

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

3
4 Surabhi S Nath^{1,2}, Franziska Brändle¹, Eric Schulz¹, Peter Dayan^{*,1,3}, Aenne Brielmann^{*1}

5
6 ¹Max Planck Institute for Biological Cybernetics, Tübingen, Germany

7 ²Max Planck School of Cognition, Stephanstrasse 1a, Leipzig, Germany

8 ³University 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
25 Foundation)—461354985.

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

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

28 Correspondence concerning this article should be addressed to Surabhi S Nath, AGPD, Max
29 Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany.
30 E-mail: surabhi.nath@tuebingen.mpg.de

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 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 depend on the number of distinct elements in the pattern along with the
47 elements’ local spatial distribution. Therefore, global and local image features are integrated
48 to determine complexity judgements. Furthermore, we found a positive linear influence of
49 complexity ratings on beauty, with a negative linear influence of disorder (asymmetry and
50 entropy of means), and a negative interaction between the two ($R^2_{\text{test}} = 0.64$). This implies that
51 there is beauty in complexity as long as there is sufficient order. Lastly, a moderated mediation
52 analysis showed that complexity mediates the influence of objective complexity on beauty,
53 implying that complexity supplies useful information over and above objective complexity.

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

55 **1. Background and Introduction**

56

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

68

69 **Relationship between Subjective and Objective Complexity:** The complexity literature
70 includes several attempts to arrive at an understanding of the basis of human ratings of
71 complexity. Although complete consensus has not been reached, two prominent streams of
72 work emerge – one based on qualitative measures which focusses on identifying separable
73 object features that contribute to experienced complexity, and the other based on quantitative
74 measures, which focuses on the objects' statistical properties.

75

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

86

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

98 Parallely, the works of Chipman (1977), Chipman and Mendelson (1979) proposed that
99 subjective complexity was determined by two components – a quantity-based component
100 derived from the number of vertices in the stimuli and a structure-based component based on
101 the structural aspects like symmetry, repetitions etc. They also indicated that these two
102 components might involve different cognitive mechanisms. Additional evidence for this dual
103 processing theory came from Ichikawa (1985) who attributed different levels of processing to
104 these two components – a lower level/primary processing for quantity-based features and a
105 higher-level processing for the discovery of structure.

106 By contrast, the second stream of work has focused on more quantitative correlates of
107 subjective complexity judgements based on statistical properties of the stimuli. Fred Attneave
108 (1919-1991) was the first to apply information theory to quantify stimulus properties in the
109 context of aesthetics . Shannon's information theory (Shannon, 1948) conceptualised "entropy"
110 as a measure of the quantity of information potentially contained in a signal. This was later
111 adopted as a measure of order and complexity in aesthetics by several authors (Arnheim, 1956,
112 1966; Bense, 1960, 1969; Moles, 1958; Schmidhuber, 2009). Low entropy implies low
113 uncertainty, high predictability, high order, and less complexity, and vice versa. Shannon
114 entropy also found its way into the creative fields of art (used to measure pattern complexity in
115 the Kolam artform Tran et al., 2021) and music (used to measure complexity in sequences of
116 tones (Delplanque et al., 2019)). The compressibility of an image could be used as a measure
117 of its complexity with less compressibility implying more complexity and vice-versa (Birkin,
118 2010; Donderi, 2006). Kolmogorov complexity is a similar quantification of data compression
119 determined by the length of the shortest computer program that produces a desired output in
120 any standard universal computer programming language, and is one of the most direct
121 applications of algorithmic information theory to stimulus description (eg. Chikhman et al.,
122 2012; Singh and Shukla, 2017). Snodgrass (1971) demonstrated the potential of information
123 measures in predicting subjective complexity on black-and-white pixel patterns. Javaheri Javid
124 recently introduced a measure of spatial complexity based on the probabilistic spatial
125 distribution of pairs of pixels and also examined the applicability of Kolmogorov complexity
126 to evaluate the complexity of pixel patterns created using aesthetic automata (Javaheri Javid,
127 2021). Corchs et al. (2016) used computational measures quantifying spatial, frequency and
128 colour properties to predict complexity of real-world stimuli. Other modern studies have used
129 computational image properties such as histogram of oriented gradients (HOG), Fourier slope
130 and fractal dimension to quantify complexity (refer to Van Geert and Wagemans, 2020 for a
131 detailed review).

132
133 **Relationship between Beauty and Complexity:** In contrast to work focussing directly on
134 complexity, beauty, and aesthetic evaluations more generally (aesthetic preference/liking,
135 pleasantness, pleasure), have been the subject of a rather large body of studies, including many
136 attempts to model the processes leading to aesthetic evaluations. Nevertheless, these two
137 streams of work are closely coupled, as complexity has been considered an important
138 contributor to aesthetic evaluations since the early days of empirical aesthetics (although most
139 modern models of aesthetic value consider the role of complexity only implicitly (Brielmann
140 and Dayan, 2022; Ligaya et al., 2020)).

141
142 Fechner's principle of "unitary connection" suggested that pleasant stimuli express a balance
143 of complexity and order (Cupchik, 1986), and work by Birkhoff mathematically formulated an
144 aesthetic measure (M) that varied positively with order (O) and negatively with complexity I ,
145 or $M = O / C$ for polygonal figures, vases, poetry and music (Birkhoff, 1933). However, Davis
146 (1936) criticized the measure M as being inappropriate for empirical test. Several studies aimed
147 at testing the applicability of M have yielded a high variance in correlations between actual
148 rankings and those given by the formula (Eysenck, 1941; Harsh and Beebe-Center, 1939). Later
149 experiments by Eysenck suggested an empirical formula for M which yielded much higher
150 correlations with subject rankings. He suggested an approximation where aesthetic preference
151 varied positively with *both* order and complexity, altering the equation to $M = O \times C$ (Eysenck,
152 1942, 1968).

153
154 Berlyne also formulated a relationship between complexity and aesthetic preference, via a more
155 general inverted-U relationship between hedonic value and arousal potential (the
156 "psychological strength" or the extent to which a stimulus is capable of raising arousal
157 (Berlyne, 1967)) of a stimulus (Berlyne, 1960). Berlyne proposed three classes of variables that
158 determined arousal potential, namely, psychophysical variables, ecological variables and
159 collative variables (Berlyne, 1971). Collative variables include subjective novelty, complexity
160 and surprise. This theory was further applied to the context of art, suggesting that the collative
161 variable of subjective complexity is one of the most significant determinants of aesthetic
162 preference. The implication was that there is a general preference for stimuli of intermediate
163 complexity, as described by the inverted-U shape.

164
165 Several subsequent works have attempted to reproduce this inverted-U relationship between
166 stimulus complexity and aesthetic preference. However, there is little evidence in support of
167 this theory, but rather contradictions (Nadal et al., 2010). While several studies supported the
168 importance of complexity in shaping aesthetic preferences (e.g., Jacobsen and Höfel (2002);
169 Jacobsen et al. (2006); Tinio and Leder (2009)), the type and directionality of the relationship
170 has not been conclusively settled to date (Nadal et al., 2010). Some studies concurred with the
171 original inverted U-relationship (Chmiel and Schubert, 2017; Eisenman, 1967; Farley and
172 Weinstock, 1980; Gordon and Gridley, 2013; Hekkert and Van Wieringen, 1990; Lakhal et al.,
173 2020; Madison and Schiölde, 2017; Marin et al., 2016; Munsinger and Kessen, 1964; Nasar,
174 2002; Nicki, 1972; Vitz, 1966), while others suggested a linear relationship (Eysenck, 1941;
175 Day, 1967; Heath et al., 2000; Javaheri Javid, 2019; Nicki and Moss, 1975; Nicki and Gale,
176 1977; Osborne and Farley, 1970; Stamps III, 2002; Taylor and Eisenman, 1964). Surprisingly,
177 a few other studies reported either a non-inverted U (Adkins and Norman, 2016; Norman et al.,
178 2010), or no relationship at all (Messinger, 1998).

179
180 The reasons for these discrepancies include theoretical and empirical challenges similar to
181 those for complexity studies, along with other, unique factors. First, there are contrasting
182 methods of defining, measuring and manipulating objective complexity (Nadal et al., 2010).

183

184 Second, most of the above-mentioned works use handcrafted stimuli. Human-generated stimuli
185 are more likely to suffer from implicit biases of the designer. Moreover, such stimuli cannot be
186 reproduced without access to the original stimuli set, nor can be manipulated easily, making it
187 difficult to falsify findings or identify factors causing the failure of generalisation. One
188 remarkable recent deviation from this pattern is the OCTA toolbox (Van Geert et al., 2022)
189 which allows the user to generate stimuli comprised of multiple elements based on a series of
190 settings that aim to manipulate order and/or complexity. However, this toolbox does not
191 compute a complexity (or order) measure based on the created stimuli's appearance but rather
192 returns the settings used to create the stimulus.

193

194 There is therefore the need for systematically-generated experimental stimuli along with well-
195 defined programmatic measures. In this work, we relinquish generality and naturalness in
196 favour of reproducibility and rigour, and develop an algorithmically defined stimulus generator
197 which is capable of systematically producing simple yet diverse abstract patterns that span a
198 range of subjective assessments of complexity and beauty. We develop various programmatic
199 pattern quantification measures defined for our pattern class including measures of density,
200 entropy, asymmetry, information in the pattern, along with a novel “intricacy” measure
201 quantifying the number of separable visual components in the pattern.

202

203 Based on this, our work has two components:

204

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

207

208 Secondly, we inspect the relationship between ratings of the beauty¹ and both the subjective
209 and objective complexity of our patterns.

210

211 1. Methods

212

213 2.1 Cellular Automata for Pattern Generation

214

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

¹ 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).

224

225 Definition: Cellular Automata

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

- 227 1. L is a $P \times Q$ grid with cells (i, j) , $1 \leq i \leq P$, $1 \leq j \leq Q$
- 228 2. $S = \{1, 2, \dots, k\}$ are the potential states of each cell $(i, j) \in L$. Therefore, each cell (i, j)
229 has a state at time t denoted by $s_{(i,j)}^t \in S$
- 230 3. $N_{(i,j)}$ is the neighbourhood of cell (i, j) which can either be von Neumann/5-cell ($N = 5$)
231 or Moore/9-cell ($N = 9$) neighbourhoods, where N is the neighbourhood size. The
232 neighbourhood is specified by a set of vectors $\{e_a\}$, $a = 1, 2, \dots, N$ such that $N_{(i,j)}$ is given
233 by $\{(i, j) + e_1, (i, j) + e_2, \dots, (i, j) + e_N\}$ and can be denoted as $\{(i_1, j_1), (i_2, j_2), \dots, (i_N, j_N)\}$.
234 A cell is always considered its own neighbour hence one of $\{e_a\}$ is the zero vector $(0,0)$.
235 Periodic boundary conditions are applied at the edges of the lattice so that complete
236 neighbourhoods exist for every cell in L .
- 237 4. f is the state-transition function which computes the state of cell (i, j) at the $t+1 = s_{(i,j)}^{t+1}$
238 as a function of the states of the cells in its neighbourhood. Hence, $s_{(i,j)}^{t+1} =$
239 $f(s_{(i_1, j_1)}^t, s_{(i_2, j_2)}^t, \dots, s_{(i_N, j_N)}^t)$, where $(i_1, j_1), (i_2, j_2), \dots, (i_N, j_N) \in N_{(i,j)}$. A quiescent state s_q
240 satisfies $f(s_q, I., s_q) = s_q$.

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

242

243 We use binary 2D CA, $S = \{0, 1\}$ (corresponding to the colours white and black). Taking
244 inspiration from pixel-art sketches, we set our grid size to 15×15 ($P = Q = 15$). We consider
245 simple rules and initial configurations. We use two classes of rules, conventionally called
246 “totalistic” (Tot) or “outer-totalistic” (Otot) (Refer to Appendix I for specific details of the
247 generation algorithm). For our initial configurations (ICs), based on work by Javaheri Javid
248 (2019), we use either a single central cell (IC_1), a small disordered central region (IC_2, also
249 known as the glider shape in Game of Life CA (Gardner, 1970)) or a fully random starting
250 grid configuration (IC_3) (Figure 1A). We limit the number of iterations to $t = 40$, and add
251 every 5th pattern to the stimuli set. Therefore, each distinct set of algorithm parameter values
252 produces 8 patterns.

253

254 We use 51 rules in total with differing combinations of neighbourhood size (5-cell/9-cell
255 neighbourhood), rule code, rule type (totalistic/outer-totalistic) and initial configuration
256 (IC_1, IC_2 or IC_3) resulting in a total of nearly 400 patterns. Since pattern evolutions may
257 enter oscillating configurations, some of the generated patterns are identical. We remove such
258 patterns from the stimuli set. Figure 1B shows an example pattern evolution based on a
259 specified rule and Figure 2 shows some of the produced patterns. We also attempted to avoid
260 any recognizable semantic content in our patterns, since complexity perception is largely
261 influenced by familiarity (Forsythe et al., 2008) which could confound our findings.

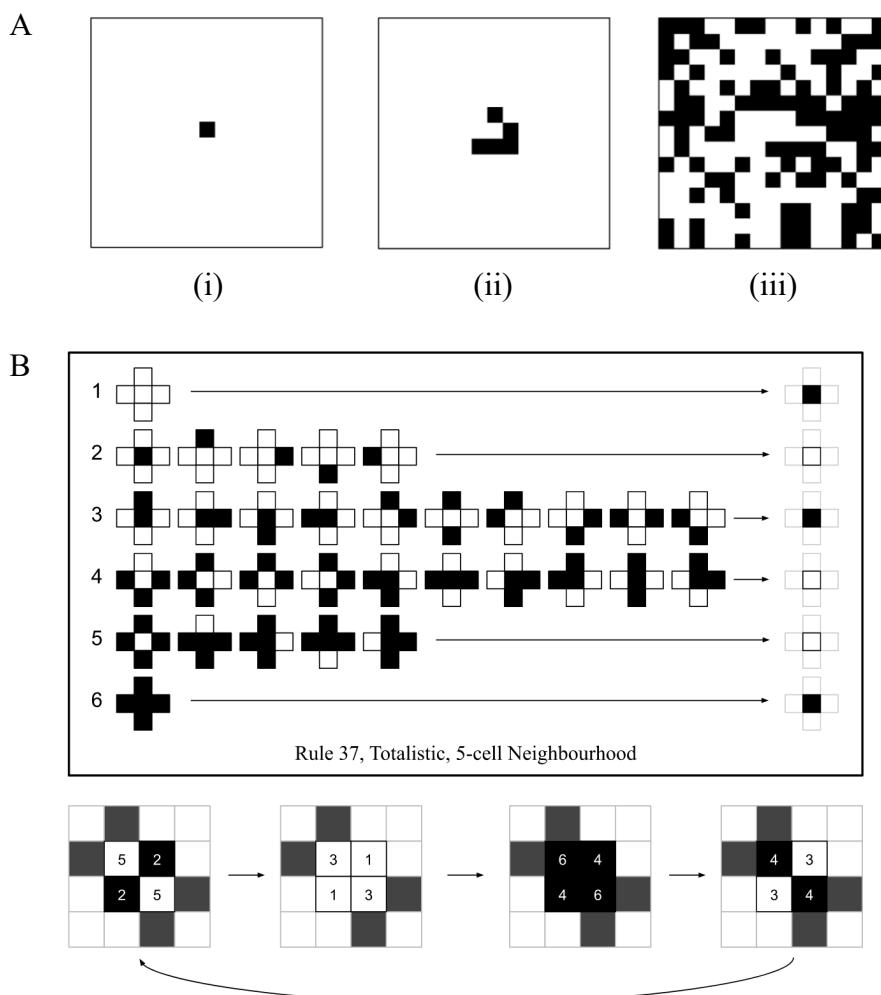
262

263

264 **Figure 1**

265 Cellular automata (CA) (A) 3 initial configurations (ICs) at $t = 0$: (i) IC_1, with a single
 266 central cell (ii) IC_2, with a small disordered region (also called the glider shape in Game of
 267 Life CA (Gardner, 1970)) (iii) IC_3, with a fully random starting configuration - the figure
 268 shown is an example random starting configuration but the pattern was different for each CA
 269 using IC_3 (B) example rule (shown in box) along with example pattern updates. The rule
 270 shown is an example of a 5-neighbourhood totalistic rule. The i^{th} row in the rule is labelled i
 271 and describes the update for a cell with exactly $i-1$ black cells in its neighbourhood,
 272 irrespective of the state of the cell at the time. In this example, row 2 defines the update for a
 273 cell with exactly 1 black cell in its neighbourhood to white irrespective of the state of the cell
 274 at the time. The example pattern trajectory shows the result of the repeated application of the
 275 rule to a starting pattern. Note that the grid size and IC used here are different from what we
 276 use for our pattern generations and are only used here for illustration purposes. In the example
 277 pattern trajectory, only the central 4 cells are updated and the outer rim of cells is kept
 278 constant. The number on the cell defines the update that is being performed on the cell based
 279 on which row in the rule matches the neighbourhood of the cell. In this example, a cell will be
 280 marked 5 if it has exactly 4 black cells in its neighbourhood and would follow the update in
 281 row 5 of the rule hence will change-to or stay white.

