar analysis to that above. We fit linear mixed effects models on the beauty ratings but now  
577 including the complexity ratings as a predictor. The objective complexity measures that best  
578 described the complexity ratings from the previous analysis were combined into one “objective  
579 complexity” measure. Based on previous suggestions that beauty is linked to both complexity  
580 and order (Section 1.2), and asymmetry and entropy quantify disorder, we combined mean  
581 asymmetry and entropy into one “disorder” measure. We also systematically examined the  
582 relationship between objective complexity, subjective complexity and beauty and disorder  
583 using moderated mediation analysis (using PROCESS and *mediation* library (Hayes, 2017;  
584 Tingley et al., 2014)).

585  
586            **3. Results**

587  
588            **3.1 Metric for Complexity**

589  
590 Table 1 presents a summary of our main models and their performance (averaged over 3 cross  
591 validation folds). Our dependent variable, complexity ratings, are abbreviated as CR. We use  
592 the Wilkinson-Rogers notation to report our models (Wilkinson and Rogers, 1973).

593

594 **Table 1**595 *A Models of complexity ratings*

Id	Model	Significance	AIC	BIC	$R^2$		RMSE	
					Train	Test	Train	Test
1	CR ~ 1 + 1 Participant		8545.4	8563.6	0.16	0.13	0.92	0.93
2	CR ~ 1 + 1 Participant + 1 Set		8546.3	8570.5	0.16	0.13	0.92	0.93
3	CR ~ LSC + 1 Participant	LSC*	7552.0	7576.2	0.40	0.37	0.78	0.79
4	CR ~ LSC + LSC Participant	LSC*	7491.3	7527.5	0.43	0.39	0.75	0.78
5	CR ~ KC + 1 Participant	Intercept* KC*	7703.8	7728.0	0.37	0.34	0.80	0.81
6	CR ~ LSC + density + 1 Participant	LSC* density*	7551.3	7581.5	0.40	0.37	0.78	0.79
7	CR ~ LSC + entropy + 1 Participant	LSC* entropy*	7536.6	7566.8	0.40	0.38	0.77	0.79
8	CR ~ LSC + multiscale_entropy + 1 Participant	LSC* multiscale_entropy *	7550.7	7581.0	0.40	0.37	0.78	0.79
9	CR ~ LSC + asymm + 1 Participant	LSC* asymm*	7520.2	7550.5	0.40	0.38	0.77	0.79
10	CR ~ LSC + intricacy + 1 Participant	LSC* intricacy*	7285.1	7315.4	0.45	0.43	0.74	0.76
11	CR ~ LSC + intricacy + LSC:intricacy + 1 Participant	LSC* intricacy* LSC:intricacy*	7276.8	7313.1	0.45	0.43	0.74	0.76
12	CR ~ LSC + intricacy + LSC Participant	LSC* intricacy*	7219.9	7262.2	0.48	0.44	0.72	0.75
13	CR ~ LSC + intricacy + intricacy Participant	LSC* intricacy*	7166.3	7208.6	0.50	0.46	0.71	0.74
14	<b>CR ~ LSC + intricacy + (LSC + intricacy) Participant</b>	<b>LSC* intricacy*</b>	<b>7138.9</b>	<b>7199.4</b>	<b>0.52</b>	<b>0.47</b>	<b>0.70</b>	<b>0.73</b>
15	CR ~ LSC + intricacy + LSC:intricacy + (LSC + intricacy) Participant	LSC* intricacy* LSC:intricacy*	7133.6	7200.1	0.52	0.47	0.70	0.73
16	CR ~ LSC <sup>2</sup> + intricacy <sup>2</sup> + 1 Participant	LSC <sup>2</sup> * intricacy <sup>2</sup> *	7329.3	7359.5	0.44	0.42	0.75	0.77
17	CR ~ LSC + LSC <sup>2</sup> + intricacy + intricacy <sup>2</sup> + 1 Participant	LSC <sup>2</sup> * intricacy*	7256.7	7299.0	0.46	0.43	0.73	0.75

596 Note. CR=complexity ratings, LSC=local spatial complexity, KC=Kolmogorov complexity; \*  
 597 indicates  $p < 0.05$ . Bold indicates best model.

598

599 B *Fixed effects in the best model using fold 1 (refer to Appendix III for plots of random effects).*

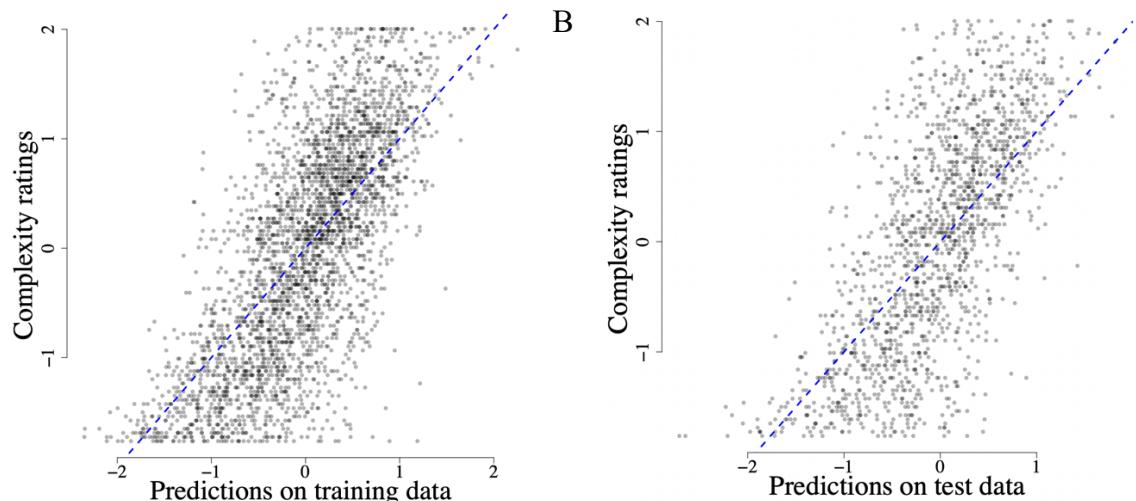
**CR ~ LSC + intricacy + (LSC + intricacy)|Participant**

Intercept (mean)	LSC	Intricacy
-0.02	0.38	0.26

Our analysis shows that subjective complexity can be predicted by a positive linear combination of LSC and intricacy measures with random slopes of LSC and intricacy, and a random intercept of participant (Table 1, row 14,  $R^2_{\text{test}} = 0.47$ , AIC = 7138.9, BIC = 7199.4). While the AIC for the model with an interaction effect is slightly lower (Table 1, row 15), the BIC is slightly higher implying possible overfitting, and so we preferred the simpler model (Table 1, row 14). Figure 7 shows the plot of predictions versus ground truth on the train and test set from cross-validation fold 1.

**Figure 7**

*Performance of the best model CR ~ LSC + intricacy + (LSC + intricacy) | Participant on (A) training and (B) test data from fold 1 for complexity ratings*



*Note.* y-axis displays z-scored complexity ratings and x-axis display their corresponding model predictions. Blue dashed line represents  $y = x$ .

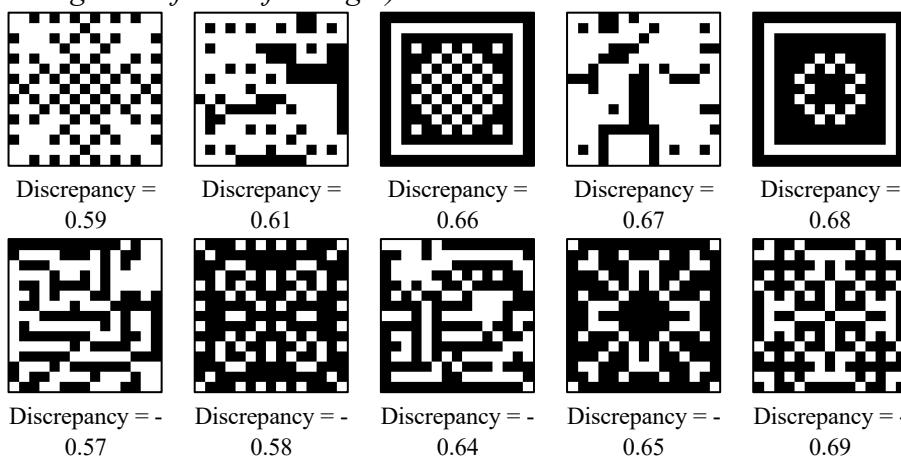
Thus, we found evidence that a weighted combination of spatial complexity and intricacy measures can reliably explain a substantial fraction of human subjective complexity ratings. This implies that people's complexity judgments depend on the local arrangement of pixels (captured by LSC) along with the global number of visual components in the pattern (captured by intricacy). The slopes for both quantities are positive indicating that a larger LSC and a larger intricacy are associated with higher complexity and vice versa. Since intricacy evaluates the number of connected components in the whole pattern, it is a global feature. On the other hand, LSC looks at pairwise pixel distributions at one scale and hence it is a local feature. This suggests that people integrate certain local and global pattern features to arrive at their complexity estimates. Furthermore, random slopes of LSC and intricacy imply that these measures influence the complexity assessments of different participants to different degrees.

627 We also visualized some of the patterns for which the difference between objective complexity  
 628 measure and subjective ratings was the largest on average across all participants (Figure 8). We  
 629 see that the measure overweights complexity in some patterns where the computed intricacy  
 630 measure is high, or where there is added structure as in framed patterns (top row in Figure 8).  
 631 Refer to Appendix II for additional analysis using an 8-neighbourhood variant of intricacy),  
 632 and underweights complexity in some of the high density and (partially) asymmetric patterns  
 633 (bottom row in Figure 8).

634

### 635 **Figure 8**

636 *Some of the patterns with largest (z-scored) overestimation of complexity by the model (row 1,  
 637 increasing in magnitude from right to left) and largest (z-scored) underestimation (row 2,  
 638 increasing in magnitude from left to right).*



639 Note. Discrepancy values denote the average z-scored difference between objective complexity  
 640 measure and subjective ratings for the pattern across all subjects. It is seen that patterns where  
 641 complexity is overestimated by the model tend to have high intricacy and patterns where  
 642 complexity is underestimated tend to have high density or partial asymmetry.

643

### 644 3.2 Relationship between Beauty and Complexity Ratings

645

646 Table 2 summarises our main models and their performance (averaged over 3 cross validation  
 647 folds from data split 1). Our dependent variable, beauty ratings, are abbreviated as BR. The  
 648 disorder measure is a weighted combination of asymmetry and entropy where the weights are  
 649 obtained from the fixed effects of asymmetry and entropy respectively in the model  $BR \sim CR$   
 650 + *asymmetry* + *entropy* +  $I|Participant$ . Objective complexity (OC) was evaluated as the  
 651 weighted combination of LSC and intricacy, where the weights were obtained from the  
 652 regression coefficients (including random effects) of the best performing model (Table 1, row  
 653 12) fit on complexity ratings from cross validation fold 1.

654

### 655 **Table 2**

656 *A Models of beauty ratings*

Id	Model	Significance	AIC	BIC	$R^2$		RMSE	
					Train	Test	Train	Test

1	BR ~ CR + CR*	8046.4	8070.6	0.29	0.25	0.85	0.87
1 Participant							
2	BR ~ CR + disorder + CR*	6619.6	6649.8	0.56	0.53	0.67	0.69
1 Participant	disorder*						
3	BR ~ CR + disorder + CR*	6214.1	6256.5	0.64	0.59	0.60	0.64
disorder Participant	disorder*						
4	BR ~ CR + disorder + CR*	6340.1	6382.4	0.62	0.58	0.62	0.65
CR Participant	disorder*						
5	BR ~ CR + disorder + CR*	5942.6	6003.0	0.69	0.63	0.56	0.61
(CR + disorder) Participant							
6	BR ~ CR + disorder + CR*	6524.6	6560.9	0.57	0.54	0.66	0.68
CR:disorder + disorder*							
1 Participant	CR:disorder*						
7	<b>BR ~ CR + disorder + CR*</b>	<b>5846.4</b>	<b>5912.9</b>	<b>0.70</b>	<b>0.65</b>	<b>0.55</b>	<b>0.60</b>
<b>CR:disorder + (CR + disorder) Participant</b>	<b>CR:disorder*</b>						
8	BR ~ CR <sup>2</sup> + disorder + disorder*	7099.9	7130.2	0.48	0.45	0.72	0.74
1 Participant							
9	BR ~ CR + CR <sup>2</sup> + CR*	6628.0	6664.2	0.56	0.53	0.67	0.69
disorder + 1 Participant	disorder*						
10	BR ~ OC + disorder + OC*	6844.1	6874.3	0.52	0.50	0.69	0.71
1 Participant	disorder*						
11	BR ~ OC + disorder + OC*	6356.5	6417.0	0.64	0.58	0.60	0.65
(OC + disorder) Participant							
12	BR ~ OC + disorder + OC*	6205.1	6271.6	0.66	0.60	0.58	0.63
OC:disorder + (OC + disorder) Participant	OC:disorder*						

657

658 B *Fixed effects in the best model using fold 1 (refer to Appendix III for plots of random effects).*  
**BR ~ CR + disorder + CR:disorder + (CR + disorder)|Participant**

Intercept (mean)	CR	disorder	CR:disorder
-0.04	0.34	-0.51	-0.12

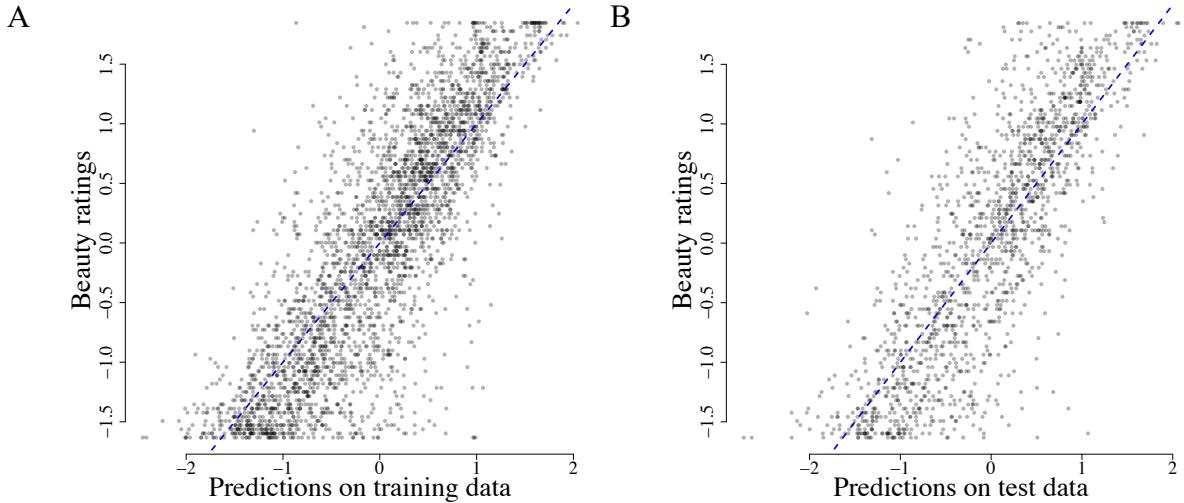
659

660 We found that the beauty ratings are well predicted by complexity ratings, disorder and their  
661 interaction, along with a random intercept of participant and random slopes of subjective  
662 complexity and disorder (Table 2, row 7,  $R^2_{\text{test}} = 0.65$ , AIC = 5846.4, BIC = 5912.9). Figure 9  
663 shows the plot of predictions versus ground truth on a train and test set from cross-validation  
664 fold 1.