282



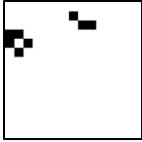
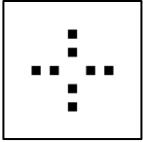
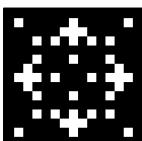
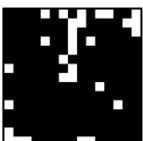
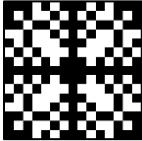
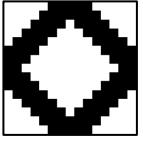
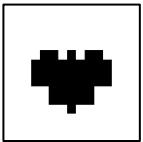
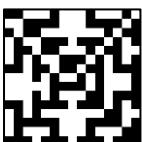
283

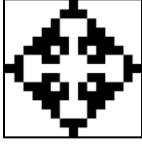
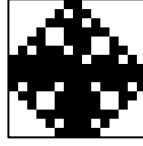
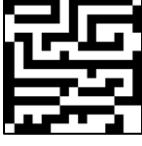
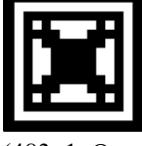
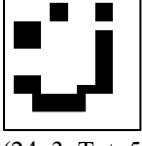
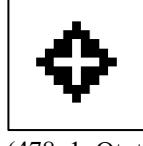
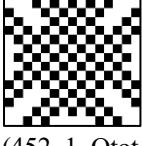
284

285 **Figure 2**

286 Examples of cellular automata generated patterns with (rule code, IC, rule type, N, iteration).
 287 Some (manually identified) visual feature categories of diversity are exemplified, each with 2
 288 example patterns – (1) the proportion of black pixels which can be low or high, (2) degree of
 289 symmetry including full symmetry (about horizontal, vertical, diagonal and rotational axes),
 290 unidirectional symmetry (about either horizontal or vertical axes), partial asymmetry or full
 291 asymmetry, (3) shape of the pattern, for example, diamond or square, (4) style of the pattern,
 292 such as, maze-like, block-like or framed, and (5) number of components in a pattern where a
 293 component is defined as a connected set of black or white cells where diagonal neighbours of
 294 the same colour are not considered part of the same component.

295

Category	Sub-category	Example 1	Example 2
Proportion of black pixels	Low		
	High		
Symmetry	Full symmetry (about horizontal, vertical, diagonal and rotational axes)		
	Unidirectional symmetry (about either horizontal or vertical axes)		
Partial asymmetry			
			
Full asymmetry			

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)	(478, 1, Otot, 5, 5)
High			
		(452, 1, Otot, 5, 25)	(469, 1, Otot, 5, 25)

296

297 CA provide a principled and structured method for pattern generation since it is fully
 298 specified by $\langle L, S, N, f \rangle$, we can deterministically reproduce all our patterns with algorithm
 299 parameters of state space, grid size, neighbourhood size, state-transition function, along with
 300 initial configuration and timestep. Moreover, from Figure 2, we see that the produced patterns
 301 are diverse and vary in multiple features, such as proportion of black pixels, symmetry, shape
 302 or number of components (while this diversity is hard to predict at the outset, some of it is

303 pre-determinable – for example, IC_1 results in fully symmetric patterns).

304

305 2.2 Computational Measures for Pattern Quantification

306

307 We defined six measures that could be potential influencers of subjective ratings: density,
308 entropy, local spatial complexity, Kolmogorov complexity (approximate), intricacy, and
309 symmetry; these are described below. We chose these six measures since they are commonly
310 studied in the literature as potential determinants of complexity and beauty judgements
311 (Arnheim, 1956, 1966; Attneave, 1957; Bense, 1960, 1969; Chikhman et al., 2012; Damiano
312 et al., 2021; Fan et al., 2022; Friedenberg and Liby, 2016; Gartus and Leder, 2017; Javaheri
313 Javid, 2016; Moles, 1958; Nadal, 2007; Rigau, 2008; Schmidhuber, 2009; Singh and Shukla,
314 2017; Silva, 2021; Snodgrass, 1971). We do not consider other popular measures such as
315 number of vertices or edges since as remarked in the previous section, they are hard to define
316 for our CA patterns.

317

- 318 1. Density: Density is defined as the proportion of black pixels in the pattern.

319

- 320 2. We define three different measures of entropy:

321

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

324

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

325

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

329

- 330 (ii) Mean Entropy: Since the entropy measure does not take spatial arrangement
331 into consideration, we compute density entropies at all scales and average
332 them (Eq. 2).

333

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

334

335 where ns is the number of scales (in our case, $ns = P = Q = 15$), nw_s is the
336 number of sliding windows at scale s (defined with overlap $nw_s = (15 - s +$
337 $1)^2$), $P(b)_{s,w}$ and $P(w)_{s,w}$ are the proportions of black and white pixels in the
338 sliding window w at scale s respectively. $P(w)_{s,w} = 1 - P(b)_{s,w}$. The range
339 of this quantity is from 0 to 1 and the unit of measurement is bits.

340

- 341 (iii) Entropy of Means: Huber (2011) proposed a measure of entropy based on
342 spatial proximity which calculated the entropy of a pattern at different levels

343 of smoothing (sliding window averages), resulting in an entropy profile. We
 344 calculate the mean of the entropy profile (Eq. 3).
 345

$$346 - \frac{1}{ns} \sum_{s=1}^{ns} \text{entropy}(\text{mean}(w_{s,1}), \text{mean}(w_{s,2}), \dots, \text{mean}(w_{s,nw_s})) \quad (3)$$

347 where ns is the number of scales (in our case, $ns = P = Q = 15$), nw_s is the
 348 number of sliding windows at scale s (defined with overlap $nw_s = (15 - s +$
 349 $1)^2$), $\text{mean}(w_{a,b})$ is the mean of the b^{th} window at the a^{th} scale, or the mean
 350 of the b^{th} window of size $a \times a$. The mean of a window is simply the sum of
 351 the pixels in the window with black pixels as 1 and white pixels as 0 divided
 352 by the square of the window size. The entropy of the list of S distinct window
 353 means with proportions p_1, p_2, \dots, p_S respectively is calculated as:

$$355 -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_S \log_2 p_S \quad (4)$$

357 In practice, the range of this quantity between 0 and 3.5, and the unit of
 358 measurement is bits.

- 359
- 360 3. Local Spatial Complexity: Local Spatial Complexity (LSC) is defined as the mean
 361 information gain of pixels having homogeneous (same colour) or heterogeneous
 362 (different colour) neighbouring pixels (Javaheri Javid, 2019). This measure takes the
 363 *local* probabilistic spatial distribution of pixels into consideration. The average spatial
 364 complexity across 8 directions of pixel-neighbour pairs is evaluated (Eq. 2). However,
 365 this measure is implemented only across one scale giving it the name *local* spatial
 366 complexity.

$$367 LSC = \frac{1}{8} \sum_{d=1}^8 \bar{G}_d = -\frac{1}{8} \sum_{d=1}^8 \sum_{s_1=1}^2 \sum_{s_2=1}^2 P(s_1, s_2)_d \log_2 P(s_1|s_2)_d \quad (4)$$

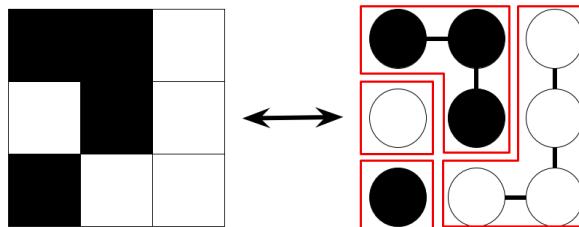
368 Here, d denotes the direction. The direction specifies the positional relationship between
 369 s_1 and s_2 . Assuming s_1 refers to the cell (i, j) , when $d = 1$, s_2 refers to $(i + 1, j)$; when
 370 $d = 2$, s_2 refers to $(i - 1, j)$; when $d = 3$, s_2 refers to $(i, j + 1)$; when $d = 4$, s_2 refers to
 371 $(i, j - 1)$; when $d = 5$, s_2 refers to $(i + 1, j + 1)$; when $d = 6$, s_2 refers to $(i + 1, j - 1)$;
 372 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
 373 denote the state combination (black-black, black-white, white-black or white-white)
 374 under consideration. $P(s_1, s_2)_d$ is the joint probability that a pixel pair (along direction
 375 d) has states (s_1, s_2) . $P(s_1|s_2)_d$ is the conditional probability that a pixel has state s_1
 376 given its neighbouring pixel (along direction d) has state s_2 . \bar{G}_d is the mean information
 377 gain across all state combinations for direction d . LSC is the mean of the 8 directional
 378 \bar{G}_d 's, or the mean of $\{\bar{G}_1, \bar{G}_2, \dots, \bar{G}_8\}$ measured in bits.
 379

- 380
- 381 4. Kolmogorov Complexity: Kolmogorov complexity (KC) is based on algorithmic
 382 information theory. It is defined as the length of the shortest computer program that
 383 produces the desired pattern and can be used as a measure of pattern complexity.
 384 Kolmogorov complexity is uncomputable, and methods to compute it only estimate an
 385 upper bound. Some such methods are the LZ78 universal compression algorithm (Ziv
 386 and Lempel, 1978) and the Block Decomposition Method (Zenil et al., 2018). In this
 387 work, we use the Block Decomposition Method.
- 388
- 389 5. Intricacy: To quantify the number of elements in a pattern, we introduce an intricacy
 390 measure using a graph-based approach (Authors, 2022). The pattern is encoded as a
 391 graph with each pixel as a node. Edges are added between neighbouring pixels of the
 392 same colour. We considered up, down, right and left (non-diagonal) adjacent pixels as
 393 valid neighbours. Depth first search is performed to count the number of connected
 394 components which is used as our intricacy measure. This value is the sum of black
 395 components and white components in the pattern. This procedure is shown in Figure 3
 396 for an example 3×3 grid. For 15×15 stimuli, the range of intricacy values is 1 (for
 397 an all-white or all-black pattern) to 225 (for a checkerboard).
- 398

399 **Figure 3**

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

404



405
 406 *Note. Diagonally adjacent cells of the same colour are not considered part of the same*
 407 *connected component.*

408

- 409 6. Asymmetry: We use three measures to capture the global and local asymmetry in the
 410 patterns. For global asymmetry, we restrict ourselves to the horizontal and vertical
 411 directions as they are most readily perceived by humans (Giannouli, 2013). Horizontal
 412 asymmetry (Hasymm) and vertical asymmetry (Vasymm) were computed as the
 413 percentages of mismatches in the horizontal and vertical directions respectively (refer
 414 to appendix AII for a mathematical description and illustrated example of these
 415 symmetries). Our third asymmetry measure, local asymmetry was computed as the
 416 average difference in mean information gains as specified by the LSC, along 4
 417 directions.
- 418

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

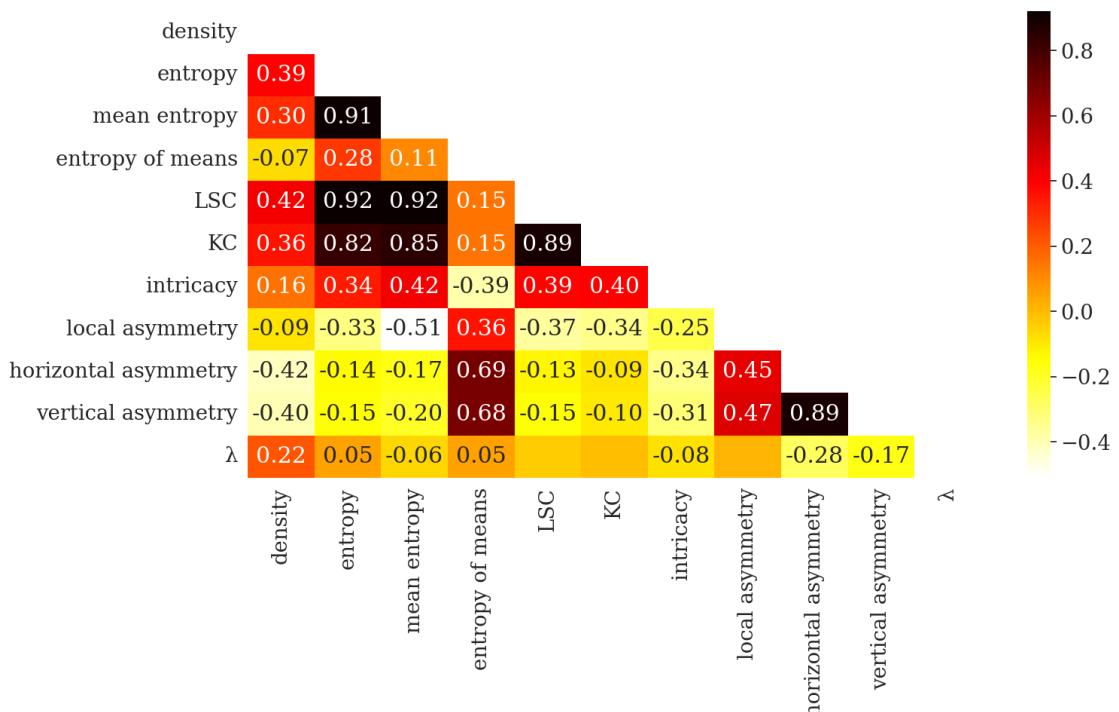
427

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

438

439 **Figure 4**

440 *Pearson correlations between the measures – density, entropy, mean entropy, entropy of
 441 means, local spatial complexity (LSC), Kolmogorov complexity (KC), intricacy, local
 442 asymmetry, horizontal asymmetry, vertical asymmetry and number of rules (λ).*



443

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

446

447 2.3 Pattern Rating Experiment

448

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

452

453 1. Design

454

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

468

469 2. Stimuli

470

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

480

481 3. Procedure

482

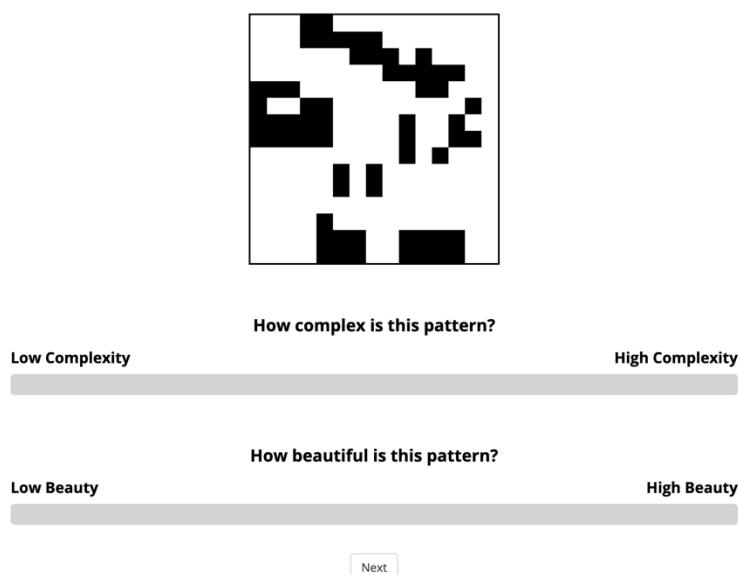
483 The experiment opened with a welcome screen, followed by a consent form and data protection
484 form. An overview page provided the task instructions. Participants then proceeded to the
485 ratings where they were shown a pattern at the top of the screen with 2 sliders positioned below
486 the pattern. The two sliders recorded responses to the two questions – "How complex is this
487 pattern?" and "How beautiful is this pattern?" (Figure 5). The pattern was displayed in the size

488 300 px × 300 px irrespective of screen size. People could use either laptops or desktops to take
489 part in the experiment. The order of questions remained the same across trials and across
490 participants.