665

## 666 **Figure 9**

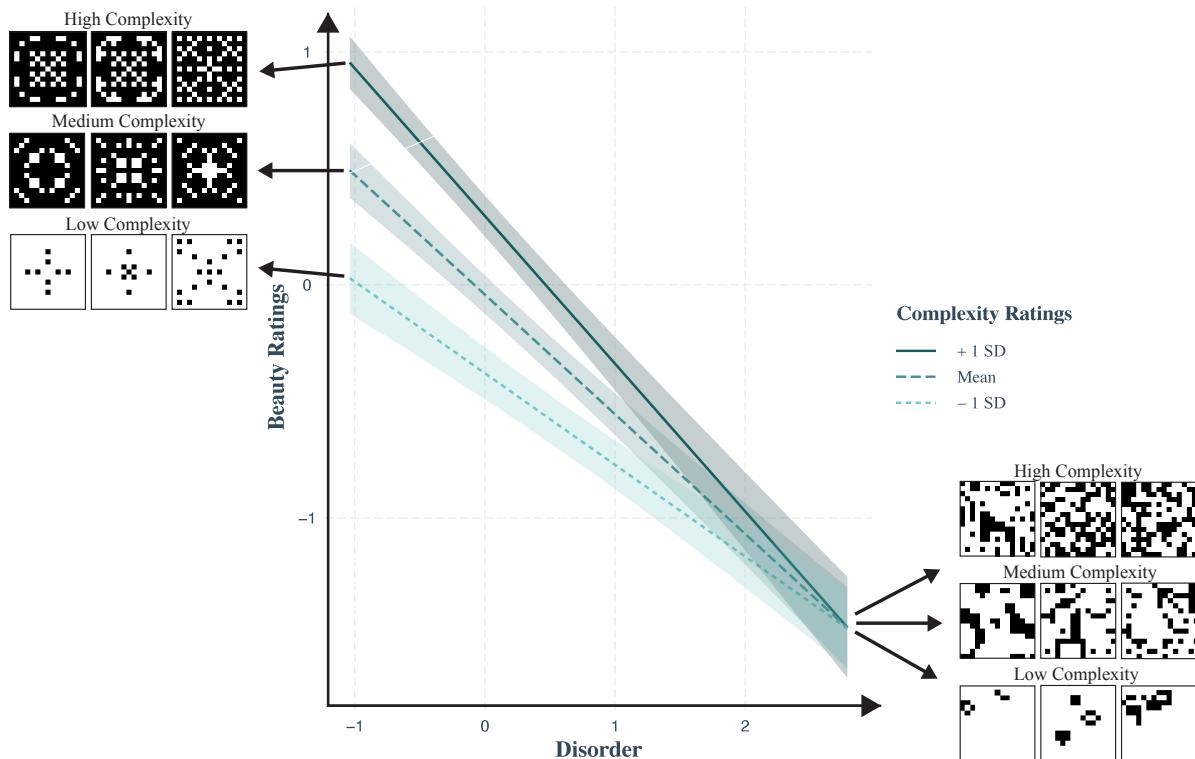
667 *Performance of the best model BR ~ CR \* disorder + (CR + disorder) | Participant*  
668 *performance on (A) training and (B) test data from fold 1 for beauty ratings.*



669 Note. Y-axis displays z-scored beauty ratings and x-axis display their corresponding model  
 670 predictions. Blue dashed line represents  $y = x$ .

671  
 672 Beauty ratings correlated positively with complexity ratings but negatively with disorder (mean  
 673 asymmetry and entropy). Studying the interaction effect (Figure 10) suggested that high  
 674 complexity is considered beautiful as long as the amount of disorder is low; and when the  
 675 disorder is high, beauty is low irrespective of complexity. In other words, people prefer  
 676 complexity (or find it more beautiful) while ensuring order. Further, the random slopes of  
 677 subjective complexity and disorder imply that different people may have different degrees of  
 678 preference for complexity and/or dislike towards disorder. Adding a quadratic effect of  
 679 complexity to the model however, did not lead to a significant performance enhancement  
 680 (Table 2, rows 8-9).

681  
 682 **Figure 10**  
 683 *Visualisation of the interaction effect between complexity ratings and disorder. The*  
 684 *relationship between beauty ratings (y-axis) and disorder (x-axis) is shown at three levels of*  
 685 *complexity ratings (+1 SD, mean and -1 SD). Example patterns at three levels of mean*  
 686 *complexity are shown in the left and right panels. The procedure for obtaining the example*  
 687 *patterns is as follows. 10% (22) patterns (each) with the highest, and lowest disorder are*  
 688 *identified and sorted by mean complexity rating. For the left and right panel examples*  
 689 *respectively, the top 3, mid 3 and bottom 3 patterns of the sorted 10% lowest and highest*  
 690 *disorder patterns are then picked as examples of high, medium and low mean complexity. It is*  
 691 *found that at low disorder, beauty of a pattern is generally low and increases with increasing*  
 692 *complexity. On the other hand, at high disorder, beauty of a pattern is generally low and*  
 693 *decreases with increasing complexity.*



694  
695 In sum, across both analyses, we observe a dissociation between the factors influencing  
696 perceived complexity and beauty. While subjective complexity can be explained by a  
697 combination of local spatial complexity and intricacy, beauty is explained by an interaction  
698 between subjective complexity and disorder.  
699

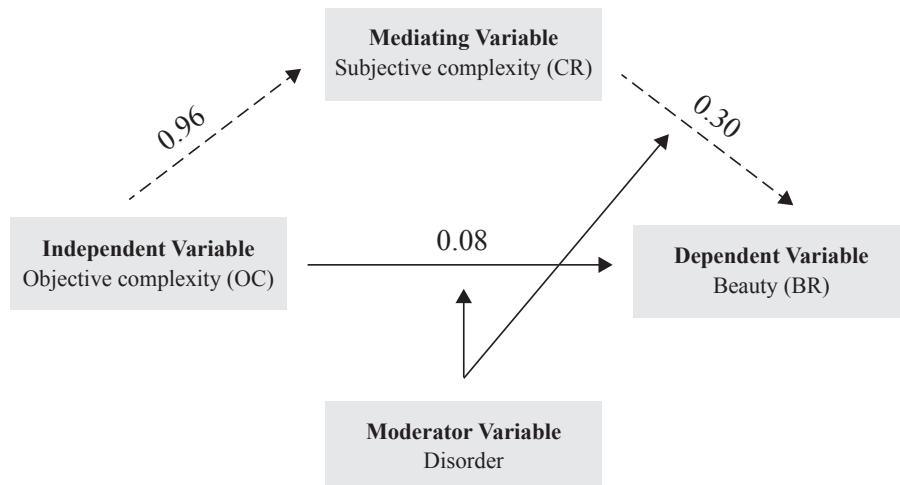
700 3.3 Relationship between Objective Complexity, Subjective Complexity and Beauty –  
701 Moderated Mediation Analysis

702 We also attempted to find an objective model of beauty, without the explicit use of the  
703 complexity ratings. Our previous analysis implies that 60% of the variance in beauty ratings is  
704 explained by a combination of purely objective measures, namely – objective complexity (a  
705 combination of LSC and intricacy, which together predict subjective complexity), and disorder  
706 (asymmetry and entropy) (Table 2, row 12). We examine the relationship between objective  
707 complexity, disorder, subjective complexity and beauty more formally using moderated  
708 mediation analysis.  
709

710 Figure 11 shows the underlying moderated mediation structure of the variables of interest from  
711 our data. We performed moderated mediation analysis using PROCESS macro model 15 in R.  
712 Due to limited data and to avoid overfitting, we only used the average model description (*i.e.*,  
713 excluding random effects) for our analysis. We fit the PROCESS model on data from cross-  
714 validation data fold 2.  
715

716  
717 **Figure 11**

718 *Mediation structure. Objective complexity (OC) interacts with Disorder to predict Beauty*  
 719 *(BR). Subjective complexity (CR) can be predicted by OC. In the presence of CR, OC is not a*  
 720 *significant predictor of BR. Therefore, CR mediates the influence of OC on BR. Further,*  
 721 *disorder did not influence the mediation between OC, CR and BR.*



722  
 723 *Note.* The values represent the coefficients between (1) objective complexity and complexity  
 724 ratings (independent variable and mediating variable), (2) complexity ratings and beauty  
 725 (mediating variable and dependent variable), and the direct effect between objective  
 726 complexity and beauty.

727  
 728 In line with our finding from Figure 10, there was a significant conditional direct effect of  
 729 objective complexity on beauty at low and mean disorder, but a non-significant effect at high  
 730 disorder. There was a significant conditional indirect effect of objective complexity on beauty  
 731 at all levels of disorder. However, the index of moderated mediation was found to be non-  
 732 significant. This implies that the interaction effect with disorder did not influence the mediation  
 733 between objective complexity, subjective complexity and beauty. We studied these direct and  
 734 indirect effects more closely using regression models now including random effects while  
 735 ignoring the interaction effect with disorder. The OC computation here excludes the random  
 736 effects since the random effects structure in the model includes subject specific slope and  
 737 intercepts. We use the cross-validation data fold 1 to fit these models. Table 3 shows the results  
 738 from the regressions.

739

740 **Table 3**

741 *Regressions to study mediation*

S. no.	Model	Significance	OC slope coefficient
1	BR ~ OC + disorder + (OC + disorder) Participant	OC* disorder*	0.38
2	CR ~ OC + OC Participant	OC*	0.96
3	BR ~ CR + OC + disorder + (CR + OC + disorder) Participant	CR* disorder*	0.08

742 Note. BR=beauty ratings, OC=objecive complexity, CR=complexity ratings

743

744 **Table 4**

745 *Results of mediation analysis*

	Estimate	95% CI	p-value
ACME	0.31	[0.22, 0.41]	0.000
ADE	0.08	[-0.02, 0.19]	0.12

746 Note. ACME=average causal mediation effect, ADE=average direct effect

747

748 From Table 3, we find that OC is no longer a significant predictor in the presence of CR, and  
749 the slope coefficient of OC is largely reduced (Table 3, row 3, compared to row 1). To examine  
750 if this mediation effect is significant, we use the *mediate()* function in R under the *mediation*  
751 library. Table 4 summarises the results for Average Causal Mediation Effect (ACME) and the  
752 Average Direct Effect (ADE). We find a significant average causal mediation effect. This  
753 means there is a significant indirect effect of objective complexity on beauty that goes through  
754 the mediator subjective complexity. Further, there is a non-significant direct effect of objective  
755 complexity on beauty ratings. Therefore, we can conclude that the effect of objective  
756 complexity on beauty is mediated by subjective complexity. This implies that subjective  
757 complexity can provide useful information towards the prediction of beauty over and above  
758 what can be explained by objective complexity.

759

760 **4. Discussion**

761 A large body of work has attempted to assess subjective complexity and study its relationship  
762 with beauty, but has been subjected to a fair share of contradictions and an overall lack of  
763 consensus. The incomplete agreement about the relationship between beauty and complexity  
764 in the literature is difficult to resolve due to the predominant use of non-programmatic  
765 measures and hand-crafted stimuli which are difficult to reproduce or manipulate.

766 To address these challenges, we stepped back from the difficulties of using natural scenes to  
767 create a foundation for future investigations based on a very simple class of patterns which  
768 admits algorithmically transparent measures – we used cellular automata, which provided us a  
769 systematic method of algorithmically generating families of 2D binary pixel patterns. Our work  
770 is one of the few studies to use algorithmic stimuli for studying aesthetic judgements aside  
771 from the recently released OCTA Toolbox (Van Geert et al., 2022). The OCTA toolbox helps  
772 create stimuli varying in order and complexity across different dimensions. Alongside, they  
773 also provide a set of programmatic measures for stimuli quantification. There are some major  
774 differences between the toolbox and our methods. The OCTA stimuli can vary across coarse-  
775 grained features such as position, number, shapes, colours, sizes and orientations of objects,  
776 while our CA-generated patterns comprise of finer-level discriminations of pattern properties.  
777 The measures that OCTA supplies correspond mostly to the pre-defined settings that the user  
778 chooses to generate their pattern rather than the pattern as it appears. In contrast, our measures  
779 capture the statistical and spatial properties of patterns as they appear to the observer.

780 Our produced patterns are diverse across multiple dimensions such as, proportion of black  
781 pixels, degree of symmetry, shape, style or number of components, they are similar in nature  
782 to graphic stimuli used previously, for example by Chipman (1977) or Jacobsen and Höffel  
783 (2002). This makes our stimuli comparable to previous studies but achieved using modern  
784 algorithmic methods.

785  
786 We developed a set of six programmatic objective measures for pattern quantification. These  
787 measures included density, entropy, local spatial complexity, Kolmogorov complexity, and  
788 local and global asymmetry. These measures have been considered frequently in past studies  
789 (Arnheim, 1956, 1966; Attneave, 1957; Bense, 1960, 1969; Chikhman et al., 2012; Damiano  
790 et al., 2021; Fan et al., 2022; Friedenberg and Liby, 2016; Gartus and Leder, 2017; Javaheri  
791 Javid, 2016; Moles, 1958; Nadal, 2007; Rigau, 2008; Schmidhuber, 2009; Singh and Shukla,  
792 2017; Silva, 2021; Snodgrass, 1971). We also introduced a novel intricacy measure which  
793 quantified the number of visual components, or groups of same coloured pixels in a pattern  
794 using a graph-based approach. This quantity can be compared to many previous works that  
795 discussed the role of the number of elements (Berlyne et al., 1968; Hall, 1969; Roberts, 2007)  
796 as important factors governing complexity judgement, where elements have implied lines,  
797 intersections or geometric figures. Our intricacy measure presents a new method of quantifying  
798 this factor in pixel patterns.

799  
800 We recorded 80 participants' subjective assessments of complexity and beauty of our patterns.  
801 It is important to note that though our patterns are algorithmically generated, our stimuli set is  
802 partially handcrafted in the selection of patterns to include in the study. However, the selection,  
803 which is mainly for the sake of efficiency, is supported by measures, CA rules and pre-  
804 identified features of patterns (Figure 2) which helps reduce human bias.

805  
806 **4.1 Relationship between Objective and Subjective Complexity**

807  
808 Using linear mixed effects regression, we found that a positive weighted combination of spatial  
809 complexity and intricacy (including random slopes of local spatial complexity (LSC) and  
810 intricacy, and a random intercept of participant) was an effective predictor ( $R^2 = 0.47$ ) of  
811 subjective complexity ratings. This result is consistent with existing results suggesting the  
812 number of elements and their spatial arrangement are good predictors of subjective complexity  
813 (Berlyne, 1960; Berlyne et al., 1968; Nadal, 2007;). Moreover, the result is also in line with the  
814 literature suggesting that two aspects of processing may be involved in complexity perception  
815 – a quantity-based component focussing on the number of visual features and a structure-based  
816 component focussing on the distribution and organization of visual features (Chipman, 1977;  
817 Ichikawa, 1985; Nadal et al. (2010); Van Geert and Wagemans (2020)). Moreover, LSC is a  
818 local property averaged over the entire pattern, whereas intricacy is computed at a global level  
819 using a graphical representation of the entire pattern. Our result implies that these measures  
820 add complementary information which are suitably integrated to give rise to human complexity  
821 evaluations. In terms of performance, our model is comparable, though lower, to results  
822 obtained in other works adopting similar binary pixel/algorithmic stimuli, for example  
823 Chipman (1977, coefficient of multiple correlation = 0.85) or Gartus and Leder (2017,  $R^2_{train} \sim$

824 0.66). The difference can be attributed to (a) the larger size and greater diversity of our stimuli  
825 set and (b) the importance we give to model simplicity and generalisability to unseen data over  
826 pure model fit, *i.e.* we optimise for BIC and  $R^2_{\text{test}}$  over correlation coefficients, AIC or  $R^2_{\text{train}}$ .  
827

828 Contrary to prior work (Arnoult, 1960; Attneave, 1957; Day, 1967; Eisenman and Gellens,  
829 1968; Marin and Leder, 2013; Redies and Brachmann, 2017), neither symmetry nor entropy  
830 related to subjective complexity. This could be explained by the simplified black-and-white  
831 pixel nature of our stimuli as opposed to natural scenes, or the significant correlations between  
832 these measures and intricacy. Further, we observed a disparity between participant strategy  
833 responses and our metric – participants indicated that their ratings depended on their ability to  
834 create or replicate the pattern (which could be seen as a direct link to algorithmic complexity),  
835 while our approximation of Kolmogorov complexity (KC) was not predictive of subjective  
836 complexity as per our model. This could again be due to the large underlying correlations of  
837 KC, asymmetry and entropy with LSC or intricacy, in turn masking their effect.  
838

#### 839 4.2 Relationship between Subjective Complexity and Beauty

840  
841 In contrast, however, asymmetry and entropy did relate to beauty judgements. Beauty ratings  
842 correlated positively with subjective complexity and negatively with asymmetry and entropy.  
843 More specifically, the conventional single-scale entropy measure performed better than  
844 multiscale entropy suggesting that beauty perception in such patterns is explained better by  
845 global features (Table AIII.2). Through this, our work has lent support for a monotonic  
846 relationship between beauty and complexity. This goes against the inverted-U like dependence  
847 proposed by Berlyne. However, one reason for this, as stated above and by others (Krupinski  
848 and Locher, 1988; Nicki, Lee, and Moss, 1981; Stamps III, 2002), could be that our stimuli are  
849 so simple in nature as to lie in the lower quantiles of complexity. If true, then we would only  
850 expect to be able to reproduce the first half of the inverted-U curve, and as a result could not  
851 falsify a linear relationship.  
852

853 A weighted combination of complexity ratings, disorder (itself a weighted combination of  
854 asymmetry and entropy) and their interaction (along with random intercepts for participants  
855 and random slopes for subjective complexity and disorder) effectively modelled beauty ratings.  
856 This concurs with the work proposing that beauty lies in the balance between order and  
857 complexity (refer to Van Geert and Wagemans, 2020 for a review). For example, Arnheim  
858 (1966, p. 124) stated: “Complexity without order produces confusion. Order without  
859 complexity causes boredom”. However, based on our linear model (with interactions), we  
860 cannot lend support to Birkhoff’s (1933) proposed  $M = O / C$  relationship, or Eysenck’s (1941,  
861 1942)  $M = O \times C$  relationship. The interaction effect we found emphasises the relative  
862 influence of order and complexity on beauty – complexity is beautiful, as long as the degree of  
863 disorder is low, and at high disorder, beauty is consistently low, irrespective of complexity. In  
864 other words, order is necessary for beauty and complexity adds to beauty once order is present.  
865 Here we can return to our initial question about why geometric tile designs are more beautiful  
866 than chess boards or QR-codes—tile designs are more beautiful than chess boards as they are  
867 more complex, but more beautiful than QR-codes as they are less disordered. This result is at

868 odds with Van Geert and Wagemans' (2021) suggestion that a balance between order and  
869 complexity involves no interaction. They, however, used real world images of neatly organized  
870 compositions and recorded aesthetic preferences using a 2-choice task. They further collected  
871 complexity and order ratings for the images, which were found to be uncorrelated. The  
872 contradiction between their and our findings underlines the challenge of comparing results  
873 across varying stimuli types and task designs and stresses the fact that the added value of the  
874 interaction term can depend on the stimuli and the specific operationalizations of order and  
875 complexity used (Van Geert and Wagemans, 2020).

876

#### 877 4.3 Relationship between Objective Complexity, Subjective Complexity and Beauty

878

879 While a combination of pure objective measures of objective complexity (LSC and intricacy),  
880 and disorder (asymmetry and entropy) was able to explain 60% of the variance in beauty  
881 ratings, formal analysis of the relationship between subjective complexity, objective  
882 complexity and beauty using moderated mediation analysis revealed that subjective complexity  
883 mediates the influence of objective complexity on beauty at all levels of disorder. This indicates  
884 that subjective complexity encodes information beyond what is expressed in terms of objective  
885 complexity measures. For this reason, some views criticize such methods attempting to  
886 quantify subjective complexity in objective terms. Heckhausen (1964) argued that relating  
887 subjective complexity to simple visual properties of stimuli as done by information theory  
888 approaches is insufficient. He claimed that the subjective complexity does not solely depend  
889 on the complexity of the stimulus but also on the way it is perceived. Attneave (1957) also  
890 suggested that people's perception of complexity is not a mere reflection of the visual stimuli.  
891 This explanation aligns with Gestalt philosophies of "perceptual organisation" often  
892 summarised as "whole is greater than sum of the parts". This need was also highlighted by  
893 Berlyne who had claimed that complexity was a property of both the physical stimulus  
894 properties and the processes within the subject. Having said that, one must clarify here that  
895 using only the average model description in the PROCESS moderated mediation analysis may  
896 have removed the effect of disorder's influence on the relation between complexity and beauty  
897 for different people. Further, no causal implications can be made from this moderated  
898 mediation analysis, and the relationship between subjective and objective quantities observed  
899 here are purely correlational. Our work is also restricted just to *how* complexity relates to  
900 beauty without looking into *why* complexity relates to beauty the way it does.

901

#### 902 4.4 Limitations and Future Directions

903

904 Our methods have limitations. Since the beauty rating was recorded along with the complexity  
905 rating, there might have been an anchoring effect which could have yielded a spurious  
906 correlation. Also, our dataset size was found to be insufficient to fit some models with a larger  
907 number of parameters (for example the complexity model with both LSC and intricacy random  
908 slopes).

909

910 Further, there are limitations on the generality associated with the specific type of stimuli. Our  
911 measures are defined specifically for 2D binary pixel patterns. While our CA generation

912 algorithm can be manipulated to produce newer families of patterns with increased size or  
913 added colours, and our measures can be adapted to apply to patterns with such transformations,  
914 this would only achieve small-scale generalisation. There are certain advantages of testing for  
915 small-scale generalisation – it provides a systematic method to identify causes for the  
916 breakdown of generalisation. For example, if the complexity metric fails to explain variance  
917 upon adding a third colour, we can isolate the factor of colour, across which generalisation  
918 does not hold. Explicit tests for this have not been addressed in this article and would be a part  
919 of future work.

920  
921 However, even if our patterns and measures generalize in the above mentioned ways, they are  
922 far from large-scale generalization to naturalistic stimuli or artworks for several reasons: (1)  
923 their statistics are very different from those of natural scenes, (2) they are overly simplistic,  
924 allowing for only limited colours in a restricted grid size and being devoid of overt semantic  
925 content, and (3) the nature of these pixel patterns makes it hard to define several popular  
926 measures of complexity such as number of vertices, edges, lines, or curves. Such a choice of  
927 stimuli therefore renders some of the prevalent contradictions regarding suitable objective  
928 measures irrelevant because these measures do not apply. This raises the concern as to whether  
929 the complexity measure we arrived at remains correct in more ecologically valid settings. As  
930 mentioned above, the definitions of LSC and intricacy can readily be extended to larger stimuli  
931 with more colours easily, but it will be necessary to study explicitly how predictive they are of  
932 subjective complexity for richer stimuli. Modern methods such as diffusion models for  
933 producing photorealistic images could be used as programmatic generators of images of  
934 potentially varying subjective complexity. Equally, convolutional neural networks could be  
935 used as feature extractors in place of our manually defined complexity measures (for example,  
936 Iigaya et al., 2020). Using such methods may, however, come at the expense of losing  
937 interpretability. Finding a middle ground would be another important focus of our future work.  
938

939 Finally, we note that across both complexity and beauty judgements, the within-participant  
940 ratings were highly consistent for repeats which indicates that our stimuli are able to elicit  
941 robust subjective judgments in participants. However, there were between-participant  
942 differences and a large amount of the variation was estimated by random effects – we saw large  
943 performance gains from adding random slopes of LSC and intricacy in our complexity models,  
944 and random slopes of subjective complexity and disorder in our beauty models per participant  
945 (refer to Appendix AIII for plots of the random effects). A more thorough analysis of individual  
946 differences is another important target for future work.

947  
948 4.5 Conclusion  
949  
950 Our work develops a class of diverse, algorithmic 2D binary pixel patterns which can be  
951 reproduced and manipulated easily. A set of programmatic pattern quantification measures  
952 were used to understand the relationship between objective complexity, subjective complexity,  
953 and beauty. We found that people’s complexity ratings depend on the local arrangement of  
954 pixels along with the global number of components, and that there is beauty in complexity as  
955 long as there is sufficient order. We also noted that subjective complexity cannot be explained

956 fully with objective complexity measures. Through this, our work showcases the usefulness of  
957 computational methods to understand the link between assessments of complexity and beauty.  
958

## 959 References

960  
961 Adamatzky, A., & Martínez, G. J. (Eds.). (2016). *Designing beauty: the art of cellular*  
962 *automata* (Vol. 20). Springer.

963 Adkins, O. C., & Norman, J. F. (2016). The visual aesthetics of snowflakes. *Perception*, 45(11),  
964 1304-1319.

965 Aitken, P. P. (1974). Judgments of pleasingness and interestingness as functions of visual  
966 complexity. *Journal of Experimental Psychology*, 103(2), 240.

967 Arnheim, R. (1956). *Art and visual perception: A psychology of the creative eye*. Univ of  
968 California Press.

969 Arnheim, R. (1966). Towards a psychology of art/entropy and art an essay on disorder and  
970 order. *The Regents of the University of California*, 160.

971 Arnoult, M. D. (1960). Prediction of perceptual responses from structural characteristics of the  
972 stimulus. *Perceptual and Motor Skills*, 11(3), 261-268.

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

975 Authors. (2022, August). Spatial complexity and intricacy predict subjective complexity.  
976 In *The Visual Science of Art Conference (VSAC 2022)*.

977 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models  
978 Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

979 Bense, M. (1960). Programmierung des Schönen Allgemeine Texttheorie Und Textästhetik.

980 Bense, M. (1969). *Kleine abstrakte ästhetik*. Edition Rot.

981 Berlyne, D. E. (1960). Conflict, arousal, and curiosity.

982 Berlyne, D. E. (1967). Arousal and reinforcement. In *Nebraska symposium on motivation*.  
983 University of Nebraska Press.

984 Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. (1968). The dimensionality of visual  
985 complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue*  
986 *canadienne de psychologie*, 22(5), 376.

987 Berlyne, D. E. (1973). Aesthetics and psychobiology. *Journal of Aesthetics and Art Criticism*,  
988 31(4).

989

- 1004 Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like  
1005 curvature or do they dislike angularity?. *British Journal of Psychology*, 107(1), 154-178.
- 1006
- 1007 Birkhoff, G. D. (1933). Aesthetic Measure.
- 1008
- 1009 Birkin, G. (2010). *Aesthetic complexity: practice and perception in art & design*. Nottingham  
1010 Trent University (United Kingdom).
- 1011
- 1012 Briellmann, A. A., & Dayan, P. (2022). A computational model of aesthetic value.  
1013 *Psychological Review*.
- 1014
- 1015 Chamberlain, R. (2022). The interplay of objective and subjective factors in empirical  
1016 aesthetics. In *Human Perception of Visual Information* (pp. 115-132). Springer, Cham.
- 1017
- 1018 Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity  
1019 of images: Experimental and computational estimates compared. *Perception*, 41(6), 631-647.
- 1020
- 1021 Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental  
1022 Psychology: General*, 106(3), 269.
- 1023
- 1024 Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on  
1025 complexity judgments in children and adults. *Journal of Experimental Psychology: Human  
1026 Perception and Performance*, 5(2), 365.
- 1027
- 1028 Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of  
1029 the literature. *Psychology of Music*, 45(6), 886-909.
- 1030
- 1031 Corchs, S. E., Ciocca, G., Bricolo, E., & Gasparini, F. (2016). Predicting complexity perception  
1032 of real world images. *PloS one*, 11(6), e0157986.
- 1033
- 1034 Cupchik, G. C. (1986). A decade after Berlyne: New directions in experimental aesthetics.  
1035 *Poetics*, 15(4-6), 345-369.
- 1036
- 1037 Damiano, C., Wilder, J., Zhou, E. Y., Walther, D. B., & Wagemans, J. (2021). The role of local  
1038 and global symmetry in pleasure, interest, and complexity judgments of natural  
1039 scenes. *Psychology of Aesthetics, Creativity, and the Arts*.
- 1040
- 1041 Davis, R. C. (1936). An evaluation and test of Birkhoff's aesthetic measure formula. *The  
1042 Journal of General Psychology*, 15(2), 231-240.
- 1043
- 1044 Day, H. Y. (1967). Evaluations of subjective complexity, pleasingness and interestingness for  
1045 a series of random polygons varying in complexity. *Perception & Psychophysics*, 2(7), 281-  
1046 286.
- 1047
- 1048 Day, H. (1968). The importance of symmetry and complexity in the evaluation of complexity,  
1049 interest and pleasingness. *Psychonomic Science*, 10(10), 339-340.
- 1050
- 1051 Delplanque, J., De Loof, E., Janssens, C., & Verguts, T. (2019). The sound of beauty: How  
1052 complexity determines aesthetic preference. *Acta Psychologica*, 192, 146-152.

- 1053  
1054 Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity.  
1055 *Perception*, 35(6), 823-835.
- 1056  
1057 Eisenman, R. (1967). Complexity-simplicity: I. Preference for symmetry and rejection of  
1058 complexity. *Psychonomic Science*, 8(4), 169-170.
- 1059  
1060 Eisenman, R., & Gellens, H. K. (1968). Preferences for complexity-simplicity and symmetry-  
1061 asymmetry. *Perceptual and Motor Skills*, 26(3), 888-890.
- 1062  
1063 Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological  
1064 Review*, 48(1), 83.
- 1065  
1066 Eysenck, H. J. (1942). The experimental study of the'good Gestalt'—a new approach.  
1067 *Psychological Review*, 49(4), 344.
- 1068  
1069 Eysenck, H. J. (1968). An experimental study of aesthetic preference for polygonal figures.  
1070 *The Journal of General Psychology*, 79(1), 3-17.
- 1071  
1072 Fechner, G. T. (1876). *Vorschule der aesthetik* (Vol. 1). Breitkopf & Härtel.
- 1073  
1074 Fan, Z. B., Li, Y. N., Zhang, K., Yu, J., & Huang, M. L. (2022). Measuring and evaluating the  
1075 visual complexity of Chinese ink paintings. *The Computer Journal*, 65(8), 1964-1976.
- 1076  
1077 Farley, F. H., & Weinstock, C. A. (1980). Experimental aesthetics: Children's complexity  
1078 preference in original art and photoreproductions. *Bulletin of the Psychonomic Society*, 15(3),  
1079 194-196.
- 1080  
1081 Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of  
1082 complexity and familiarity in basic-level picture processing. *Behavior research methods*,  
1083 40(1), 116-129.
- 1084  
1085 Friedenberg, J., & Liby, B. (2016). Perceived beauty of random texture patterns: A preference  
1086 for complexity. *Acta psychologica*, 168, 41-49.
- 1087  
1088 Gardner, M. (1970). The Fantastic Combinations of John Conway's New Solitaire Game'Life.  
1089 *Sc. Am.*, 223, 20-123.
- 1090  
1091 Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of  
1092 broken symmetries. *I-Perception*, 4(5), 361-364.
- 1093  
1094 Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns  
1095 using computational measures: The influence of mirror symmetry on complexity  
1096 perception. *PloS one*, 12(11), e0185276.
- 1097  
1098 Giannouli, V. (2013). Visual symmetry perception. *Encephalos*, 50, 31-42.
- 1099  
1100 Gordon, J., & Gridley, M. C. (2013). Musical preferences as a function of stimulus complexity  
1101 of piano jazz. *Creativity Research Journal*, 25(1), 143-146.