491

492 **Figure 5**

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



498

499

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

511

512 4. Participants

513

514 80 participants from Prolific (50 female, 29 male, 1 other; mean age = 32.3, min age = 18, max
515 age = 79; 20 participants per set) took part in the study. 2 participants failed all attention checks
516 and were removed from the analysis leaving us with 78 participants. All participants were based
517 in the United States, were fluent English speakers and had not previously participated in the

518 experiment. The average study completion time was 14.35 minutes. Each participant was paid
519 £3.50. All experiments were approved by the ethics committee of the University of Tübingen.
520

521 5. Analysis

522

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

526

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

535

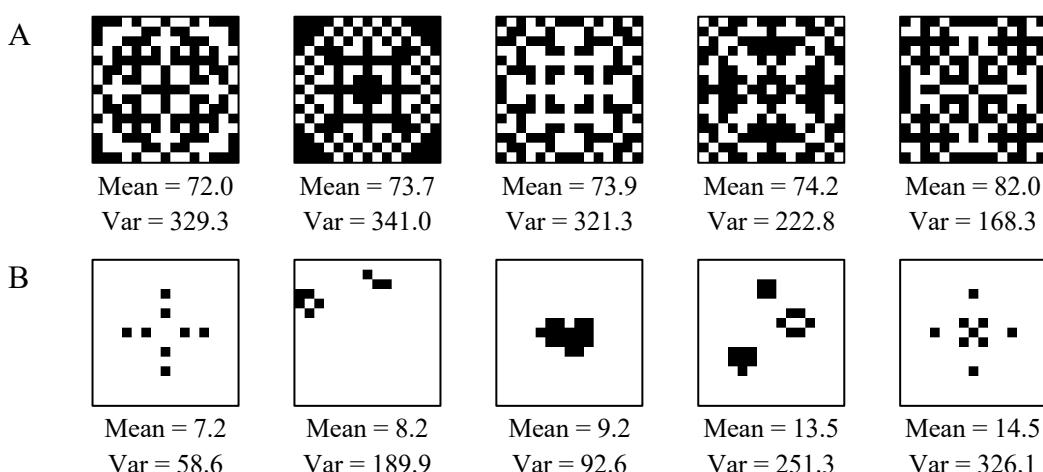
536 Figures 6A and 6B show the patterns with the highest and lowest average complexity ratings
537 and Figures 6C and 6D show the patterns with the highest and lowest average beauty ratings.
538 We also studied variance in ratings per pattern and categorized patterns into “high agreement”
539 (low variance)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

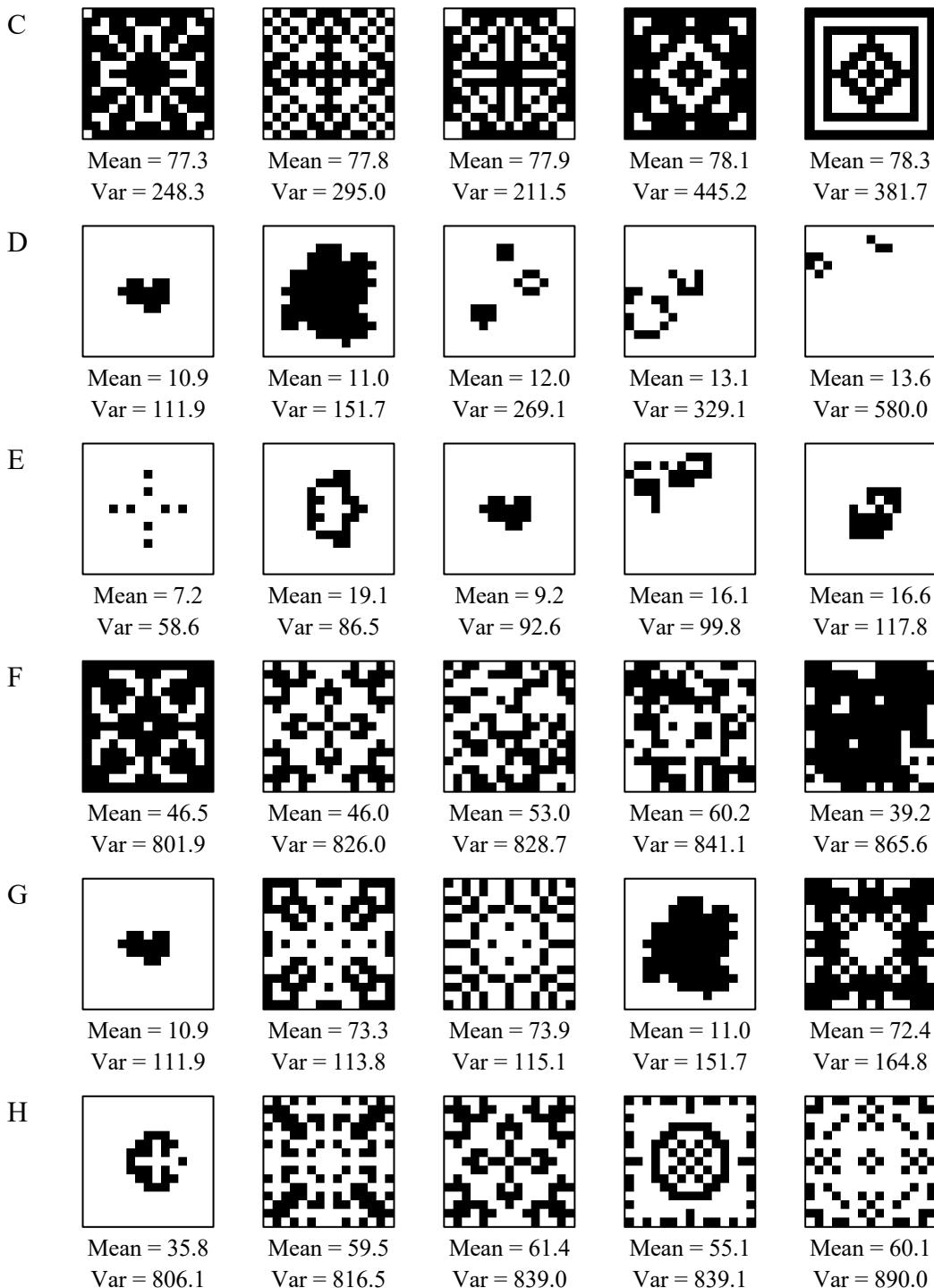
541

542 **Figure 6**

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

549





550

551 *Metric for Complexity*

552

553 We fit regression models on participant complexity ratings using R (Version 4.1.1, library lme4
 554 (Bates et al., 2015), function lmer()). We used an incremental, bottom up approach (*i.e.*, starting
 555 from the predictors that had high linear correlations with our dependent variable and adding
 556 predictors one by one to our models (Table 1)), to arrive at the objective complexity metric that
 557 best predicted subjective complexity. High linear correlations between the ratings and the
 558 measures justify the use of linear models; hence we fit linear mixed effects models. We

559 performed cross validation by splitting our data into 3 stratified folds. Each test fold had 20
 560 randomly sampled ratings from every participant and each training fold had the remaining 40
 561 ratings from every participant. The seed was set to 20 for all our experiments. We z-scored all
 562 the predictor variables. We also created additional variables by squaring each predictor to
 563 check for quadratic effects. We used the *bobyqa* optimizer in R with all other parameters same
 564 as default. The three asymmetry measures were combined into one mean asymmetry measure
 565 because of their high correlations. We evaluated our model using 3-fold cross validation and
 566 reported the average Akaike information criterion (AIC), Bayesian information criterion (BIC)
 567 and the average R^2 values across the three folds. We also evaluated Root Mean Square Error
 568 (RMSE) values on train and test sets. In addition, we checked for multicollinearity using
 569 Variance Inflation Factors (VIF). We experimented with random intercepts, random slopes,
 570 quadratic and interaction effects in our models. The best model was defined as the one that had
 571 significant predictors and resulted in a low BIC and high R^2 while ensuring VIFs did not exceed
 572 5. If a model achieved higher R^2 at the expense of higher BIC, we prioritized model simplicity.
 573

Relationship between Beauty and Complexity Ratings

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

585 2. Results

586 3.1 Metric for Complexity

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

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	0.91	0.93
2	CR ~ 1 + 1 Participant + 1 Set		8546.3	8570.5	0.16	0.13	0.91	0.93
3	CR ~ LSC + 1 Participant	LSC*	7551.9	7576.1	0.39	0.37	0.77	0.79

4	$CR \sim LSC + LSC Participant$	LSC*	7491.2	7527.5	0.43	0.39	0.75	0.78
5	$CR \sim KC + 1 Participant$	Intercept* KC*	7703.8	7728.0	0.36	0.34	0.79	0.81
6	$CR \sim LSC + density + 1 Participant$	LSC* density*	7551.3	7581.5	0.39	0.37	0.77	0.79
7	$CR \sim LSC + entropy + 1 Participant$	LSC* entropy*	7536.6	7566.8	0.40	0.37	0.77	0.79
8	$CR \sim LSC + mean_entropy + 1 Participant$	LSC* mean_entropy*	7550.7	7580.9	0.40	0.37	0.77	0.79
9	$CR \sim LSC + entropy_of_means + 1 Participant$	LSC* entropy_of_means*	7443.8	7474.0	0.41	0.39	0.76	0.79
10	$CR \sim LSC + asymm + 1 Participant$	LSC* asymm*	7525.6	7555.9	0.40	0.38	0.77	0.79
11	$CR \sim LSC + intricacy + 1 Participant$	LSC* intricacy*	7285.1	7315.3	0.45	0.42	0.74	0.76
12	$CR \sim LSC + intricacy + LSC:intricacy + 1 Participant$	LSC* intricacy* LSC:intricacy*	7276.8	7313.0	0.45	0.43	0.74	0.75
13	$CR \sim LSC + intricacy + LSC Participant$	LSC* intricacy*	7219.8	7262.2	0.48	0.44	0.71	0.74
14	$CR \sim LSC + intricacy + intricacy Participant$	LSC* intricacy*	7166.3	7208.6	0.50	0.46	0.70	0.73
15	$CR \sim LSC + intricacy + (LSC + intricacy) Participant$	LSC* intricacy*	7138.8	7199.3	0.51	0.47	0.69	0.73
16	$CR \sim LSC + intricacy + LSC:intricacy + (LSC + intricacy) Participant$	LSC* intricacy* LSC:intricacy*	7133.5	7200.0	0.51	0.47	0.69	0.73
17	$CR \sim LSC^2 + intricacy^2 + 1 Participant$	LSC^2* intricacy ² *	7329.3	7359.5	0.44	0.41	0.74	0.76
18	$CR \sim LSC + LSC^2 + intricacy + intricacy^2 + 1 Participant$	LSC^2* intricacy* intricacy ² *	7256.6	7299.0	0.45	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).*

600

$CR \sim LSC + intricacy + (LSC + intricacy)|Participant$

Intercept (mean)	LSC	Intricacy
-0.02	0.37	0.25

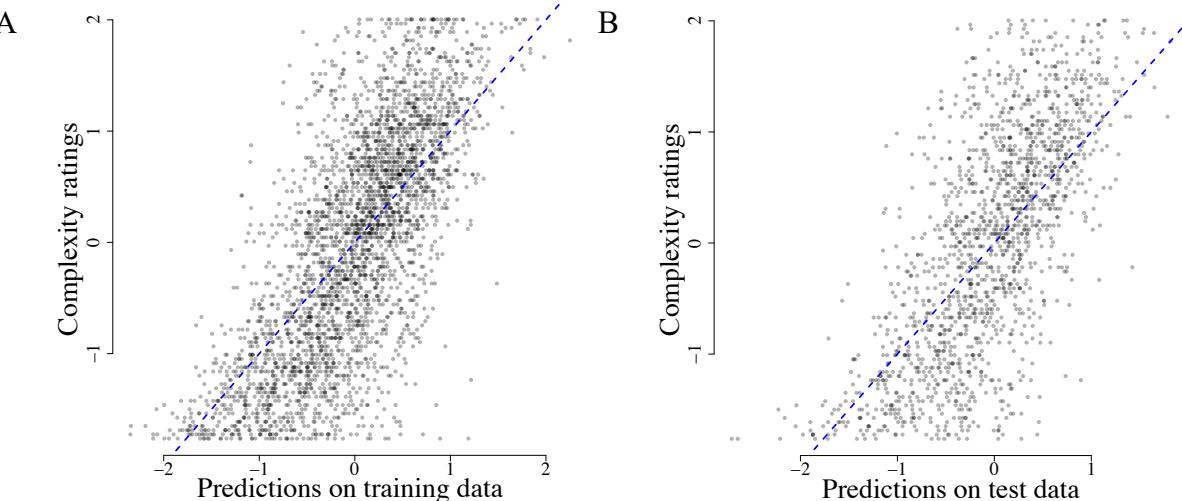
601

602 Our analysis shows that subjective complexity can be predicted by a positive linear
603 combination of LSC and intricacy measures with random slopes of LSC and intricacy, and a
604 random intercept of participant (Table 1, row 15, $R^2_{\text{test}} = 0.47$, AIC = 7138.8, BIC = 7199.3).
605 While the AIC for the model with an interaction effect is slightly lower (Table 1, row 16), the
606 BIC is slightly higher implying possible overfitting, and so we preferred the simpler model
607 (Table 1, row 15). Figure 7 shows the plot of predictions versus ground truth on the train and
608 test set from cross-validation fold 1.

609

610 **Figure 7**

611 *Performance of the best model $CR \sim LSC + intricacy + (LSC + intricacy) | \text{Participant}$ on (A)*
612 *training and (B) test data from fold 1 for complexity ratings*



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

615

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

627

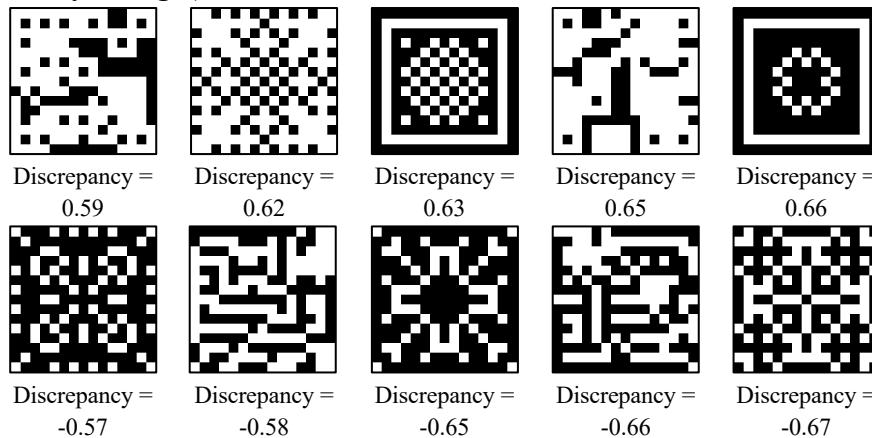
628 We also visualized some of the patterns for which the difference between objective complexity
629 measure and subjective ratings was the largest on average across all participants (Figure 8). We
630 see that the measure overweights complexity in some patterns where the computed intricacy
631 measure is high (top row in Figure 8. Refer to Appendix II for additional analysis using an 8-

neighbourhood variant of intricacy), and underweights complexity in some of the high density and (partially) asymmetric patterns (bottom row in Figure 8).

634

635 **Figure 8**

636 *Patterns with largest (z-scored) overestimation of complexity by the model (row 1, increasing
637 in magnitude from right to left) and largest (z-scored) underestimation (row 2, increasing in
638 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 of means where the
649 weights are obtained from the fixed effects of asymmetry and entropy of means respectively in
650 the model $BR \sim CR + \text{asymmetry} + \text{entropy_of_means} + I|Participant$.