- 1102  
1103 Güçlütürk, Y., Jacobs, R. H., & Lier, R. V. (2016). Liking versus complexity: Decomposing  
1104 the inverted U-curve. *Frontiers in Human Neuroscience*, 10, 112.  
1105  
1106 Hall, A. C. (1969). Measures of the complexity of random black and white and coloured  
1107 stimuli. *Perceptual and motor skills*, 29(3), 773-774.  
1108  
1109 Harsh, C. M., Beebe-Center, J. G., & Beebe-Center, R. (1939). Further evidence regarding  
1110 preferential judgment of polygonal forms. *The Journal of Psychology*, 7(2), 343-350.  
1111  
1112 Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.  
1113  
1114 Heath, T., Smith, S. G., & Lim, B. (2000). Tall buildings and the urban skyline: The effect of  
1115 visual complexity on preferences. *Environment and behavior*, 32(4), 541-556.  
1116  
1117 Heckhausen, H. (1964). Complexity in perception: Phenomenal criteria and information  
1118 theoretic calculus—A note on D. Berlynes” Complexity Effects.”.  
1119  
1120 Hekkert, P., & Van Wieringen, P. C. (1990). Complexity and prototypicality as determinants  
1121 of the appraisal of cubist paintings. *British journal of psychology*, 81(4), 483-495.  
1122  
1123 Ichikawa, S. (1985). Quantitative and structural factors in the judgment of pattern complexity.  
1124 *Perception & psychophysics*, 38(2), 101-109.  
1125  
1126 Igaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., & O'Doherty, J. P. (2020). Aesthetic preference  
1127 for art emerges from a weighted integration over hierarchically structured visual features in the  
1128 brain. *BioRxiv*.  
1129  
1130 Jacobsen, T. (2010). Beauty and the brain: culture, history and individual differences in  
1131 aesthetic appreciation. *Journal of anatomy*, 216(2), 184-191.  
1132  
1133 Jacobsen, T., & Höfel, L. E. A. (2002). Aesthetic judgments of novel graphic patterns:  
1134 Analyses of individual judgments. *Perceptual and motor skills*, 95(3), 755-766.  
1135  
1136 Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of  
1137 aesthetic judgment of beauty. *Neuroimage*, 29(1), 276-285.  
1138  
1139 Javaheri Javid, M. A., Blackwell, T., Zimmer, R., & Al-Rifaie, M. M. (2016, March).  
1140 Correlation between human aesthetic judgement and spatial complexity measure.  
1141 In *International Conference on Computational Intelligence in Music, Sound, Art and*  
1142 *Design* (pp. 79-91). Springer, Cham.  
1143  
1144 Javaheri Javid, M. A. (2019). *Aesthetic Automata: Synthesis and Simulation of Aesthetic*  
1145 *Behaviour in Cellular Automata* (Doctoral dissertation, Goldsmiths, University of London).  
1146  
1147 Javaheri Javid, M. A. (2021, April). Aesthetic Evaluation of Cellular Automata Configurations  
1148 Using Spatial Complexity and Kolmogorov Complexity. In *International Conference on*  
1149 *Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 147-160).  
1150 Springer, Cham.  
1151

- 1152  
1153 Krupinski, E., & Locher, P. (1988). Skin conductance and aesthetic evaluative responses to  
1154 nonrepresentational works of art varying in symmetry. *Bulletin of the Psychonomic Society*,  
1155 26(4), 355-358.  
1156  
1157 Lakhal, S., Darmon, A., Bouchaud, J. P., & Benzaquen, M. (2020). Beauty and structural  
1158 complexity. *Physical Review Research*, 2(2), 022058.  
1159  
1160 Langton, C. G. (1986). Studying artificial life with cellular automata. *Physica D: Nonlinear*  
1161 *Phenomena*, 22(1-3), 120-149.  
1162  
1163 Li, W., Packard, N. H., & Langton, C. G. (1990). Transition phenomena in cellular automata  
1164 rule space. *Physica D: Nonlinear Phenomena*, 45(1-3), 77-94.  
1165  
1166 Machotka, P. (1980). Daniel Berlyne's contributions to empirical aesthetics. *Motivation and*  
1167 *Emotion*, 4(2), 113-121.  
1168  
1169 Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music  
1170 regardless of its complexity: Implications for the appreciation and aesthetics of music.  
1171 *Frontiers in neuroscience*, 11, 147.  
1172  
1173 Marin, M. M., & Leder, H. (2013). Examining complexity across domains: relating subjective  
1174 and objective measures of affective environmental scenes, paintings and music. *PloS one*, 8(8),  
1175 e72412.  
1176  
1177 Marin, M. M., Lampatz, A., Wandl, M., & Leder, H. (2016). Berlyne revisited: Evidence for  
1178 the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers*  
1179 *in human neuroscience*, 10, 536.  
1180  
1181 McWhinnie, H. J. (1971). A review of selected aspects of empirical aesthetics III. *Journal of*  
1182 *Aesthetic Education*, 5(4), 115-126.  
1183  
1184 Messinger, S. M. (1998). Pleasure and complexity: Berlyne revisited. *The Journal of*  
1185 *Psychology*, 132(5), 558-560.  
1186  
1187 Moles, A. A. (1958). Theorie de l'information et perception esthetique.  
1188  
1189 Munsinger, H., & Kessen, W. (1964). Uncertainty, structure, and preference. *Psychological*  
1190 *Monographs: General and Applied*, 78(9), 1.  
1191  
1192 Nadal, M., & Vartanian, O. (2021). Empirical Aesthetics: An overview.  
1193  
1194 Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty  
1195 appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173-  
1196 191.  
1197  
1198 Nasar, J. L. (2002). What design for a presidential library? Complexity, typicality, order, and  
1199 historical significance. *Empirical Studies of the Arts*, 20(1), 83-99.  
1200

- 1201 Nicki, R. M. (1972). Arousal increment and degree of complexity as incentive. *British Journal*  
1202 *of Psychology*, 63(2), 165-171.
- 1203
- 1204 Nicki, R. M., & Moss, V. (1975). Preference for non-representational art as a function of  
1205 various measures of complexity. *Canadian Journal of Psychology/Revue 37hannon3737e de*  
1206 *psychologie*, 29(3), 237.
- 1207
- 1208 Nicki, R. M., & Gale, A. (1977). EEG, measures of complexity, and preference for  
1209 nonrepresentational works of art. *Perception*, 6(3), 281-286.
- 1210
- 1211 Nicki, R. M., Lee, P. L., & Moss, V. (1981). Ambiguity, cubist works of art, and preference.  
1212 *Acta Psychologica*, 49(1), 27-41.
- 1213
- 1214 Norman, J. F., Beers, A., & Phillips, F. (2010). Fechner's Aesthetics Revisited. *Seeing and*  
1215 *Perceiving*, 23(3), 263-271.
- 1216
- 1217 Osborne, J. W., & Farley, F. H. (1970). The relationship between aesthetic preference and  
1218 visual complexity in 37hannon37 art. *Psychonomic Science*, 19(2), 69-70.
- 1219
- 1220 Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human  
1221 preference. *Annual review of psychology*, 64, 77-107.
- 1222
- 1223 Redies, C., Brachmann, A., & Wagemans, J. (2017). High entropy of edge orientations  
1224 characterizes visual artworks from diverse cultural backgrounds. *Vision Research*, 133, 130-  
1225 144.
- 1226
- 1227 Rigau, J., Feixas, M., & Sbert, M. (2008). Informational aesthetics measures. *IEEE computer*  
1228 *graphics and applications*, 28(2), 24-34.
- 1229
- 1230 Roberts, M. N. (2007). Complexity and aesthetic preference for diverse visual stimuli.  
1231 *Doctoral), Universitat de les Illes Balears, Palma, Spain.*
- 1232
- 1233 Schmidhuber, J. (2009). Art & science as by-products of the search for novel patterns, or data  
1234 compressible in unknown yet learnable ways. *Multiple ways to design research. Research*  
1235 *cases that reshape the design discipline*, Swiss Design Network-Et al. Edizioni, 98-112.
- 1236
- 1237 Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical*  
1238 *journal*, 27(3), 379-423.
- 1239
- 1240 Silva, J. M., Pratas, D., Antunes, R., Matos, S., & Pinho, A. J. (2021). Automatic analysis of  
1241 artistic paintings using information-based measures. *Pattern Recognition*, 114, 107864.
- 1242
- 1243 Singh, S., & Shukla, D. (2017). A review on various measures for finding image complexity.  
1244 *International Journal of Scientific Research Engineering & Technology. ISSN*, 2278-0882.
- 1245
- 1246 Snodgrass, J. G. (1971). Objective and subjective complexity measures for a new population  
1247 of patterns. *Perception & Psychophysics*, 10(4), 217-224.
- 1248

- 1249 Stamps III, A. E. (2002). Entropy, visual diversity, and preference. *The Journal of general*  
1250 *psychology*, 129(3), 300-320.
- 1251
- 1252 Sun, Z., & Firestone, C. (2022). Beautiful on the inside: Aesthetic preferences and the skeletal  
1253 complexity of shapes. *Perception*, 1, 15.
- 1254
- 1255 Taylor, R. E., & Eisenman, R. (1964). Perception and production of complexity by creative art  
1256 students. *The Journal of Psychology*, 57(1), 239-242.
- 1257
- 1258 Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package  
1259 for causal mediation analysis.
- 1260
- 1261 Tinio, P. P., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry,  
1262 complexity, and the jaws of massive familiarization. *Acta psychologica*, 130(3), 241-250.
- 1263
- 1264 Tran, N. H., Waring, T., Atmaca, S., & Beheim, B. A. (2021). Entropy trade-offs in artistic  
1265 design: A case study of Tamil kolam. *Evolutionary Human Sciences*, 3.
- 1266
- 1267 Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation.  
1268 *Psychology of aesthetics, creativity, and the arts*, 14(2), 135.
- 1269
- 1270 Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly  
1271 organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, 15(3), 484.
- 1272
- 1273 Van Geert, E., Bossens, C., & Wagemans, J. (2022). The Order & Complexity Toolbox for  
1274 Aesthetics (OCTA): A systematic approach to study the relations between order, complexity,  
1275 and aesthetic appreciation. *Behavior Research Methods*, 1-24.
- 1276
- 1277 Van Lier, R., Van Der Helm, P., & Leeuwenberg, E. (1994). Integrating global and local  
1278 aspects of visual occlusion. *Perception*, 23(8), 883-903.
- 1279
- 1280 Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral science*,  
1281 11(2), 105-114.
- 1282
- 1283 Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis  
1284 of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(3), 392-  
1285 399.
- 1286
- 1287 William, A. Huber (2011). Measuring entropy/ information/ patterns of a 2d binary matrix.
- 1288
- 1289 Wolfram, S. (2002). A new kind of science, Wolfram Media, Inc. *Champaign, IL*.
- 1290
- 1291 Zenil, H., Hernández-Orozco, S., Kiani, N. A., Soler-Toscano, F., Rueda-Toicen, A., & Tegnér,  
1292 J. (2018). A decomposition method for global evaluation of 38hannon entropy and local  
1293 estimations of algorithmic complexity. *Entropy*, 20(8), 605.
- 1294
- 1295 Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding.  
1296 *IEEE transactions on Information Theory*, 24(5), 530-536.

## 1297 Appendix I – Cellular Automata Pattern Generation

1298  
1299 The number of potential state-transition update functions for a 2D CA with n states and an N-  
1300 cell neighbourhood is  $n^{n^N}$ . Hence, for our 2D binary CA with 5-cell and 9-cell neighbourhoods,  
1301 the number of possible state-transition update functions are  $2^{2^5} = 2^{32} = 4 \times 10^9$  and  $2^{2^9} =$   
1302  $2^{512} = 10^{152}$  respectively. Moreover, the total number of initial configurations for an n-state  
1303  $P \times Q$  grid is  $n^{P \times Q}$ . In our binary  $15 \times 15$  grid, this would be  $2^{15 \times 15} = 2^{225}$  possible initial  
1304 configurations. Since these are very large spaces, we consider simplified versions of rules  
1305 (Wolfram, 1983). The rules are set such that the state of a cell depends only on the sum of the  
1306 states of cells in its neighbourhood (SCN). Such rules can be of two types: “totalistic” (Tot)  
1307 where the state of the cell  $(i, j)$  at time  $t+1$  depends only on SCN at time t, or “outer-totalistic”  
1308 (Otot) where the state of the cell  $(i, j)$  at time  $t+1$  depends on both SCN and the value of cell  $(i,$   
1309  $j)$  at time t. These rules can be expressed as decimal “rule codes” given by  $\sum_{i=0}^N (f(i) \times 2^i)$  for  
1310 totalistic rules and  $\sum_{i=0}^N \sum_{j=0}^1 (f(i, j) \times 2^{2i+j})$  for outer-totalistic rules.

1311  
1312 It is a well-known fact that unlike 1D/elementary CA, the limiting behaviour of 2D CA is  
1313 undecidable. Patterns might terminate, oscillate, tend towards full randomness or tend towards  
1314 order depending on the combination of algorithm parameters. It is also not possible to reverse-  
1315 engineer rules that result in particular types of pattern outputs. Therefore, we referred to work  
1316 that has elucidated some rules with their evolving behaviours (Packard and Wolfram, 1985;  
1317 Wolfram, 2002) while selecting our set of rules and fixing the number of iterations based on  
1318 our grid size to  $T = 40$ .

1319  
1320 The generation script was coded in Python (Version 3.8.8). For each cell, either the 5-cell or a  
1321 9-cell neighbourhood was determined and the sum of the SCN was computed. The grid was  
1322 wrapped around such that the cells on the boundary considered cells on the opposite boundary  
1323 as neighbours. The updated state of the cell was obtained as a function of the computed sum as  
1324 indicated by the rule code read in binary. The pseudocode is given here.

### 1325 1326 ALGORITHM: Cellular Automata Pattern Generation

**Input:**

*Rule code: integer,*  
*N: integer,*  
*Grid with IC, G: 2D array,*  
*Rule type (tot/Otot): string,*  
*T: integer*

**Output:** 8 patterns (one every 5<sup>th</sup> iteration)

- 1 *Bin\_rule\_code*  $\leftarrow$  binary form of decimal rule code
- 2 *Powers*  $\leftarrow$  list of indices where *bin\_rule\_code* = 1 when read in reverse
- 3 **For** timestep  $t = 1$  to  $T$
- 4     *G'*  $\leftarrow$  Copy of *G*
- 5     **For** each cell  $(i, j)$  in *G'*
- 6         **If**  $N = 5$

```

7      Neigh  $\leftarrow$  list of neighbourhood cells =  $[(i,j), (i+1,j), (i,j+1), (i-1,j), (i,j-1)]$ 
8      Else if  $N = 9$ 
9          Neigh  $\leftarrow$  list of neighbourhood cells =  $[(i,j), (i+1,j), (i,j+1), (i-1,j), (i,j-1), (i-1,j-1), (i-1,j+1), (i+1,j-1), (i+1,j+1)]$ 
10     End
11     Neigh_sum  $\leftarrow$  sum of the states of the cells in Neigh
12     If rule type is "tot"
13         New_state  $\leftarrow$  1 if Neigh_sum in Powers, else 0
14     Else if rule type is "Otot"
15         New_state  $\leftarrow$  1 if  $2 \times$  Neigh_sum + G[i,j] in Powers, else 0
16     End
17     G'[i,j] = New_state
18     End
19     G  $\leftarrow$  Copy of G'
20     If  $t \% 5 = 0$ 
21         Save G
22 End

```

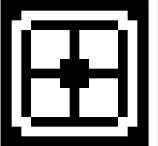
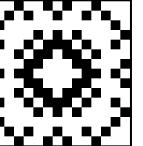
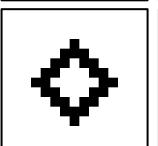
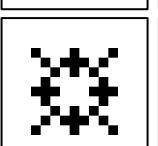
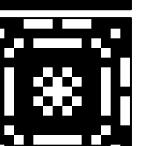
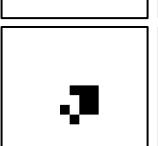
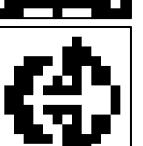
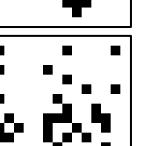
1327

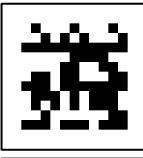
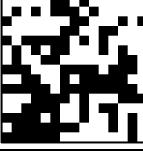
1328 Some example rules used by us along with the produced patterns are listed in Figure AI.1.

1329

### 1330 Figure AI.1.

1331 *Rules used for cellular automata*

S. no.	Rule code	Neighbourhood size	Rule (Tot/Otot)	type	IC	Pattern (iteration 5, 20)
1	451	5	Otot		1	 
2	510	5	Otot		1	 
3	15822	9	Otot		1	 
4	736	9	Otot		2	 
5	85507	5	Otot		2	 

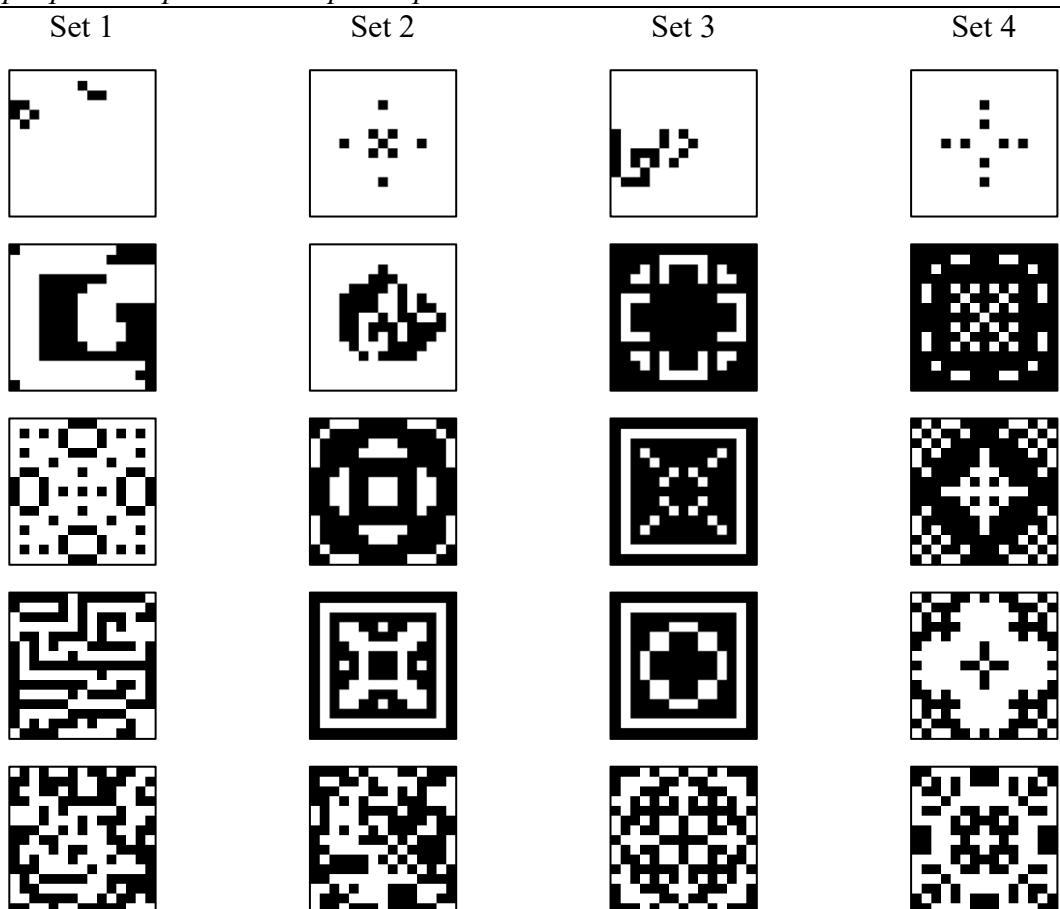
6	15822	9	Otot	2		
7	736	9	Otot	3		
8	196623	9	Otot	3		
9	52	5	Tot	3		

1332

1333 Figure AI.2 displays example patterns presented to participants.

**Figure AI.2.**

1335 *Example patterns presented to participants in the 4 sets*



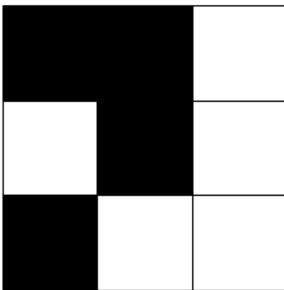
1336

## 1337 Appendix II – Objective Complexity Measures

1338  
1339 Here we provide a detailed description of the measures defined in Section 2.2 along with  
1340 illustrated examples using a simple pattern (Figure AII.1).

### 1341 1342 **Figure AII.1**

1343 *Example pattern to show measure computations.*



1344  
1345 Note: The pattern shown is same as the one used to show intricacy computation in Figure 3.

### 1346 1347 **Density**

1348 The density of the pattern is the proportion of black pixels.

1349 Here this would be,

1350

### 1351 1352 **Figure AII.2**

1353 *Density computation*

$$\frac{\text{Number of } \blacksquare}{\text{Number of } \blacksquare + \text{Number of } \square} = \frac{4}{4+5} = 0.44$$

1354

### 1355 **Entropies**

1356

- 1357 1. Entropy: Entropy is calculated at a single scale as given by Eq. 1. While  $P(b)$  is simply  
1358 the density from above,  $P(w)$  is  $(1 - \text{density})$  as  $P(b) + P(w) = 1$  for our patterns.

1359

1360 Therefore, here, entropy is:

1361  $-0.44 \log_2 0.44 - (1 - 0.44) \log_2 (1 - 0.44) = 0.52 + 0.46 = 0.98$

1362

- 1363 2. Multiscale entropy: Multiscale entropy averages entropy calculated at different scale  
1364 for different windows as illustrated in Figure AII.3.

1365

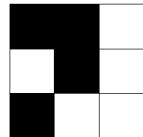
1366 Number of scales = Size of the grid = 3

1367  
1368  
1369  
1370  
1371

**Figure AII.3**  
*Multiscale entropy computation*

1372  
1373  
1374  
1375  
1376

3 × 3 windows:



Entropy = 0.98

Entropy at this scale = 0.98

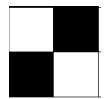
2 × 2 windows:



Entropy = 0.81



Entropy = 1



Entropy = 1



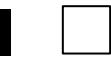
Entropy = 0.81

1377 Mean entropy at this scale = 0.91

1378

1 × 1 windows:

1380



1381

Entropy for all = 0

1382

Mean entropy at this scale = 0

1383

$$1384 \quad \text{Multiscale entropy} = \frac{0.98+0.91+0}{3} = 0.63$$

1385

**Local spatial complexity**

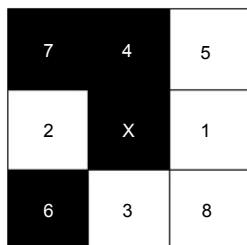
1386

1387 The LSC computations are shown in Table AII.1 for the Figure AII.1 as labelled in AII.5.

1388

**Figure AII.5**

1389 Example pattern with directions labelled for LSC computation.



1393

1394

1395

**Table AII.1**

1396 LSC computations for the example pattern

1397

Direction	States	$P(s_1, s_2)_d$	$P(s_1 s_2)_d$	$P(s_1, s_2)\log_2 P(s_1 s_2)_d$
1	$s_1 = 0, s_2 = 0$ (white, white)	0.16	0.25	0.33
	$s_1 = 0, s_2 = 1$ (white, black)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 0$ (black, white)	0.5	0.75	0.20
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.5	0.16
	Sum = 0.87			
2	$s_1 = 0, s_2 = 0$ (white, white)	0.16	0.5	0.16
	$s_1 = 0, s_2 = 1$ (white, black)	0.5	0.75	0.20
	$s_1 = 1, s_2 = 0$ (black, white)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.25	0.33
	Sum = 0.87			
3	$s_1 = 0, s_2 = 0$ (white, white)	0.33	0.5	0.33
	$s_1 = 0, s_2 = 1$ (white, black)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 0$ (black, white)	0.33	0.5	0.33
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.5	0.16
	Sum = 0.87			
4	$s_1 = 0, s_2 = 0$ (white, white)	0.33	0.66	0.19
	$s_1 = 0, s_2 = 1$ (white, black)	0.33	0.66	0.19
	$s_1 = 1, s_2 = 0$ (black, white)	0.16	0.33	0.26
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.33	0.26
	Sum = 0.99			
5	$s_1 = 0, s_2 = 0$ (white, white)	0.16	0.33	0.26
	$s_1 = 0, s_2 = 1$ (white, black)	0	0	NA (treated as 0)
	$s_1 = 1, s_2 = 0$ (black, white)	0.33	0.66	0.19
	Sum = 0.91			

	$s_1 = 1, s_2 = 1$ (black, black)	0.16	1	0
Sum = 0.45				
6	$s_1 = 0, s_2 = 0$ (white, white)	0.16	0.5	0.16
	$s_1 = 0, s_2 = 1$ (white, black)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 0$ (black, white)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.5	0.16
Sum = 0.66				
7	$s_1 = 0, s_2 = 0$ (white, white)	0.16	0.5	0.16
	$s_1 = 0, s_2 = 1$ (white, black)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 0$ (black, white)	0.16	0.5	0.16
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.5	0.16
Sum = 0.66				
8	$s_1 = 0, s_2 = 0$ (white, white)	0.16	1	0
	$s_1 = 0, s_2 = 1$ (white, black)	0.33	0.66	0.19
	$s_1 = 1, s_2 = 0$ (black, white)	0	0	NA (treated as 0)
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.33	0.26
Sum = 0.45				
Total sum/number of directions = <b>0.73</b>				

1398

**1399 Asymmetry**

1400

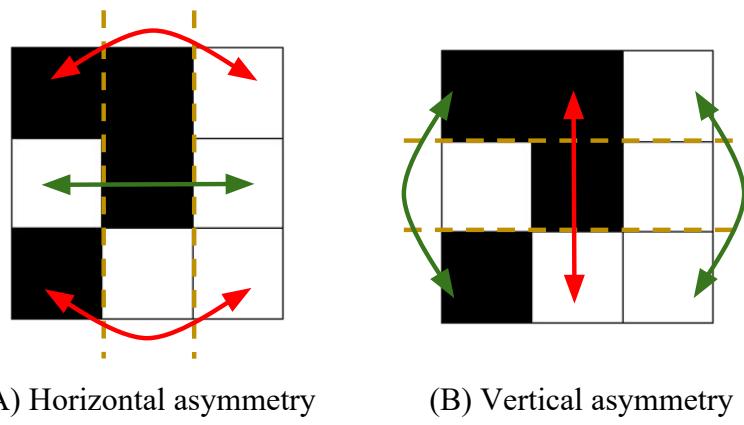
1401 Horizontal asymmetry measures asymmetry about the horizontal axis. Mathematically, for  
 1402 each cell  $(i, j)$  where  $i$  represents the cell row and  $j$  represents the cell column and assuming 0-  
 1403 indexing, horizontal symmetry compares the percentage of mismatches between  $(i, j)$  and  $(i, N-j-1)$  for  $j$  in  $[0, N/2]$ . On the other hand, vertical asymmetry measures asymmetry about the  
 1404 vertical axis and compares the percentage of mismatches between  $(i, j)$  and  $(N-i-1, j)$  for  $i$  in  
 1405  $[0, N/2]$ . Figure AII.6 illustrates the computation.

1406

**1407 Figure AII.6**

1408 *Horizontal (A) and vertical (B) asymmetry computation*

1410



(A) Horizontal asymmetry

(B) Vertical asymmetry

1411

1412

1413 In addition to the complexity measure mentioned in Section 2.2, we implemented two other  
1414 computational complexity measures: an 8-neighbourhood version of our intricacy measure and  
1415 a hierarchical quad-tree measure. The motivation for and computation of these measures are  
1416 presented here.

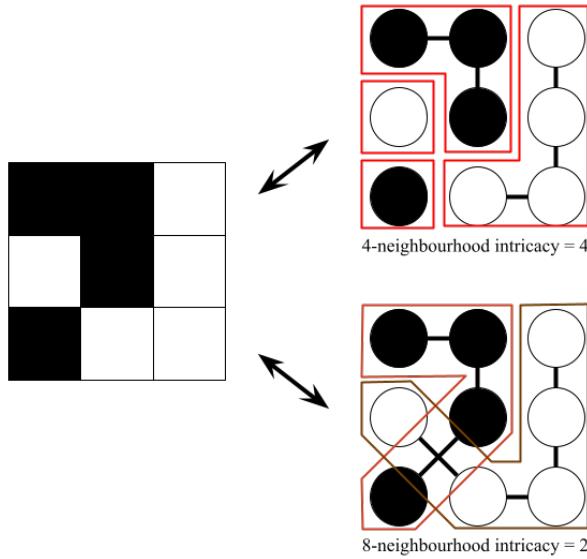
1417

- 1418 1. **8-neighbourhood intricacy measure:** In Figure 8, we presented some of the patterns  
1419 for which our complexity measure is most erroneous. Complexity is overestimated in  
1420 some patterns with high evaluated intricacy – potentially because diagonal relationships  
1421 do not contribute to connectedness. To tackle this, we implemented an 8-  
1422 neighbourhood version of intricacy wherein the graphs consider all 8 neighbours. An  
1423 edge is inserted between two neighbours if they are the same colour. This measure  
1424 would in turn result in a lower value of intricacy compared to the 4-neighbourhood  
1425 version. Figure AII.7 illustrates the computation of the 4-neighbourhood vs. the  
1426 8-neighbourhood intricacy. 8-neighbourhood intricacy was only weakly correlated to 4-  
1427 neighbourhood intricacy ( $r = 0.26$ ,  $p < 0.01$ ,  $CI = [0.23, 0.29]$ ) and other measures.  
1428

1429

### Figure AII.7

1430 4-neighbourhood and 8-neighbourhood intricacy computation for an example pattern.  
1431 The graphs on right are constructed from the pattern on left. Boxes indicate connected  
1432 components. Here, 4-neighbourhood intricacy = 4 and 8-neighbourhood intricacy = 2.  
1433



1434  
1435  
1436

1437 2. **Quadtree:** Since most of our measures (except multiscale entropy) are either local or  
1438 global, we programmed a hierarchical measure of complexity, namely quadtrees.  
1439 Quadtrees are hierarchical data structures and are commonly used for representing  
1440 images (Finkel and Bentley, 1974). A quadtree is based on the principle of recursive  
1441 decomposition where the pattern is repeatedly divided into four sub-patterns. The  
1442 condition for subdivision is based on the number of distinct states in the graph. If all  
1443 cells are of the same colour, the algorithm stops, if not, the graph is subdivided into  
1444 four sub-graphs and the same rule is recursively applied to each sub-graph. The number  
1445 of times the graph undergoes subdivisions is evaluated as a measure of complexity.  
1446 This measure however is not ideal for our pattern with grid size  $15 \times 15$  where the  
1447 width and height are not powers of 4 and as a result the four sub-patterns are of unequal  
1448 dimensions. Moreover, the measure was found to be highly correlated with LSC ( $r =$   
1449 0.88,  $p < 0.01$ , CI = [0.87, 0.88]).  
1450

1451 3. **Information measure H:** Snodgrass (1971) proposed two information theory measure  
1452 –  $H$  and  $H'$  to quantify complexity of  $9 \times 9$  binary pixel matrices.  $H$  was found to be  
1453 superior in predicting complexity across a series of experiments. Therefore, we tested  
1454 the applicability of  $H$  for our  $15 \times 15$  patterns. The measure is calculated as the entropy  
1455 of the distribution of submatrices at different scales (refer to Snodgrass, 1971 for more  
1456 details and examples). Based on the results from Snodgrass (1971), we only evaluate  $H$   
1457 at scales  $2 \times 2$  ( $H_{\_2}$ ) and the mean over all scales ( $H_{\_mean}$ ). Both  $H_{\_2}$  and  $H_{\_mean}$   
1458 were found to be highly correlated with LSC ( $r = 0.99$ ,  $p < 0.01$ , CI = [0.99, 0.99];  $r =$   
1459 0.88,  $p < 0.01$ , CI = [0.87, 0.88]), KC ( $r = 0.89$ ,  $p < 0.01$ , CI = [0.88, 0.89];  $r = 0.82$ ,  $p$   
1460  $< 0.01$ , CI = [0.81, 0.83]), entropy ( $r = 0.90$ ,  $p < 0.01$ , CI = [0.89, 0.90];  $r = 0.79$ ,  $p <$   
1461 0.01, CI = [0.78, 0.80]) and multiscale entropy ( $r = 0.92$ ,  $p < 0.01$ , CI = [0.91, 0.92];  $r =$   
1462 0.88,  $p < 0.01$ , CI = [0.87, 0.89]).  
1463

1464 Appendix III, section 2 reports the performance of these measures in predicting subjective  
 1465 complexity and Table AII.2 shows some example patterns along with the corresponding  
 1466 computed measures.