651

652 **Table 2**

653 *A Models of beauty ratings*

Id	Model	Significance	AIC	BIC	R^2		RMSE	
					Train	Test	Train	Test
1	$BR \sim CR + CR^*$ $1 Participant$		8046.4	8070.6	0.29	0.25	0.85	0.87
2	$BR \sim CR + \text{disorder} + CR^*$ $1 Participant$	disorder^*	6449.6	6479.8	0.58	0.55	0.65	0.67
3	$BR \sim CR + \text{disorder} + CR^*$ $\text{disorder} Participant$	disorder^*	5985	6027.3	0.67	0.63	0.58	0.61
4	$BR \sim CR + \text{disorder} + CR^*$ $CR:\text{disorder} + \text{disorder}^*$ $\text{disorder} Participant$	$CR:\text{disorder}^*$	5919.1	5967.4	0.68	0.64	0.57	0.60

5	$BR \sim CR^2 + disorder + disorder Participant$	6872.3	6902.6	0.52	0.49	0.69	0.71
6	$BR \sim CR + CR^2 + disorder + 1 Participant$	6457.9	6494.2	0.58	0.55	0.65	0.67
7	$BR \sim LSC + intricacy + disorder + 1 Participant$	6685.8	6722	0.55	0.52	0.67	0.69
8	$BR \sim LSC + intricacy + disorder + intricacy Participant$	6185	6233.4	0.65	0.61	0.59	0.63
9	$BR \sim LSC + intricacy + disorder + (LSC + intricacy):disorder + LSC:disorder Participant$	6088.1	6148.6	0.66	0.62	0.58	0.62

654

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

656

BR ~ CR + disorder + CR:disorder|Participant

Intercept (mean)	CR	disorder	CR:disorder
-0.02	0.27	-0.57	-0.11

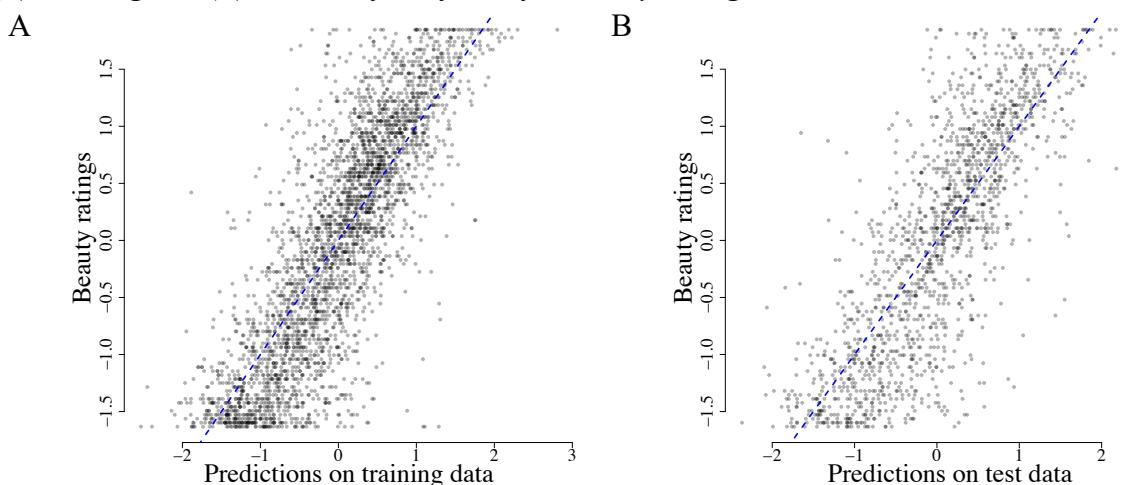
657

658 We found that the beauty ratings are well predicted by complexity ratings, disorder and their
 659 interaction, along with a random intercept of participant and a random slope of disorder (Table
 660 2, row 4, $R^2 = 0.64$). Figure 9 shows the plot of predictions versus ground truth on a train and
 661 test set from cross-validation fold 1.

662

Figure 9

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



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

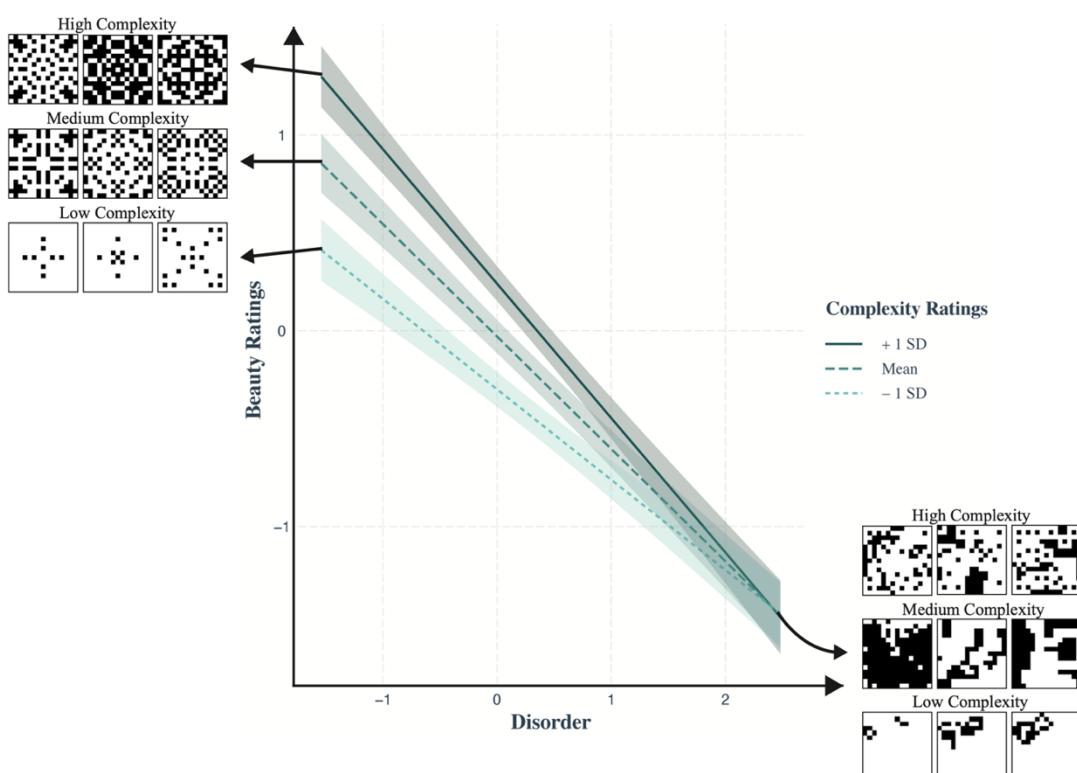
668

669 Beauty ratings correlated positively with complexity ratings but negatively with disorder (mean
 670 asymmetry and entropy). Studying the interaction effect (Figure 10) suggested that high
 671 complexity is considered beautiful as long as the amount of disorder is low; and when the
 672 disorder is high, beauty is low irrespective of complexity. In other words, people prefer
 673 complexity (or find it more beautiful) while ensuring order. Further, the random slope of
 674 disorder implies that different people may have a different degree of perceived dislike towards
 675 disorder. Adding a quadratic effect of complexity to the model however, did not lead to a
 676 significant performance enhancement (Table 2, rows 5-6).

677

678 **Figure 10**

679 *Visualisation of the interaction effect between complexity ratings and disorder. The*
 680 *relationship between beauty ratings (y-axis) and disorder (x-axis) is shown at three levels of*
 681 *complexity ratings (high, medium and low). Example patterns at the three levels of complexity*
 682 *are shown in the left and right panels. The procedure for obtaining the example patterns is as*
 683 *follows. First, all the patterns are binned by disorder –bins 1 – 9 linearly defined based on*
 684 *pattern disorder value). For the left and right panel examples respectively, patterns with the*
 685 *two lowest and highest disorder bins are shown, sorted in ascending order by mean complexity*
 686 *ratings. The last 3 (+1 SD), mid 3 (Mean) and first 3 (-1 SD) patterns are then picked as*
 687 *examples of high, medium and low complexity respectively. It is found that at high disorder,*
 688 *beauty of a pattern is always low, but at low disorder, beauty of the pattern increases with*
 689 *increasing complexity.*



690
 691

692 In sum, across both analyses, we observe a dissociation between the factors influencing
 693 perceived complexity and beauty. While subjective complexity can be explained by an

694 integration of objective measures that encode local and global image features, beauty is
695 explained by an interaction between complexity and disorder.

696

697 2.3 3.3 Relationship between Objective Complexity, Subjective Complexity and Beauty – 698 Moderated Mediation Analysis

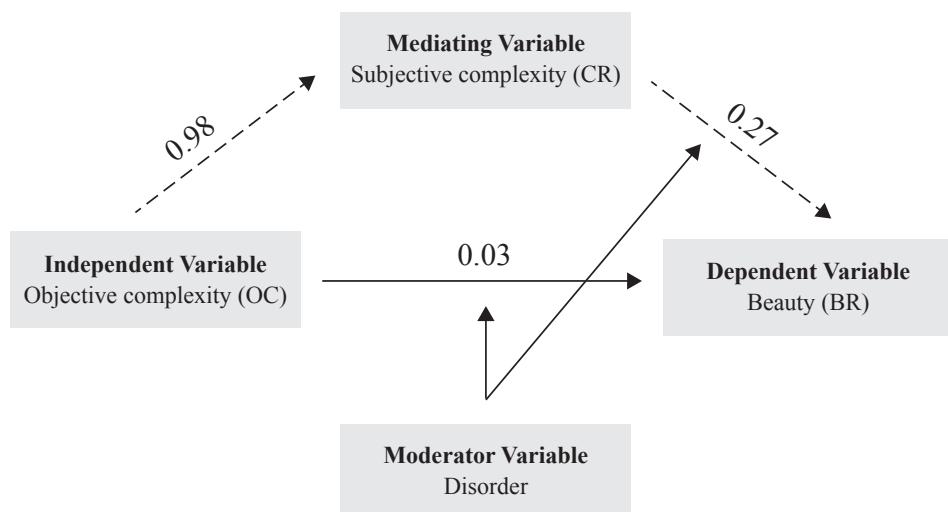
700 We also attempted to find an objective model of beauty, without the explicit use of the
701 complexity ratings. Our previous analysis implies that beauty is well explained by a
702 combination of purely objective measures, namely – LSC, intricacy (which together predict
703 subjective complexity), and disorder. This model explains 62% of the variance in ratings (From
704 Table 2, row 8). We examine the relationship between objective complexity, disorder,
705 subjective complexity and beauty more formally using moderated mediation analysis.

707 Figure 11 shows the underlying moderated mediation structure of the variables of interest from
708 our data. Objective complexity (OC) was evaluated as the weighted combination of LSC and
709 intricacy, where the weights were obtained from the regression coefficients (including random
710 effects) of the best performing model (Table 1, row 12) fit on complexity ratings from cross
711 validation fold 1. We performed moderated mediation analysis using PROCESS macro model
712 15 in R. Due to limited data and to avoid overfitting, we only used the average model
713 description (*i.e.*, excluding random effects) for our analysis. We fit the PROCESS model on
714 data from cross-validation data fold 2.

715

716 **Figure 11**

717 *Mediation structure. Objective complexity (OC) interacts with Disorder to predict Beauty*
718 *(BR). Subjective complexity (CR) can be predicted by OC. In the presence of CR, OC is not a*
719 *significant predictor of BR. Therefore, CR mediates the influence of OC on BR. Further,*
720 *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 disorder, but a non-significant effect at high disorder.
730 There was a significant conditional indirect effect of objective complexity on beauty. However,
731 the index of moderated mediation was found to be non-significant. This implies that the
732 interaction effect with disorder did not influence the mediation between objective complexity,
733 subjective complexity and beauty. We studied these direct and indirect effects more closely
734 using regressions now including random effects while ignoring the interaction effect with
735 disorder. The OC computation here excludes the random effects since the random effects
736 structure in the model includes subject specific slope and intercepts. We use the cross-
737 validation data fold 1 to fit these models. Table 3 shows the results from the regressions.
738

739 **Table 3**

740 *Regressions to study mediation*

S. no.	Model	Significance	OC slope coefficient
1	BR ~ OC + disorder + OC* disorder Participant	disorder*	0.31
2	CR ~ OC + intricacy Participant	OC*	0.98
3	BR ~ CR + OC + disorder + CR* disorder Participant	disorder*	0.03

741 Note. BR=beauty ratings, OC=objective complexity, CR=complexity ratings

742 **Table 4**

743 *Results of mediation analysis*

	Estimate	95% CI	p-value
ACME	0.27	[0.22, 0.32]	0.000
ADE	0.03	[-0.03, 0.11]	0.29

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

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

759 **3. Discussion**
760

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 course-
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 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; Attnave, 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 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. Therefore, these measures add
820 complementary information which are suitably integrated to give rise to human complexity
821 evaluations.

822

823 Contrary to prior work (Arnoult, 1960; Attneave, 1957; Day, 1967; Eisenman and Gellens,
824 1968; Marin and Leder, 2013; Redies and Brachmann, 2017), neither symmetry nor entropy
825 related to subjective complexity. This could be explained by the simplified black-and-white
826 pixel nature of our stimuli as opposed to natural scenes, or the significant correlations between
827 these measures and intricacy. Further, we observed a disparity between participant strategy
828 responses and our metric – participants indicated that their ratings depended on their ability to
829 create or replicate the pattern (which could be seen as a direct link to algorithmic complexity),
830 while our approximate Kolmogorov complexity (KC) was not predictive of subjective
831 complexity as per our model. This could again be due to the large underlying correlations of
832 KC, asymmetry and entropy with LSC or intricacy, in turn masking their effect.

833

834 Relationship between Subjective Complexity and Beauty

835

836 In contrast, however, asymmetry and entropy did relate to beauty judgements. Beauty ratings
837 correlated positively with subjective complexity and negatively with asymmetry and entropy.
838 More specifically, the entropy-of-means measure performed better than conventional single-
839 scale entropy or mean entropy (Appendix AII shows some examples of patterns with all
840 measures). Through this, our work has lent support for a monotonic relationship between
841 beauty and complexity. This goes against the inverted-U like dependence proposed by Berlyne.
842 However, one reason for this, as stated above and by others (Krupinski and Locher, 1988;
843 Nicki, Lee, and Moss, 1981; Stamps III, 2002), could be that our stimuli are so simple in nature

844 as to lie in the lower quantiles of complexity. If true, then we would only expect to be able to
845 reproduce the first half of the inverted-U curve, and as a result could not falsify a linear
846 relationship.

847

848 A weighted combination of complexity ratings, disorder (itself a weighted combination of
849 asymmetry and entropy) and their interaction (along with random slopes for disorder and
850 random intercepts for participants) effectively modelled beauty ratings. This means that there
851 is beauty in complexity as long as the disorder is low, or alternatively, order is necessary for
852 beauty and complexity adds to beauty once order is present. This concurs with the work
853 proposing that beauty lies in the balance between order and complexity (refer to Van Geert and
854 Wagemans, 2020 for a review). For example, Arnheim (1966, p. 124) stated: “Complexity
855 without order produces confusion. Order without complexity causes boredom”. However,
856 based on our linear model (with interactions), we cannot lend support to Birkhoff’s (1933)
857 proposed $M = O / C$ relationship, or Eysenck’s (1941, 1942) $M = O \times C$ relationship. The
858 interaction effect we found emphasises the relative influence of order and complexity on beauty
859 – complexity is beautiful, as long as the degree of disorder is low, and at high disorder, beauty
860 is consistently low. Here we can return to our initial question about why geometric tile designs
861 are more beautiful than chess boards or QR-codes—tile designs are more beautiful than chess
862 boards as they are more complex, but more beautiful than QR-codes as they are less disordered.
863 This result is at odds with Van Geert and Wageman’s (2021) suggestion that a balance between
864 order and complexity involves no interaction. They, however used real world images of neatly
865 organized compositions and recorded fascination and soothingness judgements in a 2-choice
866 task. The contradiction between their and our findings underlines the challenge of comparing
867 results across varying stimuli types and task designs and stresses the fact that the added value
868 of the interaction term can depend on the stimuli and the specific operationalizations of order
869 and complexity used (Van Geert and Wagemans, 2020).