1467

**1468 Table AII.2**

**1469 Example patterns along with computed measures**

	Density	Entropy	Multiscale entropy	H_2, H_mean	LSC	KC	Asymmetry (horizontal, vertical, local)	Intricacy (4, 8 neighbour hood)	Quadtree	N	Rule type	IC	$\lambda$
	1.00	0.0	0.0	0.0, 0.0	0.0	25.17	0, 0, 0	1, 1	0	5	Otot	1	9
	0.004	0.04	0.05	0.18, 2.74	0.04	48.35	0, 0, 0	2, 2	4	-	-	1	-
	0.49	0.99	0.93	1.0, 0.93	0.0	33.43	0, 0, 0	255, 2	84	5	Otot	1	5
	0.10	0.48	0.55	1.74, 3.97	0.43	134.2	58.3, 0, 0	3	22	9	Otot	2	4
	0.74	0.81	0.80	3.21, 4.49	0.81	257.8	0, 0, 0	16	54	5	Otot	1	8
	0.45	0.99	0.89	3.96, 4.65	0.96	278.7	53.3, 46.6, 0.004	30	65	9	Otot	3	6

1470 *Note.* The generation parameters (N, rule type, IC, along with rule number (hence also  $\lambda$ ) and  
 1471 iteration number) specified are one such combination that can result in the corresponding  
 1472 shown pattern. However, there may be other combinations of these parameters that may give  
 1473 rise to the same pattern.

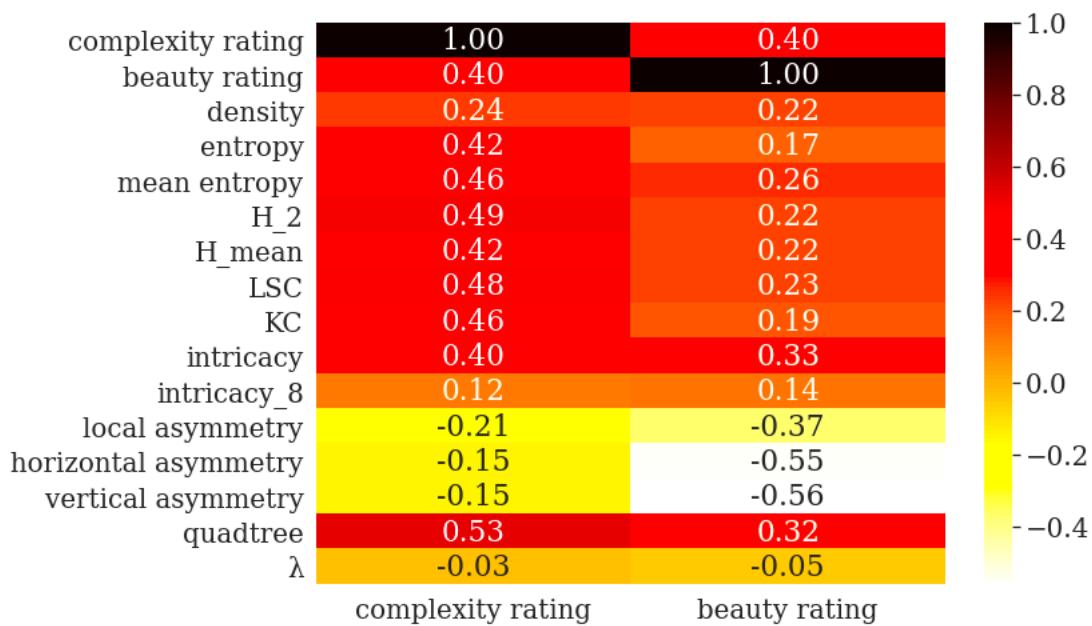
1474 **Appendix III – Supplementary Analysis**

1475  
1476 **1. Correlations between measures and ratings**

1477  
1478 From Figure AIII.1 we see that human complexity ratings are positively correlated with LSC  
1479 ( $r = 0.48$ ,  $p < 0.01$ ,  $CI = [0.37, 0.42]$ ), KC ( $r = 0.46$ ,  $p < 0.01$ ,  $CI = [0.46, 0.51]$ ), quadtree ( $r =$   
1480  $0.53$ ,  $p < 0.01$ ,  $CI = [0.5, 0.55]$ ), 4-nieghbourhood intricacy ( $r = 0.40$ ,  $p < 0.01$ ,  $CI = [0.38,$   
1481  $0.43]$ ) and H\_2 ( $r = 0.49$ ,  $p < 0.01$ ,  $CI = [0.47, 0.51]$ ) measures. The ratings were not highly  
1482 correlated with asymmetry or density. On the other hand, the beauty ratings were highly  
1483 negatively correlated with the three asymmetry measures ( $r = -0.37$ ,  $p < 0.01$ ,  $CI = [-0.39, -$   
1484  $0.34]$ ;  $r = -0.55$ ,  $p < 0.01$ ,  $CI = [-0.57, -0.52]$ ;  $r = -0.56$ ,  $p < 0.01$ ,  $CI = [-0.58, -0.54]$ ) and  
1485 entropy ( $r = -0.51$ ,  $p < 0.01$ ,  $CI = [-0.53, -0.49]$ ) and positively with complexity ratings ( $r =$   
1486  $0.40$ ,  $p < 0.01$ ,  $CI = [0.37, 0.42]$ ).

1487 **Figure AIII.1**

1488 *Correlation between ratings and computational measures*



1491  
1492 **2. Fitness of models of complexity ratings.**

1493  
1494 Here we report the full list of complexity models we experimented with along with variances  
1495 in the performance measures.

1496 **Table AIII.1**

1497  
1498 *(A) Models of complexity ratings*

Id	Model
1	$CR \sim 1 + (1   Participant)$
2	$CR \sim 1 + (1   Participant) + (1   Set)$

3 CR ~ density + (1 | Participant)  
 4 CR ~ entropy + (1 | Participant)  
 5 CR ~ multiscale\_entropy + (1 | Participant)  
 6 CR ~ H\_2 + (1 | Participant)  
 7 CR ~ H\_mean + (1 | Participant)  
 8 CR ~ LSC + (1 | Participant)  
 9 CR ~ KC + (1 | Participant)  
 10 CR ~ asymm + (1 | Participant)  
 11 CR ~ intricacy\_4 + (1 | Participant)  
 12 CR ~ intricacy\_8 + (1 | Participant)  
 13 CR ~ quadtree + (1 | Participant)  
 14 CR ~ LSC + density + (1 | Participant)  
 15 CR ~ LSC + entropy + (1 | Participant)  
 16 CR ~ LSC + multiscale\_entropy + (1 | Participant)  
 17 CR ~ LSC + H\_2 + (1 | Participant)  
 18 CR ~ LSC + asymm + (1 | Participant)  
 19 CR ~ LSC + intricacy\_4 + (1 | Participant)  
 20 CR ~ LSC + intricacy\_8 + (1 | Participant)  
 21 CR ~ LSC + quadtree + (1 | Participant)  
 22 CR ~ LSC + H\_2 + intricacy\_4 + (1 | Participant)  
 23 CR ~ LSC + intricacy\_4 + intricacy\_8 + (1 | Participant)  
 24 CR ~ LSC \* intricacy\_4 + (1 | Participant)  
 25 CR ~ LSC + intricacy\_4 + (LSC | Participant)  
 26 CR ~ LSC + intricacy\_4 + (intricacy\_4 | Participant)  
**27 CR ~ LSC + intricacy\_4 + ((LSC + intricacy\_4) | Participant)**  
 28 CR ~ LSC \* intricacy\_4 + ((LSC + intricacy\_4) | Participant)  
 29 CR ~ H\_2 + intricacy\_4 + (1 | Participant)  
 30 CR ~ H\_2 + intricacy\_4 + ((H\_2 + intricacy\_4) | Participant)  
 31 CR ~ H\_2 \* intricacy\_4 + ((H\_2 + intricacy\_4) | Participant)  
 32 CR ~ quadtree + intricacy\_4 + (1 | Participant)  
 33 CR ~ quadtree + intricacy\_4 + (quadtree | Participant)  
 34 CR ~ quadtree + intricacy\_4 + ((intricacy\_4 + intricacy\_4) | Participant)  
 35 CR ~ LSCsq + intricacy\_4sq + (1 | Participant)  
 36 CR ~ LSC + LSCsq + intricacy\_4 + intricacy\_4sq + (1 | Participant)  
 37 CR ~ neighbourhood\_size + tot\_outertot + IC + (1 | Participant)  
 38 CR ~ trial + (1 | Participant)  
 39 CR ~ LSC + intricacy\_4 + trial + ((LSC + intricacy\_4) | Participant)  
 40 CR ~ previous\_CR + (1 | Participant)  
 41 CR ~ LSC + intricacy\_4 + previous\_CR + ((LSC + intricacy\_4) | Participant)

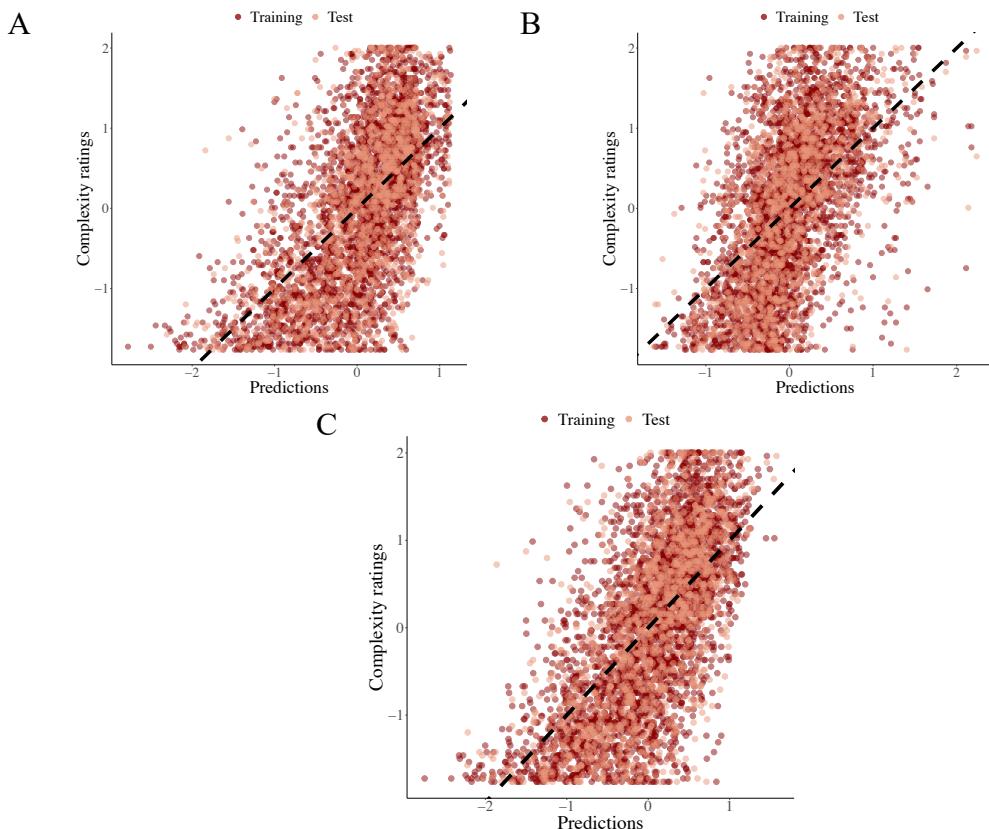
1500 Note. CR=complexity ratings, LSC=local spatial complexity, KC=Kolmogorov complexity; \*  
 1501 indicates p < 0.05. Bold indicates best model.  
 1502

1503 (B) Performance of models specified in Table AIII.1 based on AIC, BIC, R<sup>2</sup> and RMSE metrics.