870

871 Relationship between Objective Complexity, Subjective Complexity and Beauty

872

873 While a combination of pure objective measures of LSC, intricacy, and disorder (asymmetry
874 and entropy of means) was able to explain 62% of the variance in beauty ratings, formal
875 analysis of the relationship between subjective complexity, objective complexity and beauty
876 using moderated mediation analysis revealed that subjective complexity mediates the influence
877 of objective complexity on beauty at all levels of disorder. This indicates that subjective
878 complexity encodes information beyond what is expressed in terms of objective complexity
879 measures. For this reason, some views criticize such methods attempting to quantify subjective
880 complexity in objective terms. Heckhausen (1964) argued that relating subjective complexity
881 to simple visual properties of stimuli as done by information theory approaches is insufficient.
882 He claimed that the subjective complexity does not solely depend on the complexity of the
883 stimulus but also on the way it is perceived. Attneave (1957) also suggested that people’s
884 perception of complexity is not a mere reflection of the visual stimuli. This explanation aligns
885 with Gestalt philosophies of “perceptual organisation” often summarised as “whole is greater
886 than sum of the parts”. This need was also highlighted by Berlyne who had claimed that
887 complexity was a property of both the physical stimulus properties and the processes within

888 the subject. Having said that, one must clarify here that using only the average model
889 description in the PROCESS moderated mediation analysis may have removed the effect of
890 disorder's influence on the relation between complexity and beauty for different people.
891 Further, no causal implications can be made from this moderated mediation analysis, and the
892 relationship between subjective and objective quantities observed here are purely correlational.
893 Our work is also restricted just to *how* complexity relates to beauty without looking into *why*
894 complexity relates to beauty the way it does.

895

896 Limitations and Future Directions

897

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

903 Further, there are limitations on the generality associated with the specific type of stimuli. Our
904 measures are defined specifically for 2D binary pixel patterns. While our CA generation
905 algorithm can be manipulated to produce newer families of patterns with increased size or
906 added colours, and our measures can be adapted to apply to patterns with such transformations,
907 this would only achieve small-scale generalisation. There are certain advantages of testing for
908 small-scale generalisation – it provides a systematic method to identify causes for the
909 breakdown of generalisation. For example, if the complexity metric fails to explain variance
910 upon adding a third colour, we can isolate the factor of colour, across which generalisation
911 does not hold. Explicit tests for this have not been addressed in this article and would be a part
912 of future work.

914

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

932

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

941

942 Conclusion

943

944 Our work develops a class of diverse, algorithmic 2D binary pixel patterns which can be
945 reproduced and manipulated easily. A set of programmatic pattern quantification measures
946 were used to understand the relationship between objective complexity, subjective complexity,
947 and beauty. We found that people's complexity ratings depend on an integration of global and
948 local pattern properties and found that there is beauty in complexity as long as there is sufficient
949 order. We also noted that subjective complexity cannot be explained fully with objective
950 complexity measures. Through this, our work showcases the usefulness of computational
951 methods to understand the link between assessments of complexity and beauty.

952

953 **References**

954

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

957

958 Adkins, O. C., & Norman, J. F. (2016). The visual aesthetics of snowflakes. *Perception*, 45(11),
959 1304-1319.

960

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

963

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

966

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

969

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

972

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

975

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

978

- 979 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
980 Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- 981
- 982 Bense, M. (1960). Programmierung des Schönen Allgemeine Texttheorie Und Textästhetik.
- 983
- 984 Bense, M. (1969). *Kleine abstrakte ästhetik*. Edition Rot.
- 985
- 986 Berlyne, D. E. (1960). Conflict, arousal, and curiosity.
- 987
- 988 Berlyne, D. E. (1967). Arousal and reinforcement. In *Nebraska symposium on motivation*.
989 University of Nebraska Press.
- 990
- 991 Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. (1968). The dimensionality of visual
992 complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue
993 canadienne de psychologie*, 22(5), 376.
- 994
- 995 Berlyne, D. E. (1973). Aesthetics and psychobiology. *Journal of Aesthetics and Art Criticism*,
996 31(4).
- 997
- 998 Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like
999 curvature or do they dislike angularity?. *British Journal of Psychology*, 107(1), 154-178.
- 1000
- 1001 Birkhoff, G. D. (1933). Aesthetic Measure.
- 1002
- 1003 Birkin, G. (2010). *Aesthetic complexity: practice and perception in art & design*. Nottingham
1004 Trent University (United Kingdom).
- 1005
- 1006 Briellmann, A. A., & Dayan, P. (2022). A computational model of aesthetic value.
1007 *Psychological Review*.
- 1008
- 1009 Chamberlain, R. (2022). The interplay of objective and subjective factors in empirical
1010 aesthetics. In *Human Perception of Visual Information* (pp. 115-132). Springer, Cham.
- 1011
- 1012 Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity
1013 of images: Experimental and computational estimates compared. *Perception*, 41(6), 631-647.
- 1014
- 1015 Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental
1016 Psychology: General*, 106(3), 269.
- 1017
- 1018 Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on
1019 complexity judgments in children and adults. *Journal of Experimental Psychology: Human
1020 Perception and Performance*, 5(2), 365.
- 1021
- 1022 Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of
1023 the literature. *Psychology of Music*, 45(6), 886-909.
- 1024
- 1025 Corchs, S. E., Ciocca, G., Bricolo, E., & Gasparini, F. (2016). Predicting complexity perception
1026 of real world images. *PloS one*, 11(6), e0157986.
- 1027

- 1028 Cupchik, G. C. (1986). A decade after Berlyne: New directions in experimental aesthetics.
1029 *Poetics*, 15(4-6), 345-369.
- 1030
- 1031 Damiano, C., Wilder, J., Zhou, E. Y., Walther, D. B., & Wagemans, J. (2021). The role of local
1032 and global symmetry in pleasure, interest, and complexity judgments of natural
1033 scenes. *Psychology of Aesthetics, Creativity, and the Arts*.
- 1034
- 1035 Davis, R. C. (1936). An evaluation and test of Birkhoff's aesthetic measure formula. *The
1036 Journal of General Psychology*, 15(2), 231-240.
- 1037
- 1038 Day, H. Y. (1967). Evaluations of subjective complexity, pleasingness and interestingness for
1039 a series of random polygons varying in complexity. *Perception & Psychophysics*, 2(7), 281-
1040 286.
- 1041
- 1042 Day, H. (1968). The importance of symmetry and complexity in the evaluation of complexity,
1043 interest and pleasingness. *Psychonomic Science*, 10(10), 339-340.
- 1044
- 1045 Delplanque, J., De Loof, E., Janssens, C., & Verguts, T. (2019). The sound of beauty: How
1046 complexity determines aesthetic preference. *Acta Psychologica*, 192, 146-152.
- 1047
- 1048 Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity.
1049 *Perception*, 35(6), 823-835.
- 1050
- 1051 Eisenman, R. (1967). Complexity-simplicity: I. Preference for symmetry and rejection of
1052 complexity. *Psychonomic Science*, 8(4), 169-170.
- 1053
- 1054 Eisenman, R., & Gellens, H. K. (1968). Preferences for complexity-simplicity and symmetry-
1055 asymmetry. *Perceptual and Motor Skills*, 26(3), 888-890.
- 1056
- 1057 Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological
1058 Review*, 48(1), 83.
- 1059
- 1060 Eysenck, H. J. (1942). The experimental study of the 'good Gestalt'—a new approach.
1061 *Psychological Review*, 49(4), 344.
- 1062
- 1063 Eysenck, H. J. (1968). An experimental study of aesthetic preference for polygonal figures.
1064 *The Journal of General Psychology*, 79(1), 3-17.
- 1065
- 1066 Fan, Z. B., Li, Y. N., Zhang, K., Yu, J., & Huang, M. L. (2022). Measuring and evaluating the
1067 visual complexity of Chinese ink paintings. *The Computer Journal*, 65(8), 1964-1976.
- 1068
- 1069 Farley, F. H., & Weinstock, C. A. (1980). Experimental aesthetics: Children's complexity
1070 preference in original art and photoreproductions. *Bulletin of the Psychonomic Society*, 15(3),
1071 194-196.
- 1072
- 1073 Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of
1074 complexity and familiarity in basic-level picture processing. *Behavior research methods*,
1075 40(1), 116-129.
- 1076

- 1077 Friedenberg, J., & Liby, B. (2016). Perceived beauty of random texture patterns: A preference
1078 for complexity. *Acta psychologica*, 168, 41-49.
- 1079
- 1080 Gardner, M. (1970). The Fantastic Combinations of John Conway's New Solitaire Game'Life.
1081 *Sc. Am.*, 223, 20-123.
- 1082
- 1083 Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of
1084 broken symmetries. *I-Perception*, 4(5), 361-364.
- 1085
- 1086 Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns
1087 using computational measures: The influence of mirror symmetry on complexity
1088 perception. *PloS one*, 12(11), e0185276.
- 1089
- 1090 Giannouli, V. (2013). Visual symmetry perception. *Encephalos*, 50, 31-42.
- 1091
- 1092 Gordon, J., & Gridley, M. C. (2013). Musical preferences as a function of stimulus complexity
1093 of piano jazz. *Creativity Research Journal*, 25(1), 143-146.
- 1094
- 1095 Güçlütürk, Y., Jacobs, R. H., & Lier, R. V. (2016). Liking versus complexity: Decomposing
1096 the inverted U-curve. *Frontiers in Human Neuroscience*, 10, 112.
- 1097
- 1098 Hall, A. C. (1969). Measures of the complexity of random black and white and coloured
1099 stimuli. *Perceptual and motor skills*, 29(3), 773-774.
- 1100
- 1101 Harsh, C. M., Beebe-Center, J. G., & Beebe-Center, R. (1939). Further evidence regarding
1102 preferential judgment of polygonal forms. *The Journal of Psychology*, 7(2), 343-350.
- 1103
- 1104 Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- 1105
- 1106 Heath, T., Smith, S. G., & Lim, B. (2000). Tall buildings and the urban skyline: The effect of
1107 visual complexity on preferences. *Environment and behavior*, 32(4), 541-556.
- 1108
- 1109 Heckhausen, H. (1964). Complexity in perception: Phenomenal criteria and information
1110 theoretic calculus—A note on D. Berlynes" Complexity Effects.".
- 1111
- 1112 Hekkert, P., & Van Wieringen, P. C. (1990). Complexity and prototypicality as determinants
1113 of the appraisal of cubist paintings. *British journal of psychology*, 81(4), 483-495.
- 1114
- 1115 Ichikawa, S. (1985). Quantitative and structural factors in the judgment of pattern complexity.
1116 *Perception & psychophysics*, 38(2), 101-109.
- 1117
- 1118 Igaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., & O'Doherty, J. P. (2020). Aesthetic preference
1119 for art emerges from a weighted integration over hierarchically structured visual features in the
1120 brain. *BioRxiv*.
- 1121
- 1122 Jacobsen, T. (2010). Beauty and the brain: culture, history and individual differences in
1123 aesthetic appreciation. *Journal of anatomy*, 216(2), 184-191.
- 1124
- 1125

- 1126 Jacobsen, T., & Höfel, L. E. A. (2002). Aesthetic judgments of novel graphic patterns:
1127 Analyses of individual judgments. *Perceptual and motor skills*, 95(3), 755-766.
- 1128
- 1129 Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of
1130 aesthetic judgment of beauty. *Neuroimage*, 29(1), 276-285.
- 1131
- 1132 Javaheri Javid, M. A., Blackwell, T., Zimmer, R., & Al-Rifaie, M. M. (2016, March).
1133 Correlation between human aesthetic judgement and spatial complexity measure.
1134 In *International Conference on Computational Intelligence in Music, Sound, Art and*
1135 *Design* (pp. 79-91). Springer, Cham.
- 1136
- 1137 Javaheri Javid, M. A. (2019). *Aesthetic Automata: Synthesis and Simulation of Aesthetic*
1138 *Behaviour in Cellular Automata* (Doctoral dissertation, Goldsmiths, University of London).
- 1139
- 1140 Javaheri Javid, M. A. (2021, April). Aesthetic Evaluation of Cellular Automata Configurations
1141 Using Spatial Complexity and Kolmogorov Complexity. In *International Conference on*
1142 *Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 147-160).
1143 Springer, Cham.
- 1144
- 1145 Krupinski, E., & Locher, P. (1988). Skin conductance and aesthetic evaluative responses to
1146 nonrepresentational works of art varying in symmetry. *Bulletin of the Psychonomic Society*,
1147 26(4), 355-358.
- 1148
- 1149 Lakhal, S., Darmon, A., Bouchaud, J. P., & Benzaquen, M. (2020). Beauty and structural
1150 complexity. *Physical Review Research*, 2(2), 022058.
- 1151
- 1152 Langton, C. G. (1986). Studying artificial life with cellular automata. *Physica D: Nonlinear*
1153 *Phenomena*, 22(1-3), 120-149.
- 1154
- 1155 Li, W., Packard, N. H., & Langton, C. G. (1990). Transition phenomena in cellular automata
1156 rule space. *Physica D: Nonlinear Phenomena*, 45(1-3), 77-94.
- 1157
- 1158 Machotka, P. (1980). Daniel Berlyne's contributions to empirical aesthetics. *Motivation and*
1159 *Emotion*, 4(2), 113-121.
- 1160
- 1161 Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music
1162 regardless of its complexity: Implications for the appreciation and aesthetics of music.
Frontiers in neuroscience, 11, 147.
- 1163
- 1164 Marin, M. M., & Leder, H. (2013). Examining complexity across domains: relating subjective
1165 and objective measures of affective environmental scenes, paintings and music. *PloS one*, 8(8),
1166 e72412.
- 1167
- 1168 Marin, M. M., Lampatz, A., Wandl, M., & Leder, H. (2016). Berlyne revisited: Evidence for
1169 the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers*
1170 *in human neuroscience*, 10, 536.
- 1171
- 1172 McWhinnie, H. J. (1971). A review of selected aspects of empirical aesthetics III. *Journal of*
1173 *Aesthetic Education*, 5(4), 115-126.
- 1174

- 1175
1176 Messinger, S. M. (1998). Pleasure and complexity: Berlyne revisited. *The Journal of*
1177 *Psychology*, 132(5), 558-560.
- 1178
1179 Moles, A. A. (1958). Theorie de l'information et perception esthetique.
- 1180
1181 Munsinger, H., & Kessen, W. (1964). Uncertainty, structure, and preference. *Psychological*
1182 *Monographs: General and Applied*, 78(9), 1.
- 1183
1184 Nadal, M., & Vartanian, O. (2021). Empirical Aesthetics: An overview.
- 1185
1186 Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty
1187 appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173-
1188 191.
- 1189
1190 Nasar, J. L. (2002). What design for a presidential library? Complexity, typicality, order, and
1191 historical significance. *Empirical Studies of the Arts*, 20(1), 83-99.
- 1192
1193 Nicki, R. M. (1972). Arousal increment and degree of complexity as incentive. *British Journal*
1194 *of Psychology*, 63(2), 165-171.
- 1195
1196 Nicki, R. M., & Moss, V. (1975). Preference for non-representational art as a function of
1197 various measures of complexity. *Canadian Journal of Psychology/Revue 36hannon36e de*
1198 *psychologie*, 29(3), 237.
- 1199
1200 Nicki, R. M., & Gale, A. (1977). EEG, measures of complexity, and preference for
1201 nonrepresentational works of art. *Perception*, 6(3), 281-286.
- 1202
1203 Nicki, R. M., Lee, P. L., & Moss, V. (1981). Ambiguity, cubist works of art, and preference.
1204 *Acta Psychologica*, 49(1), 27-41.
- 1205
1206 Norman, J. F., Beers, A., & Phillips, F. (2010). Fechner's Aesthetics Revisited. *Seeing and*
1207 *Perceiving*, 23(3), 263-271.
- 1208
1209 Osborne, J. W., & Farley, F. H. (1970). The relationship between aesthetic preference and
1210 visual complexity in 36hannon36 art. *Psychonomic Science*, 19(2), 69-70.
- 1211
1212 Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human
1213 preference. *Annual review of psychology*, 64, 77-107.
- 1214
1215 Redies, C., Brachmann, A., & Wagemans, J. (2017). High entropy of edge orientations
1216 characterizes visual artworks from diverse cultural backgrounds. *Vision Research*, 133, 130-
1217 144.
- 1218
1219 Rigau, J., Feixas, M., & Sbert, M. (2008). Informational aesthetics measures. *IEEE computer*
1220 *graphics and applications*, 28(2), 24-34.
- 1221
1222 Roberts, M. N. (2007). Complexity and aesthetic preference for diverse visual stimuli.
1223 *Doctoral*, Universitat de les Illes Balears, Palma, Spain.