Id	AIC	BIC	AIC/BIC Var	R <sup>2</sup>				RMSE			
				Train		Test		Train		Test	
				Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	8545.5	8563.6	2317.3	0.16	1.7e-06	0.13	4.7e-06	0.92	5e-05	0.93	2e-04
2	8546.4	8570.6	2355.6	0.16	1.8e-06	0.13	2.6e-06	0.92	5e-05	0.93	2e-04
3	8331.2	8355.4	1485.5	0.22	4.2e-05	0.19	0.00019	0.88	3.3e-05	0.9	0.00013
4	7829.9	7854.1	1754.5	0.34	8e-05	0.31	0.00032	0.81	3.2e-05	0.83	0.00013
5	7658.5	7682.7	610	0.38	4.6e-05	0.35	0.00016	0.79	1e-05	0.81	3.9e-05
6	7515	7539.2	1294.4	0.4	0.00017	0.38	0.00062	0.77	2.6e-05	0.79	8.7e-05
7	7818.9	7843.1	1065.4	0.34	7.1e-05	0.32	0.00029	0.81	2e-05	0.83	8e-05
8	7552	7576.2	1151.5	0.4	0.00015	0.37	0.00053	0.78	2.3e-05	0.79	7.5e-05
9	7703.8	7728	524.2	0.37	8.6e-05	0.34	3e-04	0.8	1.1e-05	0.81	3.3e-05
10	8462.2	8486.4	2176.9	0.19	1e-06	0.16	4e-06	0.9	4.6e-05	0.92	0.00019
11	7908.3	7932.5	2347.7	0.32	5.2e-06	0.3	2.1e-05	0.82	3.8e-05	0.84	0.00017
12	8493.1	8517.3	2803.9	0.18	8.5e-06	0.15	2.7e-05	0.91	6e-05	0.92	0.00024
13	7317.2	7341.4	298.4	0.44	5.4e-05	0.42	0.00015	0.75	5.3e-06	0.76	2.1e-05
14	7551.3	7581.5	1422.4	0.4	0.00017	0.37	0.00062	0.78	2.8e-05	0.79	9.2e-05
15	7536.6	7566.8	1387.7	0.4	0.00018	0.38	0.00066	0.77	2.8e-05	0.79	9.1e-05
16	7550.7	7581	872.7	0.4	0.00012	0.37	0.00045	0.78	1.8e-05	0.79	5.4e-05
17	7518.4	7548.7	1325.8	0.4	0.00017	0.38	0.00065	0.77	2.6e-05	0.79	9.1e-05
18	7520.2	7550.5	1136.7	0.4	0.00014	0.38	0.00051	0.77	2.2e-05	0.79	7.7e-05
19	7285.1	7315.4	229.4	0.45	6.8e-05	0.43	2e-04	0.74	4.6e-06	0.76	1.5e-05
20	7556.9	7587.1	1221.1	0.4	0.00014	0.37	0.00051	0.78	2.4e-05	0.79	8e-05
21	7314.8	7345	323.4	0.44	6.7e-05	0.42	2e-04	0.75	6.4e-06	0.76	2e-05
22	7265.6	7301.9	169.1	0.45	8.3e-05	0.43	0.00026	0.74	3.9e-06	0.76	1.3e-05
23	7291.4	7327.6	197	0.45	7.1e-05	0.43	0.00021	0.74	4.1e-06	0.76	1.3e-05
24	7276.8	7313.1	180.2	0.45	7.2e-05	0.43	0.00021	0.74	3.8e-06	0.76	1.1e-05
25	7219.9	7262.2	541.1	0.48	0.00012	0.45	0.00036	0.72	1.2e-05	0.75	3.2e-05
26	7166.3	7208.6	660.8	0.5	0.00011	0.46	0.00025	0.71	1.8e-05	0.74	3.4e-05
27	<b>7138.9</b>	<b>7199.4</b>	<b>822.7</b>	<b>0.52</b>	<b>0.00015</b>	<b>0.47</b>	<b>0.00035</b>	<b>0.7</b>	<b>2.1e-05</b>	<b>0.73</b>	<b>4.2e-05</b>
28	7133.6	7200.1	798.1	0.52	0.00015	0.47	0.00036	0.7	1.8e-05	0.73	4.2e-05
29	7261.3	7291.6	172.1	0.45	8.1e-05	0.43	0.00025	0.74	3.9e-06	0.76	1.2e-05
30	7107.5	7167.9	1001.5	0.52	0.00018	0.48	0.00043	0.69	2.9e-05	0.73	5.1e-05
31	7106.7	7173.2	1012.2	0.52	0.00018	0.48	0.00044	0.69	2.7e-05	0.73	5.2e-05
32	7311.8	7342.1	264.4	0.44	4.5e-05	0.42	0.00012	0.75	4e-06	0.76	1.8e-05
33	7194.4	7236.8	179.4	0.49	5.7e-05	0.45	0.00015	0.71	1.9e-06	0.74	3.4e-06
34	7191.2	7233.5	354	0.49	7.5e-05	0.45	0.00014	0.71	8.2e-06	0.74	2.2e-05
35	7329.3	7359.5	66.1	0.44	9.5e-05	0.42	0.00029	0.75	2.2e-06	0.77	5.5e-06
36	7256.7	7299	233.3	0.46	7.1e-05	0.43	2e-04	0.74	4.8e-06	0.75	1.3e-05

37	8457.2	8499.5	1845.1	0.19	9e-06	0.16	4.3e-05	0.9	3.9e-05	0.92	0.00017
38	8550.4	8574.6	2573.3	0.16	3.6e-06	0.13	1.2e-05	0.92	5.5e-05	0.93	0.00023
39	7141.6	7208.1	909.8	0.52	0.00014	0.47	0.00032	0.7	2e-05	0.73	4.8e-05
40	8525.2	8549.4	2549	0.17	1.6e-07	0.13	4.3e-07	0.91	5.2e-05	0.93	0.00022
41	7109.5	7176	491.6	0.52	0.00014	0.47	0.00027	0.69	1.7e-05	0.73	2.7e-05

1504

1505 **Figure AIII.2**1506 *Model fit plots (A) only-LSC model (without random slopes), (B) only intricacy model (without  
1507 random slopes) and (C) the model including both LSC and intricacy (without random slopes)*1508 3. Fitness of models of beauty ratings.

1509

1510 Here we report beauty models in the main text along with variances along with the full list of  
 1511 beauty models we experimented with. *disorder* is a weighted combination of asymmetry and  
 1512 entropy, *disorder\_2* is a weighted combination of asymmetry and multiscale entropy,  
 1513 *disorder\_3* is a weighted combination of local asymmetry and entropy and *disorder\_4* is a  
 1514 weighted combination of local asymmetry and multiscale entropy. These are compared to test  
 1515 the relative role of global vs local features in predicting beauty ratings. It is seen that *disorder*  
 1516 performs the best, *i.e.*, global features of mean asymmetry (mean of horizontal and vertical  
 1517 asymmetry) and entropy (at single scale) are more suitable than local asymmetry or multiscale  
 1518 entropy.

1519

1520 **Table AIII.2**

## 1521 (A) Models of beauty ratings

Id	Model
1	$BR \sim CR + (1   Participant)$
2	$BR \sim CR + asymm + entropy + (1   Participant)$
3	$BR \sim CR + asymm + multiscale\_entropy + (1   Participant)$
4	$BR \sim CR + disorder + (1   Participant)$
5	$BR \sim CR + disorder + (disorder   Participant)$
6	$BR \sim CR + disorder + (CR   Participant)$
7	$BR \sim CR + disorder + ((CR + disorder)   Participant)$
8	$BR \sim CR * disorder + (1   Participant)$
9	$BR \sim CR * disorder + (disorder   Participant)$
10	$BR \sim CR * disorder + (CR   Participant)$
11	<b><math>BR \sim CR * disorder + ((CR + disorder)   Participant)</math></b>
12	$BR \sim CR * disorder_2 + ((CR + disorder_2)   Participant)$
13	$BR \sim CR * disorder_3 + ((CR + disorder_3)   Participant)$
14	$BR \sim CR * disorder_4 + ((CR + disorder_4)   Participant)$
15	$BR \sim CR_sq + disorder + (1   Participant)$
16	$BR \sim CR + CR_sq + disorder + (1   Participant)$
17	$BR \sim LSC + intricacy_4 + asymm + entropy + (1   Participant)$
18	$BR \sim LSC + intricacy_4 + disorder + (1   Participant)$
19	$BR \sim obj\_comp + asymm + entropy + (1   Participant)$
20	$BR \sim obj\_comp + disorder + (1   Participant)$
21	$BR \sim obj\_comp + disorder + (obj\_comp   Participant)$
22	$BR \sim obj\_comp + disorder + (disorder   Participant)$
23	$BR \sim obj\_comp + disorder + ((obj\_comp + disorder)   Participant)$
24	$BR \sim obj\_comp * disorder + ((obj\_comp + disorder)   Participant)$
25	$BR \sim trial + (1   Participant)$
26	$BR \sim CR * disorder + trial + ((obj\_comp + disorder)   Participant)$
27	$BR \sim previous_BR + (1   Participant)$
28	$BR \sim CR * disorder + previous_BR + ((obj\_comp + disorder)   Participant)$
29	$BR \sim CR * trial + (1   Participant)$

1522 Note. BR=beauty ratings, CR=complexity ratings, LSC=local spatial complexity; \* indicates p  
 1523 < 0.05. Bold indicates best model.

1524

1525 (B) Performance of models specified in Table AIII.2 based on AIC, BIC, R<sup>2</sup> and RMSE metrics.

1526

Id	AIC	BIC	AIC, BIC	R <sup>2</sup>				RMSE			
				Train		Test		Train		Test	
				Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	8046.	8070.	3052.	0.29	0.0002	0.25	0.00073	0.85	7.1e-05	0.87	0.00019
	4	6	4		5						
2	6628	6664.	5020	0.56	0.0001	0.53	0.00065	0.67	6.3e-05	0.69	0.00025
		3			5						
3	6627.	6663.	4183.	0.56	0.0001	0.53	0.00055	0.67	5.2e-05	0.69	0.00021
	3	6	7		3						
4	6619.	6649.	5035.	0.56	0.0001	0.53	0.00065	0.67	6.3e-05	0.69	0.00025
	6	8	6		5						
5	6214.	6256.	7991.	0.64	0.0001	0.59	0.00071	0.6	9.3e-05	0.64	0.00032
	1	5	1		9						
6	6340.	6382.	1946.	0.62	5.9e-05	0.58	0.00036	0.62	1.9e-05	0.65	0.00013
	1	4	9								

7	5942.	6003	1529	0.69	0.0002	0.63	1	0.56	0.0001	0.61	6e-04
	6		4.1		7			8			
8	6524.	6560.	7142.	0.57	0.0001	0.54	8e-04	0.66	8.5e-05	0.68	0.00033
	6	9	8		9			8			
9	6080.	6128.	1000	0.66	2e-04	0.62	0.00084	0.58	0.0001	0.62	0.00042
	3	6	8.7					1			
10	6268	6316.	3800.	0.63	8.9e-05	0.59	0.00056	0.61	3.5e-05	0.64	0.00024
	4	9									
<b>11</b>	<b>5846.</b>	<b>5912.</b>	<b>1585</b>	<b>0.7</b>	<b>0.0002</b>	<b>0.65</b>	<b>0.0011</b>	<b>0.55</b>	<b>0.0001</b>	<b>0.6</b>	<b>0.00065</b>
	4	9	9.6		6			7			
12	5880.	5946.	1611	0.7	0.0002	0.64	0.0011	0.55	0.0001	0.6	0.00067
	3	8	7.6		7			8			
13	7141	7207.	95.1	0.53	1.1e-05	0.46	0.00011	0.68	5.1e-07	0.73	2e-05
		5									
14	7149.	7216.	69.2	0.53	1.3e-05	0.46	0.00012	0.68	3.6e-07	0.73	2e-05
	8	3									
15	7099.	7130.	1461.	0.48	8.2e-05	0.45	0.00039	0.72	2.2e-05	0.74	1e-04
	9	2	2								
16	6628	6664.	5063.	0.56	0.0001	0.53	0.00065	0.67	6.4e-05	0.69	0.00025
	2	2			5						
17	6793.	6836.	509.5	0.53	1.7e-05	0.51	8.6e-05	0.68	6.9e-06	0.7	3.3e-05
	8	1									
18	6853.	6889.	923	0.52	4.4e-05	0.5	2e-04	0.69	1.3e-05	0.71	5.6e-05
	6	9									
19	6817.	6853.	718	0.53	3.5e-05	0.5	0.00017	0.69	1e-05	0.71	4.7e-05
	2	5									
20	6844.	6874.	922.6	0.52	4.4e-05	0.5	2e-04	0.69	1.3e-05	0.71	5.7e-05
	1	3									
21	6780.	6823	283.6	0.55	9.1e-06	0.51	0.00015	0.67	1.2e-06	0.7	4.5e-05
		7									
22	6410.	6452.	3254.	0.62	0.0001	0.57	0.00033	0.62	4.3e-05	0.66	0.00011
	2	5	8		1						
23	6356.	6417	3497.	0.64	0.0001	0.58	0.00035	0.6	5.1e-05	0.65	0.00013
	5		4		1						
24	6205.	6271.	2362.	0.66	8.1e-05	0.6	0.00025	0.58	3.6e-05	0.63	8.9e-05
	1	6	4								
25	8544.	8568.	751.7	0.17	0.0001	0.12	0.00033	0.91	2.4e-05	0.94	5.8e-05
	7	9			2						
26	6030.	6102.	1002	0.68	0.0001	0.62	0.00083	0.57	1e-04	0.62	0.00044
	1	6	6.3		8						
27	8526.	8550.	1500.	0.17	0.0001	0.12	0.00054	0.91	3.6e-05	0.94	0.00014
	1	3	1		6						
28	6000.	6073.	9664.	0.68	0.0001	0.62	0.00082	0.56	1e-04	0.62	0.00044
	8	4	4		7						
29	8060.	8097.	3075.	0.29	0.0002	0.25	0.00073	0.85	7.1e-05	0.87	0.00019
	9	2	9		5						

1527

1528

4. Mixed Effects Regressions for predicting Complexity Ratings with 8-neighbourhood Intricacy and Quadtree (refer to AII for a description of these measures)

1529

1530

1531 Table AIII.3 summarizes models of complexity ratings involving 8-neighbourhood intricacy  
 1532 and quadtree. Comparing with Table 1, we see that H\_2 performs better than H\_mean and  
 1533 achieves performance comparable to LSC. Quadtree achieves good performance as a predictor  
 1534 alone, but there are negligible gains, with a hint of overfitting when we introduce intricacy into  
 1535 the expression. Further, the 8-neighbourhood intricacy is unable to explain much variance in  
 1536 the ratings compared to the 4-neighbourhood version implying that the 4-neighbourhood  
 1537 intricacy is a more superior predictor of complexity ratings.

1538

1539 Table AIII.3: Summary of models of complexity ratings

Id	Model	Significance	AIC	BIC	$R^2$		RMSE	
					Train	Test	Train	Test
1	CR ~ H_2 + H_2* 1 Participant		7515.0	7539.2	0.40	0.38	0.77	0.79
2	CR ~ H_mean + H_mean* 1 Participant		7818.9	7843.1	0.34	0.32	0.81	0.83
3	CR ~ quadtree + quadtree* 1 Participant		7317.2	7341.4	0.44	0.42	0.75	0.76
4	CR ~ intricacy_8 + intricacy_8* + 1 Participant		8493.1	8517.3	0.18	0.15	0.91	0.92
5	CR ~ H_2 + H_2* intricacy_4 + intricacy_4* (H_2 + intricacy_4) Participant		7107.5	7167.9	0.52	0.48	0.69	0.73
6	CR ~ quadtree + quadtree* intricacy_4 + intricacy_4* (quadtree + intricacy_4) Participant		7162.2	7222.7	0.51	0.46	0.70	0.74

1540

## 1541 5. Test for trends, autocorrelation, and consistency of repeated measures in the ratings

1542

1543 To test for trends and autocorrelation (at lag 1) in the data, trial number and previous rating  
 1544 were respectively added as a predictor of complexity and beauty ratings. Table AIII.4 (A) and  
 1545 (B) report the performance of these models.

1546

## 1547 Table AIII.4

1548 (A) Summary of models of complexity ratings. CR = complexity ratings, LSC = local spatial  
 1549 complexity, prevCR = previous complexity rating (from previous trial)

1550

Id	Model	Significance	AIC	BIC	$R^2$		RMSE	
					Train	Test	Train	Test

1	$CR \sim trial + 1 Participant$	8550.4	8574.6	0.16	0.13	0.92	0.93
2	$CR \sim LSC + LSC^* intricacy_4 + trial intricacy_4^*$ + (LSC + intricacy_4) Participant	7141.6	7208.1	0.52	0.47	0.70	0.73
3	$CR \sim prevCR + prevCR^* 1 Participant$	8525.2	8549.4	0.17	0.13	0.91	0.93
4	$CR \sim LSC + LSC^* intricacy_4 + intricacy_4^*$ prevCR + (LSC + prevCR^* intricacy_4) Participant	7109.5	7176.0	0.52	0.47	0.69	0.73

1551

1552 (B) Summary of models of beauty ratings. BR = beauty ratings, OC = objective complexity

1553

Id	Model	Significance	AIC	BIC	$R^2$		RMSE	
					Train	Test	Train	Test
1	$BR \sim trial + 1 Participant$		8544.7	8568.9	0.17	0.12	0.91	0.94
2	$BR \sim CR disorder + CR^* disorder^*$ CR:disorder + CR:disorder trial + (CR + * disorder) Participant		5852.9	5925.4	0.70	0.65	0.55	0.60
3	$BR \sim prevBR + prevBR^* 1 Participant$		8526.1	8550.3	0.17	0.12	0.91	0.94
4	$BR \sim CR disorder + CR^* disorder^*$ CR:disorder + CR:disorder prevBR + (CR + * disorder) Participant prevBR^*		5835.0	5907.5	0.70	0.65	0.55	0.59

1554

1555 For predicting complexity ratings, we find that trial number is not a significant predictor,  
1556 indicating there are no significant trends in our data – as would be expected in case of boredom,  
1557 or over-familiarity. We see that previous complexity rating is a significant predictor, however,  
1558 it does not enhance performance in conjunction with LSC and intricacy. Similarly for  
1559 predicting beauty ratings, trial is not significant and previous beauty rating, though significant,  
1560 does not enhance performance beyond our best model reported in Section 3.2.