- 1224
1225 Schmidhuber, J. (2009). Art & science as by-products of the search for novel patterns, or data
1226 compressible in unknown yet learnable ways. *Multiple ways to design research. Research*
1227 *cases that reshape the design discipline*, Swiss Design Network-Et al. Edizioni, 98-112.
1228
1229 Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical*
1230 *journal*, 27(3), 379-423.
1231
1232 Silva, J. M., Pratas, D., Antunes, R., Matos, S., & Pinho, A. J. (2021). Automatic analysis of
1233 artistic paintings using information-based measures. *Pattern Recognition*, 114, 107864.
1234
1235 Singh, S., & Shukla, D. (2017). A review on various measures for finding image complexity.
1236 *International Journal of Scientific Research Engineering & Technology*. ISSN, 2278-0882.
1237
1238 Snodgrass, J. G. (1971). Objective and subjective complexity measures for a new population
1239 of patterns. *Perception & Psychophysics*, 10(4), 217-224.
1240
1241 Stamps III, A. E. (2002). Entropy, visual diversity, and preference. *The Journal of general*
1242 *psychology*, 129(3), 300-320.
1243
1244 Sun, Z., & Firestone, C. (2022). Beautiful on the inside: Aesthetic preferences and the skeletal
1245 complexity of shapes. *Perception*, 1, 15.
1246
1247 Taylor, R. E., & Eisenman, R. (1964). Perception and production of complexity by creative art
1248 students. *The Journal of Psychology*, 57(1), 239-242.
1249
1250 Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package
1251 for causal mediation analysis.
1252
1253 Tinio, P. P., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry,
1254 complexity, and the jaws of massive familiarization. *Acta psychologica*, 130(3), 241-250.
1255
1256 Tran, N. H., Waring, T., Atmaca, S., & Beheim, B. A. (2021). Entropy trade-offs in artistic
1257 design: A case study of Tamil kolam. *Evolutionary Human Sciences*, 3.
1258
1259 Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation.
1260 *Psychology of aesthetics, creativity, and the arts*, 14(2), 135.
1261
1262 Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly
1263 organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, 15(3), 484.
1264
1265 Van Geert, E., Bossens, C., & Wagemans, J. (2022). The Order & Complexity Toolbox for
1266 Aesthetics (OCTA): A systematic approach to study the relations between order, complexity,
1267 and aesthetic appreciation. *Behavior Research Methods*, 1-24.
1268
1269 Van Lier, R., Van Der Helm, P., & Leeuwenberg, E. (1994). Integrating global and local
1270 aspects of visual occlusion. *Perception*, 23(8), 883-903.
1271
1272 Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral science*,
1273 11(2), 105-114.

- 1274
1275 Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis
1276 of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(3), 392-
1277 399.
1278
1279 William, A. Huber (2011). Measuring entropy/ information/ patterns of a 2d binary matrix.
1280
1281 Wolfram, S. (2002). A new kind of science, Wolfram Media, Inc. *Champaign, IL*.
1282
1283 Zenil, H., Hernández-Orozco, S., Kiani, N. A., Soler-Toscano, F., Rueda-Toicen, A., & Tegnér,
1284 J. (2018). A decomposition method for global evaluation of 38hannon entropy and local
1285 estimations of algorithmic complexity. *Entropy*, 20(8), 605.
1286
1287 Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding.
1288 *IEEE transactions on Information Theory*, 24(5), 530-536.

1289 Appendix I – Cellular Automata Generation

1290
1291 The number of potential state-transition update functions for a 2D CA with n states and an N-cell neighbourhood is n^{n^N} . Hence, for our 2D binary CA with 5-cell and 9-cell neighbourhoods,
1292 the number of possible state-transition update functions are $2^{2^5} = 2^{32} = 4 \times 10^9$ and $2^{2^9} =$
1293 $2^{512} = 10^{152}$ respectively. Moreover, the total number of initial configurations for an n-state
1294 P × Q grid is $n^{P \times Q}$. In our binary 15×15 grid, this would be $2^{15 \times 15} = 2^{225}$ possible initial
1295 configurations. Since these are very large spaces, we consider simplified versions of rules
1296 (Wolfram, 1983). The rules are set such that the state of a cell depends only on the sum of the
1297 states of cells in its neighbourhood (SCN). Such rules can be of two types: “totalistic” (tot)
1298 where the state of the cell (i, j) at time t+1 depends only on SCN at time t, or “outer-totalistic”
1299 (Otot) where the state of the cell (i, j) at time t+1 depends on both SCN and the value of cell (i,
1300 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
1301 totalistic rules and $\sum_{i=0}^N \sum_{j=0}^1 (f(i, j) \times 2^{2i+j})$ for outer-totalistic rules.
1302

1303
1304 It is a well-known fact that unlike 1D/elementary CA, the limiting behaviour of 2D CA is
1305 undecidable. Patterns might terminate, oscillate, tend towards full randomness or tend towards
1306 order depending on the combination of algorithm parameters. It is also not possible to reverse-
1307 engineer rules that result in particular types of pattern outputs. Therefore, we referred to work
1308 that has elucidated some rules with their evolving behaviours (Packard and Wolfram, 1985;
1309 Wolfram, 2002) while selecting our set of rules and fixing the number of iterations based on
1310 our grid size to T = 40.

1311
1312 The generation script was coded in Python v3.8.8. For each cell, either the 5-cell or a 9-cell
1313 neighbourhood was determined and the sum of the SCN was computed. The grid was wrapped
1314 around such that the cells on the boundary considered cells on the opposite boundary as
1315 neighbours. The updated state of the cell was obtained as a function of the computed sum as
1316 indicated by the rule code read in binary. The pseudocode is given here.
1317

1318 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 5th iteration)

- 1 *Bin_rule_code* ← binary form of decimal rule code
- 2 *Powers* ← list of indices where *bin_rule_code* = 1 when read in reverse
- 3 **For** timestep *t* = 1 to *T*
- 4 *G'* ← 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

```

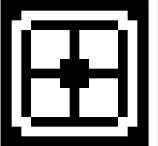
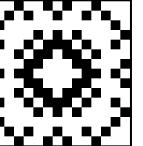
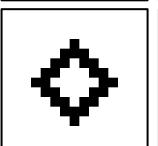
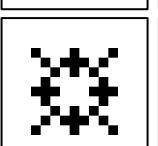
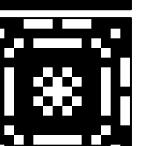
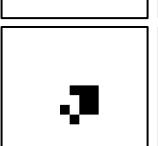
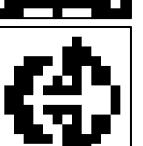
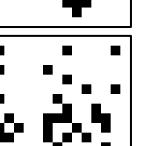
1319

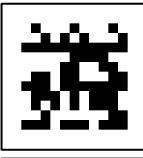
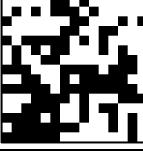
1320 Some example rules used by us along with the produced patterns are listed in Figure AI.1.

1321

1322 **Figure AI.1.**

1323 *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	 

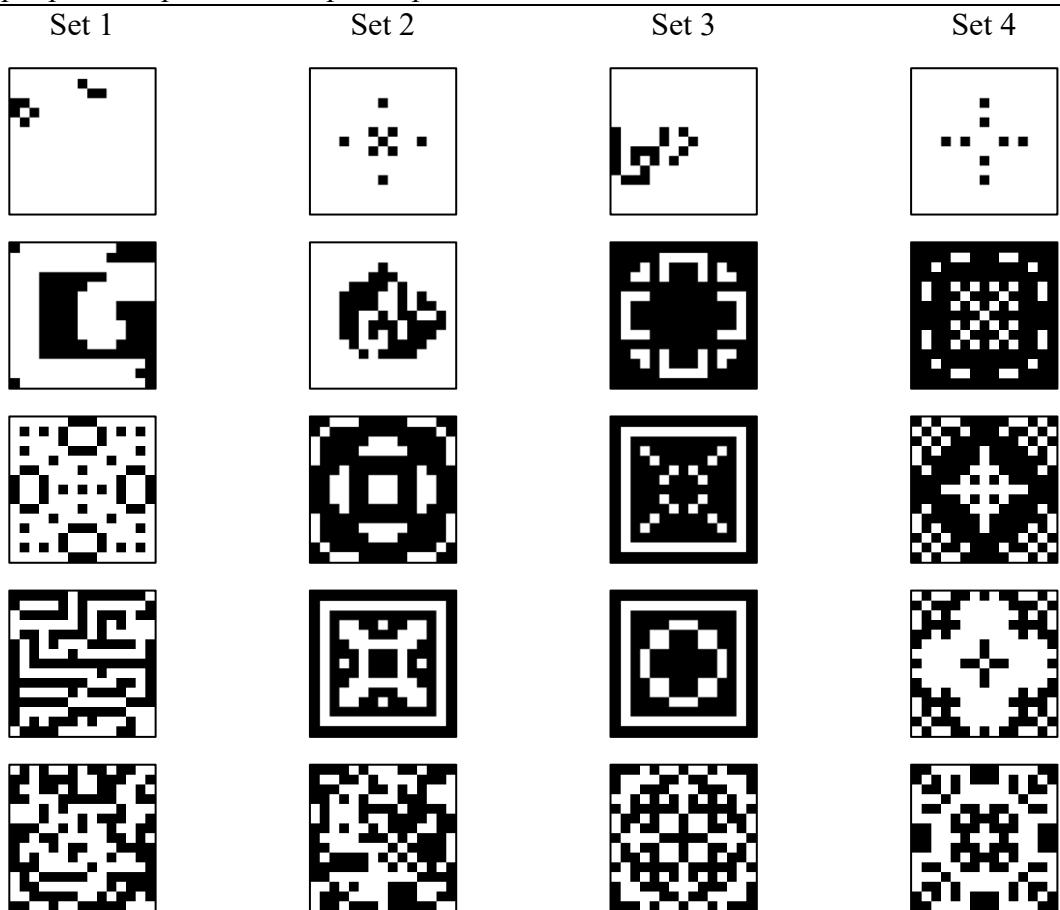
6	15822	9	Otot	2		
7	736	9	Otot	3		
8	196623	9	Otot	3		
9	52	5	Tot	3		

1324

1325 Figure AI.2 displays example patterns presented to participants.

Figure AI.2.

1327 *Example patterns presented to participants in the 4 sets*



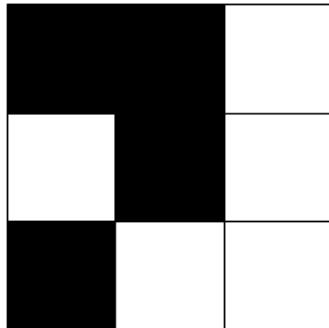
1328

1329 Appendix II – Objective Complexity Measures

1330
1331 Here we provide a detailed description of the measures defined in Section 2.2 along with
1332 illustrated examples using a simple pattern (Figure AII.1).

1333 1334 **Figure AII.1**

1335 *Example pattern to show measure computations.*



1336
1337 Note: The pattern shown is same as the one used to show intricacy computation in Figure 3.
1338

1339 **Density**

1340
1341 The density of the pattern is the proportion of black pixels.

1342 Here this would be,

1343 1344 **Figure AII.2**

1345 *Density computation*

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

1346

1347 **Entropies**

1348

- 1349 1. Entropy: Entropy is calculated at a single scale as given by Eq. 1. While $P(b)$ is simply
1350 the density from above, $P(w)$ is $(1 - \text{density})$ as $P(b) + P(w) = 1$ for our patterns.

1351

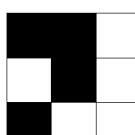
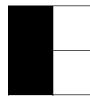
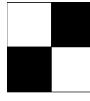
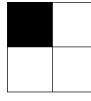
1352 Therefore, here, entropy is:

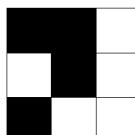
$$1353 -0.44 \log_2 0.44 - (1 - 0.44) \log_2 (1 - 0.44) = 0.52 + 0.46 = \mathbf{0.98}$$

1354

- 1355 2. Mean entropy: Mean entropy averages entropy calculated at different scale for different
1356 windows as illustrated below.

1357

1358 Number of scales = Size of the grid = 3
 1359
 1360 **Figure AII.3**
 1361 *Mean entropy computation*
 1362
 1363 3×3 windows:

 1364 Entropy = 0.98
 1365 Mean entropy at this scale = 0.98
 1366
 1367
 1368 2×2 windows:
 Entropy = 0.81  Entropy = 1  Entropy = 1  Entropy = 0.81
 1369 Mean entropy at this scale = 0.91
 1370
 1371 1×1 windows:
 1372

 1373 Entropy for all = 0
 1374 Mean entropy at this scale = 0
 1375
 1376
 1377 Mean entropy = $\frac{0.98+0.91+0}{3} = 0.63$
 1378
 1379 3. Entropy of means: Entropy of means is the mean entropy of a pattern across different
 1380 levels of smoothing (sliding window averages).
 1381
 1382 Number of scales = Size of the grid = 3
 1383
 1384 **Figure AII.4**
 1385 *Entropy of means computation*
 1386
 1387 3×3 windows:

 1388 Mean = 0.44
 1389 Entropy of means at this scale = 0
 1390
 1391

1392

2 × 2 windows:



Mean = 0.75



Mean = 0.5



Mean = 0.5



Mean = 0.25

1393

$$\text{Entropy of means at this scale} = -2 \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{2} \log_2 \frac{1}{2} = 0.811$$

1394

1395

1 × 1 windows:

Mean
= 1Mean
= 1Mean
= 0Mean
= 0Mean
= 1Mean
= 0Mean
= 1Mean
= 0Mean
= 0

1396

$$\text{Entropy of means at this scale} = -\frac{4}{9} \log_2 \frac{4}{9} - \frac{5}{9} \log_2 \frac{5}{9} = 0.52 + 0.47 = 0.99$$

1397

1398

$$\text{Mean entropy of means} = \frac{0+0.81+0.99}{3} = \mathbf{0.6}$$

1399

1400

1401 Local spatial complexity

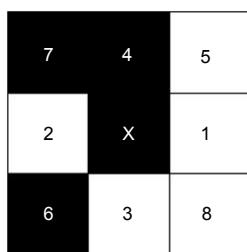
1402

1403 The LSC computations are shown in Table AII.1 for the Figure AII.1 as labelled in AII.5.

1404

1405 Figure AII.5

1406 Example pattern with directions labelled for LSC computation.



1407

1408

1409 Table AII.1

1410 LSC computations for the example pattern

1411

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 = 0.99
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 = 0.91
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		-
	$s_1 = 1, s_2 = 0$ (black, white)	0.33	0.66		0.19
	$s_1 = 1, s_2 = 1$ (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	-
	$s_1 = 1, s_2 = 1$ (black, black)	0.16	0.33	0.26
	Sum = 0.45			
	Total sum/number of directions = 0.73			

1412

1413 **Asymmetry**

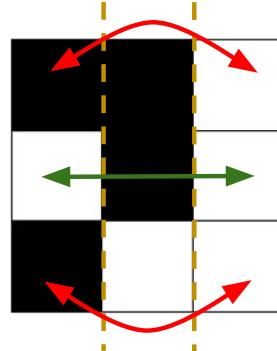
1414

1415 Horizontal asymmetry measures asymmetry about the horizontal axis. Mathematically, for
 1416 each cell (i, j) where i represents the cell row and j represents the cell column and assuming 0-
 1417 indexing, horizontal symmetry compares the percentage of mismatches between (i, j) and $(i, N-j)$ for j in $[0, N/2]$. On the other hand, vertical asymmetry measures asymmetry about the
 1418 vertical axis and compares the percentage of mismatches between (I, j) and $(N-I, j)$ for I in $[0,$
 1419 $N/2)$. Figure AII.6 illustrates the computation.

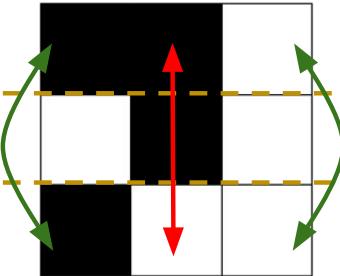
1420

1421 **Figure AII.6**1422 *Horizontal (A) and vertical (B) asymmetry computation*

1423



(A) Horizontal asymmetry



(B) Vertical asymmetry

1425

1426

1427 In addition to the complexity measure mentioned in Section 2.2, we implemented two other
 1428 computational complexity measures: an 8-neighbourhood version of our intricacy measure and
 1429 a hierarchical quad-tree measure. The motivation for and computation of these measures are
 1430 presented here.

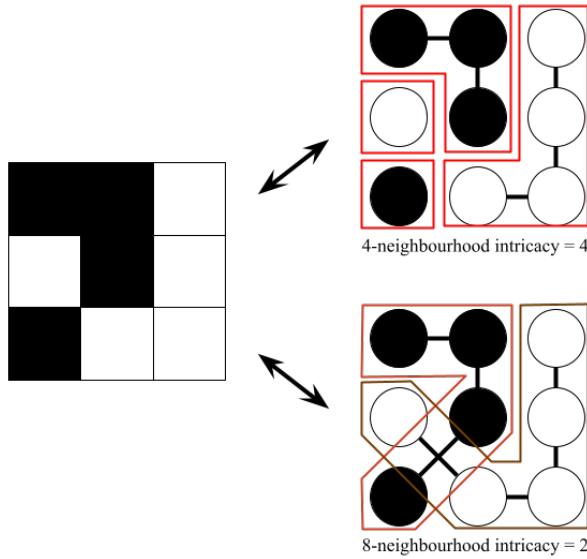
1431

1432 1. **8-neighbourhood intricacy measure:** In Figure 8, we presented some of the patterns
 1433 for which our complexity measure is most erroneous. Complexity is overestimated in
 1434 some patterns with high evaluated intricacy – potentially because diagonal relationships
 1435 do not contribute to connectedness. To tackle this, we implemented an 8-
 1436 neighbourhood version of intricacy wherein the graphs considers all 8 neighbours. An
 1437 edge is inserted between two neighbours if they are the same colour. This measure
 1438 would in turn result in a lower value of intricacy compared to the 4-neighbourhood
 1439 version. Figure AII.7 illustrates its computation of the 4-neighbourhood vs. the 8-
 1440 neighbourhood intricacy.

1441

1442 **Figure AII.7**

1443 *4-neighbourhood and 8-neighbourhood intricacy computation for an example pattern.
 1444 The graphs on right are constructed from the pattern on left. Boxes indicate connected
 1445 components. Here, 4-neighbourhood intricacy = 4 and 8-neighbourhood intricacy = 2.*
 1446



1447
1448
1449

1450 **2. Quadtree:** Since most of our measures (except entropy) are either local or global, we
1451 programmed a hierarchical measure of complexity, namely quadtrees. Quadtrees are
1452 hierarchical data structures and are commonly used for representing images (Finkel and
1453 Bentley, 1974). A quadtree is based on the principle of recursive decomposition where
1454 the pattern is repeatedly divided into four sub-patterns. The condition for subdivision
1455 is based on the number of distinct states in the graph. If all cells are of the same colour,
1456 the algorithm stops, if not, the graph is subdivided into four sub-graphs and the same
1457 rule is recursively applied to each sub-graph. The number of times the graph undergoes
1458 subdivisions is evaluated as a measure of complexity. This measure however is not
1459 ideal for our pattern with grid size 15×15 where the width and height are not powers
1460 of 4 and as a result the four sub-patterns are of unequal dimensions. Moreover, the
1461 measure was found to be highly correlated with LSC ($r = 0.88$, $p < 0.01$, $CI = [0.87,$
1462 $0.88]$).

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*

Pattern	Density	Entropy	Mean entropy	Entropy of means	LSC	KC	Asymmetry (horizontal, vertical, local)	Intricacy, 4-neigh	Intricacy, 8-neigh	Quadtree N	Rule type	IC	λ	
	1.00	0.0	0.0	0.0	0.0	25.17	0, 0, 0	1	1	0	5	Otot	1	9
	0.00 4	0.04	0.05	0.25	0.04	48.35	0, 0, 0	2	2	4	-	-	1	-
	0.49	0.99	0.93	0.5	0.0	33.43	0, 0, 0	255	2	84	5	Otot	1	5
	0.10	0.48	0.55	1.97	0.43	134.2	58.3, 0, 0	3	3	22	9	Otot	2	4

	0.74	0.81	0.80	2.16	0.81	257.8	0, 0, 0	16	16	54	5	Otot	1	8
	0.45	0.99	0.89	2.76	0.96	278.7	53.3, 46.6, 0.004	30	8	65	9	Otot	3	6

1470 *Note.* The generation parameters (N , rule type, IC, along with rule number (hence also λ) 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.

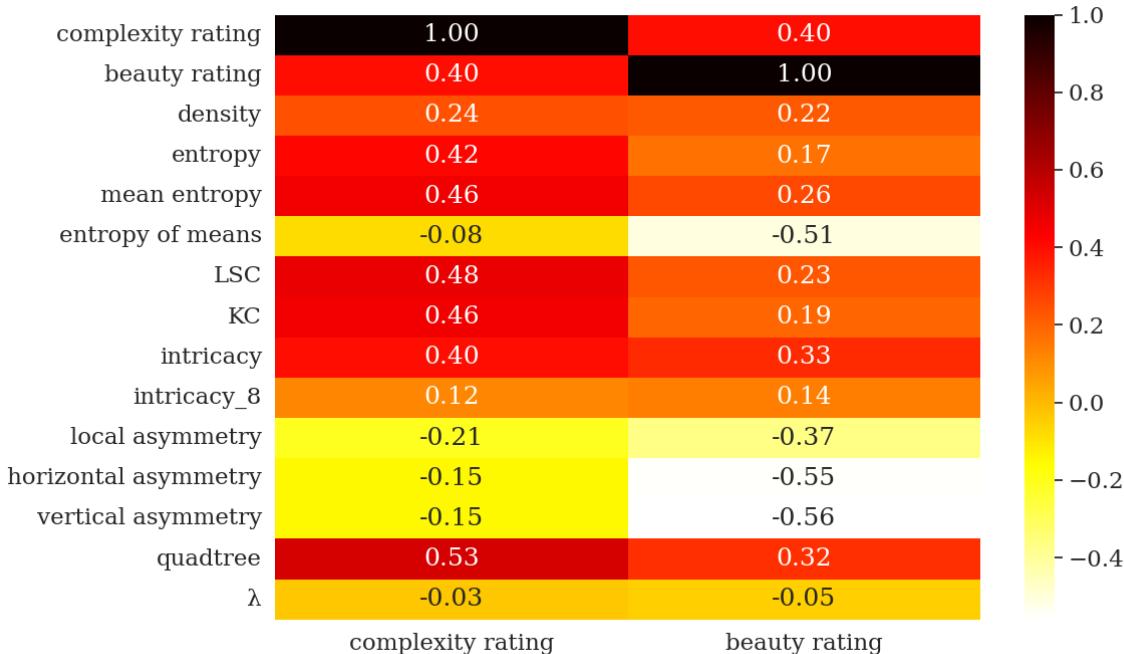
1474 **Appendix III – Supplementary Analysis**

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 , $p < 0.01$, $CI = [0.5, 0.55]$) and 4-nieghbourhood intricacy ($r = 0.40$, $p < 0.01$, $CI = [0.38,$
1481 $0.43]$) measures. The ratings were not highly correlated with asymmetry or entropy. On the
1482 other hand, the beauty ratings were highly negatively correlated with the three asymmetry
1483 measures ($r = -0.37$, $p < 0.01$, $CI = [-0.39, -0.34]$; $r = -0.55$, $p < 0.01$, $CI = [-0.57, -0.52]$; $r = -$
1484 0.56 , $p < 0.01$, $CI = [-0.58, -0.54]$) and entropy ($r = -0.51$, $p < 0.01$, $CI = [-0.53, -0.49]$) and
1485 positively with complexity ratings ($r = 0.4$, $p < 0.01$, $CI = [0.37, 0.42]$).
1486

1487 **Figure AIII.1**

1488 *Correlation between ratings and computational measures*



1489
1490
1491 **2. Fits of models of complexity ratings.**

1492
1493 Here we report the full list of complexity models we experimented with along with variances
1494 in the performance measures.
1495

1496 **Table AIII.1**

1497 *(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 ~ mean_entropy + (1 | Participant)
 6 CR ~ entropy_of_means + (1 | Participant)
 7 CR ~ LSC + (1 | Participant)
 8 CR ~ KC + (1 | Participant)
 9 CR ~ asymm + (1 | Participant)
 10 CR ~ intricacy_4 + (1 | Participant)
 11 CR ~ intricacy_8 + (1 | Participant)
 12 CR ~ quadtree + (1 | Participant)
 13 CR ~ LSC + density + (1 | Participant)
 14 CR ~ LSC + entropy + (1 | Participant)
 15 CR ~ LSC + mean_entropy + (1 | Participant)
 16 CR ~ LSC + entropy_of_means + (1 | Participant)
 17 CR ~ LSC + asymm + (1 | Participant)
 18 CR ~ LSC + intricacy_4 + (1 | Participant)
 19 CR ~ LSC + intricacy_8 + (1 | Participant)
 20 CR ~ LSC + quadtree + (1 | Participant)
 21 CR ~ LSC + intricacy_4 + intricacy_8 + (1 | Participant)
 22 CR ~ LSC + entropy_of_means + density + (1 | Participant)
 23 CR ~ LSC + intricacy_4 + LSC:intricacy_4 + (1 | Participant)
 24 CR ~ LSC + intricacy_4 + (LSC | Participant)
 25 CR ~ LSC + intricacy_4 + (intricacy_4 | Participant)
26 CR ~ LSC + intricacy_4 + ((LSC + intricacy_4) | Participant)
 27 CR ~ LSC * intricacy_4 + ((LSC + intricacy_4) | Participant)
 28 CR ~ quadtree + intricacy_4 + (1 | Participant)
 29 CR ~ quadtree + intricacy_4 + (quadtree | Participant)
 30 CR ~ quadtree + intricacy_4 + ((intricacy_4 + intricacy_4) | Participant)
 31 CR ~ LSCsq + intricacy_4sq + (1 | Participant)
 32 CR ~ LSC + LSCsq + intricacy_4 + intricacy_4sq + (1 | Participant)
 33 CR ~ neighbourhood_size + tot_outertot + IC + (1 | Participant)
 34 CR ~ trial + (1 | Participant)
 35 CR ~ LSC + intricacy_4 + trial + ((LSC + intricacy_4) | Participant)
 36 CR ~ previous_CR + (trial | Participant)
 37 CR ~ LSC + intricacy_4 + previous_CR + ((LSC + intricacy_4) | Participant)

1498 Note. CR=complexity ratings, LSC=local spatial complexity, KC=Kolmogorov complexity; *
 1499 indicates p < 0.05. Bold indicates best model.

1500

1501 *(B) Performance of models specified in Table AIII.1 based on AIC, BIC, R² and RMSE metrics.*

Id	AIC	BIC	AIC/BIC Var	R ²				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	8532.9	8557.1	2287.7	0.17	1.8e-07	0.14	2.9e-06	0.91	4.9e-05	0.93	2e-04
7	7552	7576.2	1151.5	0.4	0.00015	0.37	0.00053	0.78	2.3e-05	0.79	7.5e-05
8	7703.8	7728	524.2	0.37	8.6e-05	0.34	3e-04	0.8	1.1e-05	0.81	3.3e-05
9	8399.9	8424.1	1525.3	0.2	5.9e-06	0.17	2.3e-05	0.89	3.1e-05	0.91	0.00013
10	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