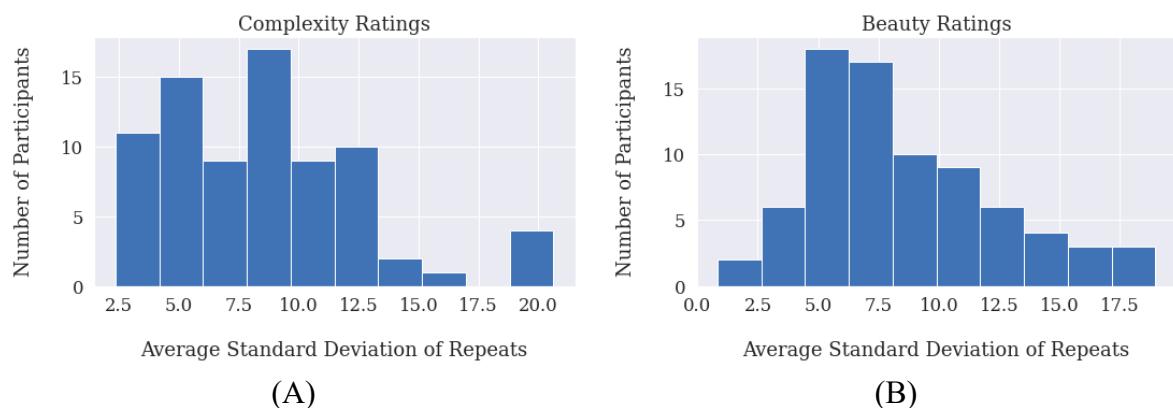
1561

1562 Further, to test for consistency of repeated responses, we plot participant mean standard  
1563 deviation in repeated measures (Figure AIII.3 (a) for complexity ratings and (b) for beauty  
1564 ratings). The range of standard deviation values is from 0 (when two values are identical) to  
1565 70.7 (when the two values are farthest apart on the scale, *i.e.*, one value is 0 and the other is  
1566 100). Since most of the average standard deviations are within ~20% of the maximum standard  
1567 deviation, we conclude that participants are largely consistent in their ratings.  
1568

1569 **Figure AIII.3**

1570 *Histogram of participant mean standard deviation in repeated measures for (A) complexity*  
1571 *ratings and (B) beauty ratings.*

1572



(A)

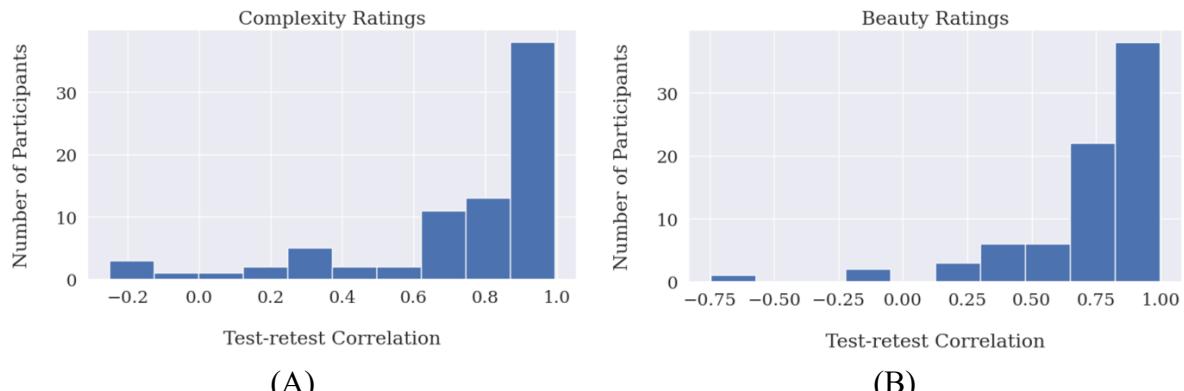
(B)

1573

1574 To test this further, we plot the distribution of test-retest correlations, *i.e.* correlations between  
1575 repeats among participants for complexity and beauty ratings (Figure AIII.4). We see that most  
1576 correlations are above 0.5 (with a mean 0.728 for complexity ratings and 0.725 for beauty  
1577 ratings) which means that the majority of participants were largely consistent in their ratings.  
1578

1579 **Figure AIII.4**

1580 *Histogram of participant test-retest correlations for (A) complexity ratings and (B) beauty*  
1581 *ratings.*



(A)

(B)

1582

- 1583 6. Average complexity ratings vs average beauty ratings (colour coded by symmetry and  
1584 entropy)

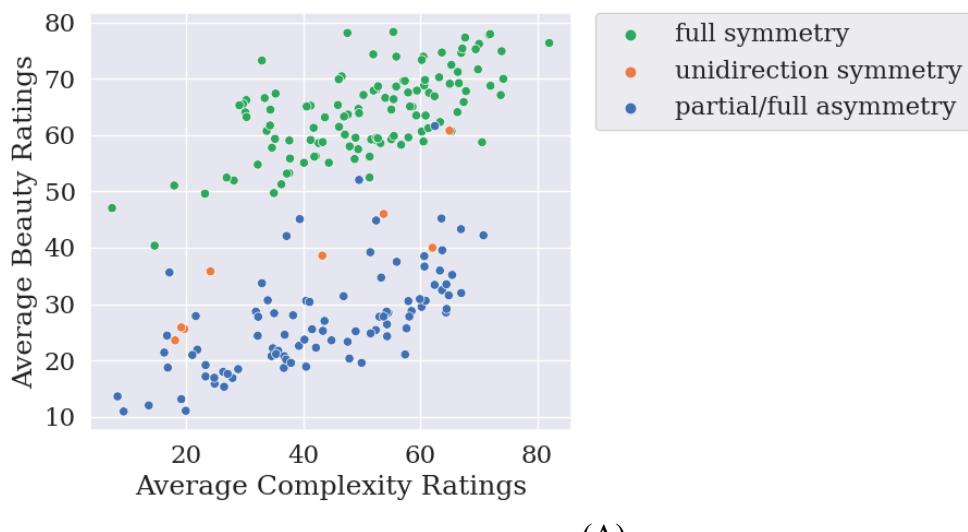
1585

1586 A plot of average beauty ratings per pattern across all participants versus average complexity  
 1587 ratings per pattern across all participants neatly underlines the role of asymmetry and entropy  
 1588 in beauty assessment (Figure AIII.5a,b). In Figure AIII.5a, the linear relationship between  
 1589 beauty and complexity is evident but we see two modes in the distribution (a Gaussian Mixture  
 1590 Model with 2 components was able to fit the average ratings well, Appendix AIII.7). The  
 1591 degree of symmetry is successfully able to explain this bimodality with fully symmetric  
 1592 patterns being rated higher on average than unidirectional symmetric (semi-symmetric)  
 1593 patterns, which are themselves rated higher on average than fully asymmetric (non symmetric)  
 1594 patterns. Further, Figure AIII.5b indicates that patterns with high entropy were largely rated as  
 1595 low beauty whereas patterns with low entropy were largely rated as high beauty unless the  
 1596 pattern was rated very low complexity.  
 1597

### 1598 **Figure AIII.5**

1599 *Average complexity ratings vs average beauty ratings per pattern across all participants*  
 1600 *labelled according to (A) degree of symmetry in the pattern (as defined in Table 1), (B) level*  
 1601 *of entropy of the pattern (tertile split of entropy)*

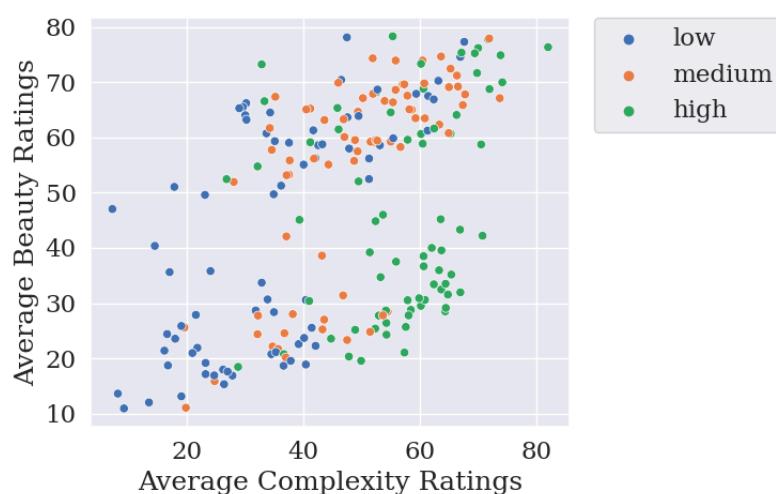
1602



1603 (A)

1604

1605



1606 (B)

1607

1608

1609     7. Gaussian Mixture Model to fit average complexity ratings versus average beauty ratings

1610

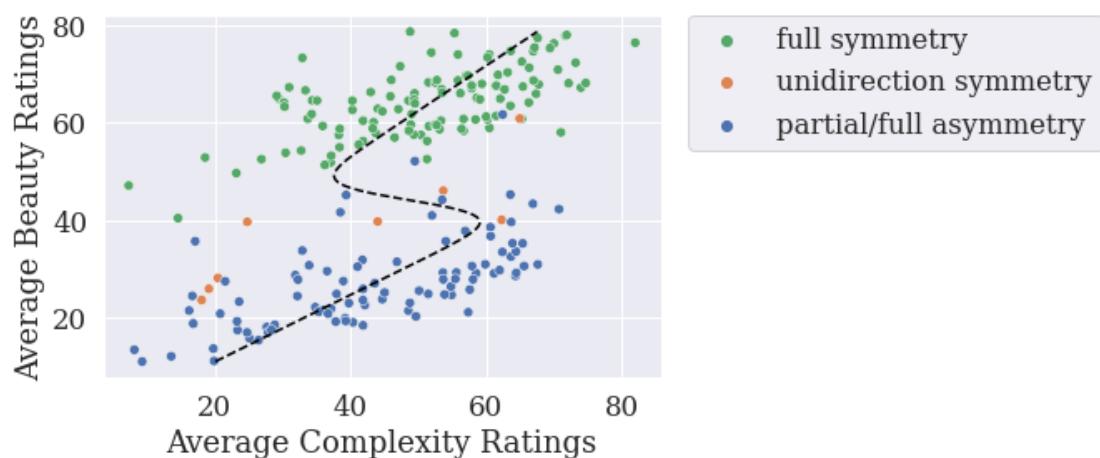
1611 As illustrated above, the plot of average complexity ratings versus average beauty ratings  
1612 reflects a bimodal distribution. We therefore used the *gmr* package in Python to perform  
1613 Gaussian Mixture Regression (GMR) with 2 modes on the average ratings. In GMR, first the  
1614 joint distribution  $p(x, y)$  is learnt, where in this case  $x$  refers to average beauty ratings and  $y$   
1615 refers to average complexity ratings. Then, the conditional  $p(y | x)$  is computed to make  
1616 predictions. GMR with 2 components is able to fit the data well (Figure AIII.6) with an RMSE  
1617 of 12.0 on a held-out test set containing 40 data points.

1618

1619 **Figure AIII.6**

1620 *GMM fit on average ratings. The dashed line represents the model fit by plotting the prediction*  
1621 *of average complexity ratings against a continuous range of average beauty ratings.*

1622



1623

1624

1625     8. Examination of Individual Differences

1626

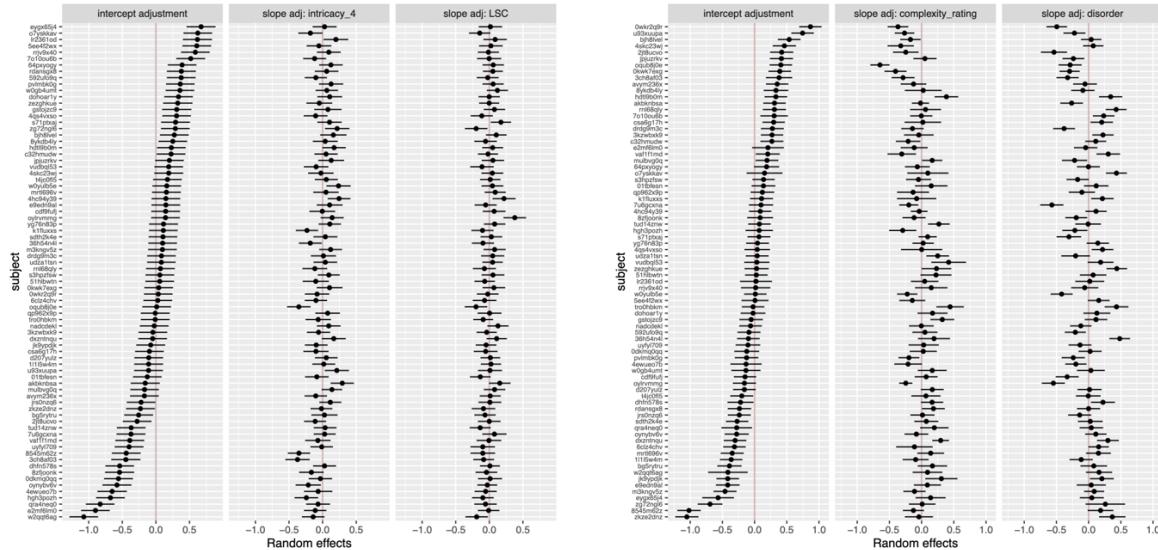
1627 Figure AIII.7 displays a plot of random effects for the best performing model of complexity  
1628 ratings (A) and beauty ratings (B). The plots show that random intercepts for participant have  
1629 a higher variance than random slopes for in both complexity and beauty models. The variation  
1630 in random intercepts of participants show that while some people rate complexity to be high  
1631 on average, some others rate them to be low, and similarly for beauty. Moreover, the random  
1632 slopes indicate that people perceive the impact of LSC, intricacy on complexity and subjective  
1633 complexity, disorder on beauty to different degrees. More thorough analysis of individual  
1634 differences is intended to be a part of our future work.

1635

1636 **Figure AIII.7**

1637 *Plot of random effects from our best performing (A) complexity model and (B) beauty model*

1638



(A) Random intercept of participant and random slopes of intricacy and LSC

(B) Random intercept of participant and random slopes of subjective complexity and disorder

1639

## 1640 References

1641

1642 Finkel, R. A., & Bentley, J. L. (1974). Quad trees a data structure for retrieval on composite  
1643 keys. *Acta informatica*, 4(1), 1-9.

1644

1645 Packard, N. H., & Wolfram, S. (1985). Two-dimensional cellular automata. *Journal of  
1646 Statistical physics*, 38(5), 901-946.

1647

1648 Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of modern  
1649 physics*, 55(3), 601.

1650

1651 Snodgrass, J. G. (1971). Objective and subjective complexity measures for a new population  
1652 of patterns. *Perception & Psychophysics*, 10(4), 217-224.