11	8493.1	8517.3	2803.9	0.18	8.5e-06	0.15	2.7e-05	0.91	6e-05	0.92	0.00024
12	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
13	7551.3	7581.5	1422.4	0.4	0.00017	0.37	0.00062	0.78	2.8e-05	0.79	9.2e-05
14	7536.6	7566.8	1387.7	0.4	0.00018	0.38	0.00066	0.77	2.8e-05	0.79	9.1e-05
15	7550.7	7581	872.7	0.4	0.00012	0.37	0.00045	0.78	1.8e-05	0.79	5.4e-05
16	7443.8	7474.1	690.5	0.42	0.00013	0.4	0.00045	0.76	1.4e-05	0.78	4.5e-05
17	7525.7	7555.9	979.4	0.4	0.00014	0.38	0.00049	0.77	1.9e-05	0.79	6.6e-05
18	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
19	7556.9	7587.1	1221.1	0.4	0.00014	0.37	0.00051	0.78	2.4e-05	0.79	8e-05
20	7314.8	7345	323.4	0.44	6.7e-05	0.42	2e-04	0.75	6.4e-06	0.76	2e-05
21	7291.4	7327.6	197	0.45	7.1e-05	0.43	0.00021	0.74	4.1e-06	0.76	1.3e-05
22	7449.9	7486.2	801.6	0.42	0.00014	0.4	0.00048	0.76	1.6e-05	0.78	5e-05
23	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
24	7219.9	7262.2	541.1	0.48	0.00012	0.45	0.00036	0.72	1.2e-05	0.75	3.2e-05

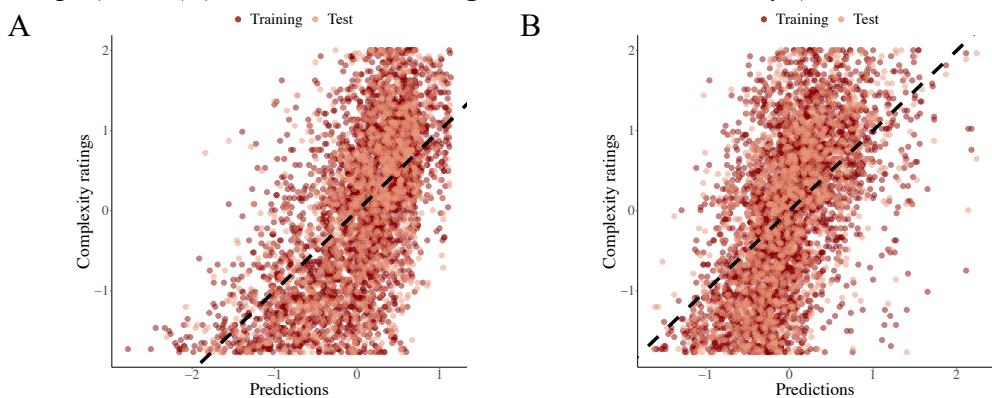
25	7166.3	7208.6	660.8	0.5	0.00011	0.46	0.00025	0.71	1.8e-05	0.74	3.4e-05
26	7138.9	7199.4	822.7	0.52	0.00015	0.47	0.00035	0.70	2.1e-05	0.73	4.2e-05
27	7133.6	7200.1	798.1	0.52	0.00015	0.47	0.00036	0.70	1.8e-05	0.73	4.2e-05
28	7311.8	7342.1	264.4	0.44	4.5e-05	0.42	0.00012	0.75	4e-06	0.76	1.8e-05
29	7194.4	7236.8	179.4	0.49	5.7e-05	0.45	0.00015	0.71	1.9e-06	0.74	3.4e-06
30	7191.2	7233.5	354	0.49	7.5e-05	0.45	0.00014	0.71	8.2e-06	0.74	2.2e-05
31	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
32	7256.7	7299	233.3	0.46	7.1e-05	0.43	2e-04	0.74	4.8e-06	0.75	1.3e-05
33	8457.2	8499.5	1845.1	0.19	9e-06	0.16	4.3e-05	0.9	3.9e-05	0.92	0.00017
34	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
35	7141.6	7208.1	909.8	0.52	0.00014	0.47	0.00032	0.7	2e-05	0.73	4.8e-05
36	8520.4	8556.7	2617.7	0.18	1.5e-05	0.14	6.6e-06	0.91	5.3e-05	0.93	0.00025
37	7109.5	7176	491.6	0.52	0.00014	0.47	0.00027	0.69	1.7e-05	0.73	2.7e-05

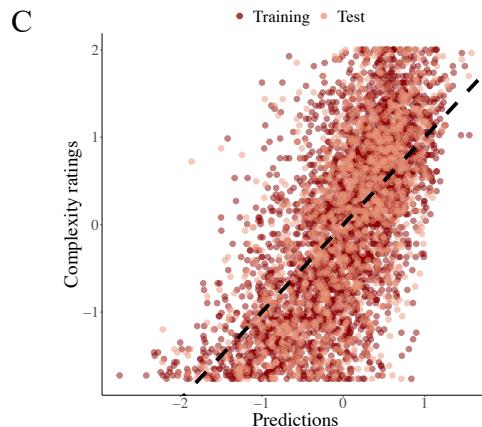
1502

1503

Figure AIII.2

1504 *Model fit plots (A) only-LSC model (without random slopes), (B) only intricacy model (without
1505 random slopes) and (C) the model including both LSC and intricacy (without random slopes)*





1506 3. Fits of models of beauty ratings.

1507

1508 Here we report beauty models in the main text along with variances along with the full list of
 1509 beauty models we experimented with.

1510

1511 **Table AIII.2**

1512 *(A) Models of beauty ratings*

Id	Model
1	$BR \sim CR + (1 Participant)$
2	$BR \sim CR + asymm + entropy + (1 Participant)$
3	$BR \sim CR + asymm + entropy + (CR Participant)$
4	$BR \sim CR + asymm + entropy + (asymm Participant)$
5	$BR \sim CR + asymm + entropy + (entropy Participant)$
6	$BR \sim CR + disorder + (1 Participant)$
7	$BR \sim CR + disorder + (disorder Participant)$
8	$BR \sim CR * disorder + (disorder Participant)$
9	$BR \sim CR * disorder_1 + (disorder_1 Participant)$
10	$BR \sim CR * disorder_2 + (disorder_2 Participant)$
11	$BR \sim CR_sq + disorder + (1 Participant)$
12	$BR \sim CR + CR_sq + disorder + (1 Participant)$
13	$BR \sim LSC + intricacy_4 + asymm + entropy + (1 Participant)$
14	$BR \sim LSC + intricacy_4 + disorder + (1 Participant)$
15	$BR \sim LSC + intricacy_4 + disorder + (disorder Participant)$
16	$BR \sim LSC + intricacy_4 + disorder + (LSC + intricacy_4):disorder + (disorder Participant)$
17	$BR \sim trial + (1 Participant)$
18	$BR \sim CR * disorder + trial + (disorder Participant)$
19	$BR \sim previous_BR + (trial Participant)$
20	$BR \sim CR * disorder + previous_BR + (disorder Participant)$
21	$BR \sim CR * trial + (1 Participant)$

1513 Note. BR=beauty ratings, CR=complexity ratings, LSC=local spatial complexity; * indicates p
 1514 < 0.05. Bold indicates best model.

1515

1516 *(B) Performance of models specified in Table AIII.2 based on AIC, BIC, R² and RMSE metrics.*

1517

Id	AIC	BIC	AIC, BIC Var	R ²				RMSE			
				Train		Test		Train		Test	
				Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	8046.4	8070.6	3052.4	0.29	0.000	0.25	0.00	0.85	7.1e-05	0.87	0.00
2	6651.4	6687.6	6156.6	0.55	0.000	0.53	0.00	0.67	7.5e-05	0.69	3e-04
3	6358.1	6406.4	1344.9	0.62	2.8e-05	0.58	0.00	0.62	8e-06	0.65	0.00
4	6261.7	6310.1	6742.3	0.64	0.000	0.58	7e-04	0.6	7e-05	0.65	0.00
5	6623.7	6672	4951.1	0.57	0.000	0.53	0.00	0.65	4.6e-05	0.69	0.00
6	6449.6	6479.8	5423.6	0.58	0.000	0.55	0.00	0.65	6.3e-05	0.67	0.00
7	5985.0	6027.3	6317.1	0.67	0.000	0.63	0.00	0.58	6e-05	0.61	0.00
8	5919.1	5967.4	6164.4	0.68	9.8e-05	0.64	0.00	0.57	5.5e-05	0.60	0.00
9	6083.0	6131.4	9223.2	0.66	0.000	0.61	0.00	0.58	9e-05	0.62	0.00
10	6099.8	6148.2	8974.6	0.66	0.000	0.61	0.00	0.59	8.8e-05	0.62	0.00
11	6872.3	6902.6	1543.3	0.52	5.4e-05	0.49	0.00	0.69	2.1e-05	0.71	8.8e-05
12	6457.9	6494.2	5480.4	0.58	0.000	0.55	0.00	0.65	6.3e-05	0.67	0.00
13	6852.0	6894.3	1332.4	0.53	3.1e-05	0.50	0.00	0.69	1.8e-05	0.71	7.1e-05
14	6685.8	6722	1414.5	0.55	3.4e-05	0.52	0.00	0.67	1.9e-05	0.69	7e-05
15	6185	6233.4	2373.3	0.65	5.3e-05	0.61	0.00	0.59	2.5e-05	0.63	9.1e-05
16	6088.1	6148.6	1446.1	0.66	3.2e-05	0.62	0.00	0.58	1.6e-05	0.62	5.7e-05
17	8544.7	8568.9	751.7	0.17	0.000	0.12	0.00	0.91	2.4e-05	0.94	5.8e-05
18	5926.7	5981.1	6272.3	0.68	9.9e-05	0.64	0.00	0.57	5.6e-05	0.6	0.00
19	8526.8	8563.1	1450	0.17	9.5e-05	0.13	5e-04	0.91	2e-05	0.94	0.00
20	5903.3	5957.7	6532.6	0.68	9.9e-05	0.64	5e-04	0.57	5.7e-05	0.60	0.00
21	8060.9	8097.2	3075.9	0.29	0.000	0.25	0.00	0.85	7.1e-05	0.87	0.00
				25		073		05		019	

1518

1519 4. Mixed Effects Regressions for predicting Complexity Ratings with 8-neighbourhood
 1520 Intricacy and Quadtree (refer to AII for a description of these measures)

1521
 1522 Table AIII.3 summarizes models of complexity ratings involving 8-neighbourhood intricacy
 1523 and quadtree. Comparing with Table 1, we see that although quadtree achieves good
 1524 performance as a predictor alone, there are negligible gains when we introduce intricacy into
 1525 the expression. Further, the slight increase in BIC when doing so indicates that the model nearly
 1526 overfits the data. Further, the 8-neighbourhood intricacy is unable to explain much variance in
 1527 the ratings compared to the 4-neighbourhood version implying that the 4-neighbourhood
 1528 intricacy is a more superior predictor of complexity ratings. Therefore, the model we propose
 1529 in Section 3.1 comprising of LSC and 4-neighbourhood intricacy along with random slopes of
 1530 LSC and intricacy, and random intercept of participant is our best predictor for complexity
 1531 ratings.

1532
 1533 Table AIII.3: Summary of models of complexity ratings

Id	Model	Significance	AIC	BIC	R ²		RMSE
					Train	Test	
1	CR ~ quadtree + quadtree* 1 Participant		7317.1	7341.3	0.41	0.74	0.76
2	CR ~ quadtree + quadtree* quadtree Participant		7199.9	7236.2	0.45	0.71	0.74
3	CR ~ intricacy_8 + intricacy_8* 1 Participant		8493.1	8517.2	0.14	0.90	0.92
4	CR ~ quadtree + quadtree* intricacy_4 + intricacy_4* 1 Participant		7311.8	7342.0	0.42	0.74	0.76
5	CR ~ quadtree + quadtree* intricacy_4 + intricacy_4* quadtree Participant		7194.4	7236.7	0.45	0.71	0.74
6	CR ~ quadtree + quadtree* intricacy_4 + intricacy_4* intricacy_4 Participant		7191.2	7233.5	0.45	0.71	0.74

1534
 1535 5. Test for trends, autocorrelation, and consistency of repeated measures in the ratings

1536
 1537 To test for trends and autocorrelation (at lag 1) in the data, trial number and previous rating
 1538 were respectively added as a predictor of complexity and beauty ratings. Table AIII.4 (A) and
 1539 (B) report the performance of these models.

1540
 1541 **Table AIII.4**

1542 (A) Summary of models of complexity ratings. CR = complexity ratings, LSC = local spatial
 1543 complexity, prevCR = previous complexity rating (from previous trial)

Id	Model	Significance	AIC	BIC	R ²	RMSE
----	-------	--------------	-----	-----	----------------	------

							Train	Test
1	CR ~ trial + 1 Participant			8550.4	8574.6	0.13	0.91	0.93
2	CR ~ LSC + intricacy_4	LSC*	+ trial + intricacy_4*	7168.4	7216.7	0.46	0.70	0.73
	intricacy_4 Participant							
3	CR ~ prevCR + prevCR*	1 Participant	8520.4	8556.6	0.13	0.90	0.92	
4	CR ~ LSC + intricacy_4	LSC*	+ prevCR + intricacy_4*	7137.3	7185.6	0.46	0.70	0.73
	intricacy_4 Participant	prevCR*						

1544

1545 (B) Summary of models of beauty ratings. BR = beauty ratings, OC = objective complexity

Id	Model	Significance	AIC	BIC	R ²	RMSE	
						Train	Test
1	BR ~ trial + 1 Participant		8550.4	8574.5	0.13	0.91	0.93
2	BR ~ CR + disorder + CR:disorder + trial + disorder Participant	*	5948.3	6002.7	0.65	0.57	0.59
3	BR ~ prevBR + prevBR*	1 Participant	8532.1	8568.4	0.13	0.91	0.93
4	BR ~ CR + disorder + CR:disorder + prevBR + disorder Participant	*	5925.4	5979.8	0.65	0.57	0.59
		CR:disorder					
		*					
		prevBR*					

1546

1547 For predicting complexity ratings, we find that trial number is not a significant predictor,
1548 indicating there are no significant trends in our data – as would be expected in case of boredom,
1549 or over-familiarity. We see that previous complexity rating is a significant predictor, however,
1550 it does not enhance performance in conjunction with LSC and intricacy. Similarly for
1551 predicting beauty ratings, trial is not significant and previous beauty rating, though significant,
1552 does not enhance performance beyond our best model reported in Section 3.2.

1553

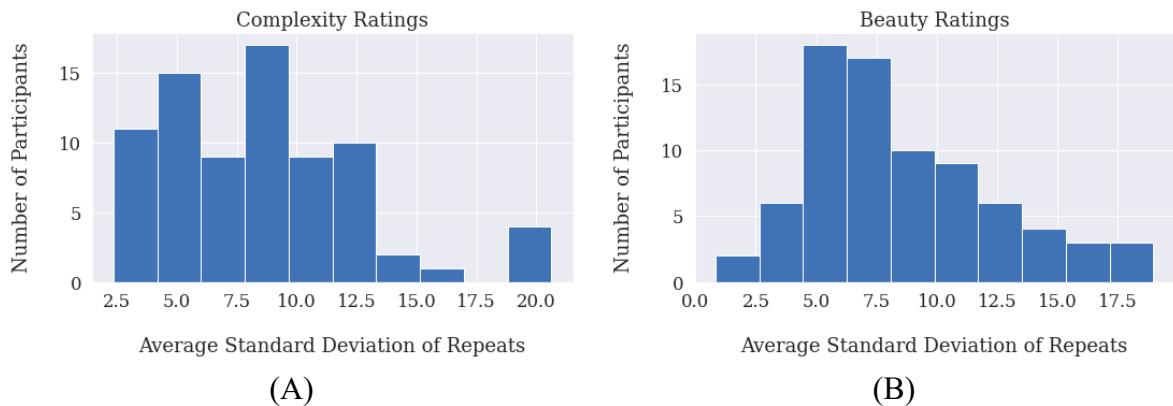
1554 Further, to test for consistency of repeated responses, we plot participant mean standard
1555 deviation in repeated measures (Figure AIII.3 (a) for complexity ratings and (b) for beauty
1556 ratings). The range of standard deviation values is from 0 (when two values are identical) to
1557 70.7 (when the two values are farthest apart on the scale, *i.e.*, one value is 0 and the other is
1558 100). Since most of the average standard deviations are within ~20% of the maximum standard
1559 deviation, we conclude that participants are largely consistent in their ratings.

1560

1561 **Figure AIII.3**

1562 *Histogram of participant mean standard deviation in repeated measures for (A) complexity*
1563 *ratings and (B) beauty ratings.*

1564



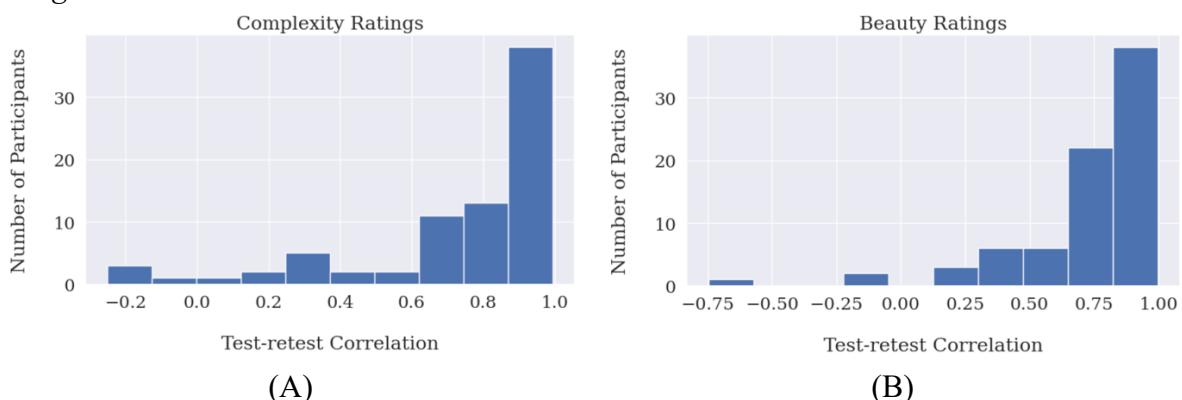
1565

1566 To test this further, we plot the distribution of test-retest correlations, *i.e.* correlations between
1567 repeats among participants for complexity and beauty ratings (Figure AIII.4). We see that most
1568 correlations are above 0.5 (with a mean 0.728 for complexity ratings and 0.725 for beauty
1569 ratings) which means that the majority of participants were largely consistent in their ratings.

1570

1571 **Figure AIII.4**

1572 *Histogram of participant test-retest correlations for (A) complexity ratings and (B) beauty*
1573 *ratings.*



1574

1575 6. Average complexity ratings vs average beauty ratings (colour coded by symmetry and
1576 entropy)

1577

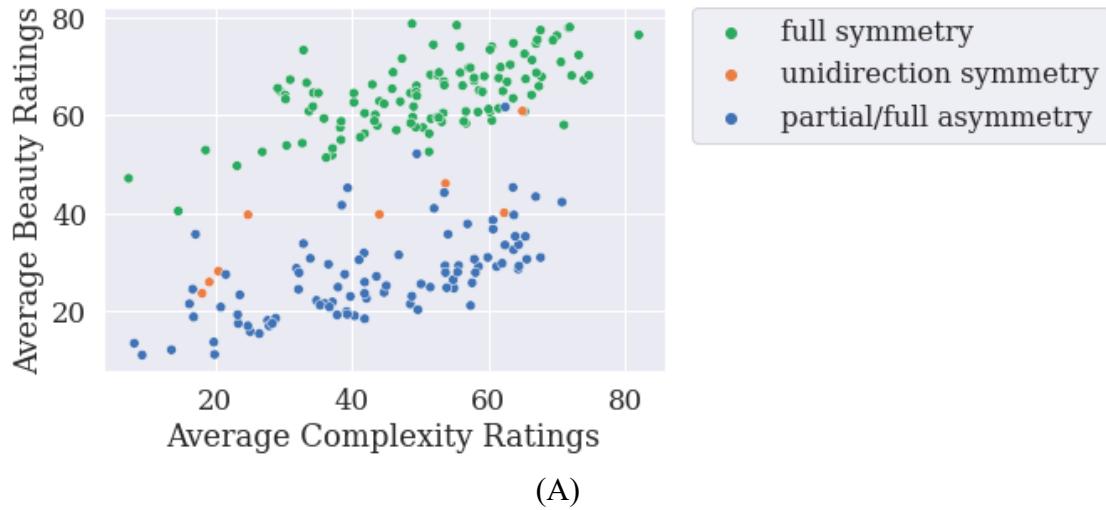
1578 A plot of average beauty ratings per pattern across all participants versus average complexity
1579 ratings per pattern across all participants neatly underlines the role of asymmetry and entropy
1580 in beauty assessment (Figure AIII.5a,b). In Figure AIII.5a, the linear relationship between
1581 beauty and complexity is evident but we see two modes in the distribution (a Gaussian Mixture
1582 Model with 2 components was able to fit the average ratings well, Appendix AIII.7). The
1583 degree of symmetry is successfully able to explain this bimodality with fully symmetric
1584 patterns being rated higher on average than unidirectional symmetric (semi-symmetric)
1585 patterns, which are themselves rated higher on average than fully asymmetric (non symmetric)

1586 patterns. Further, Figure AIII.5b indicates that patterns with high entropy were rated as low
1587 beauty whereas patterns with low entropy were rated as high beauty unless the pattern was
1588 rated very low complexity.

1589

1590 **Figure AIII.5**

1591 *Average complexity ratings vs average beauty ratings per pattern across all participants*
1592 *labelled according to (A) degree of symmetry in the pattern (as defined in Table 1), (B) level*
1593 *of entropy of the pattern (tertile split of entropy)*



1594

1595

1596

1597

1598

1599

1600 7. Gaussian Mixture Model to fit average complexity ratings versus average beauty ratings

1601

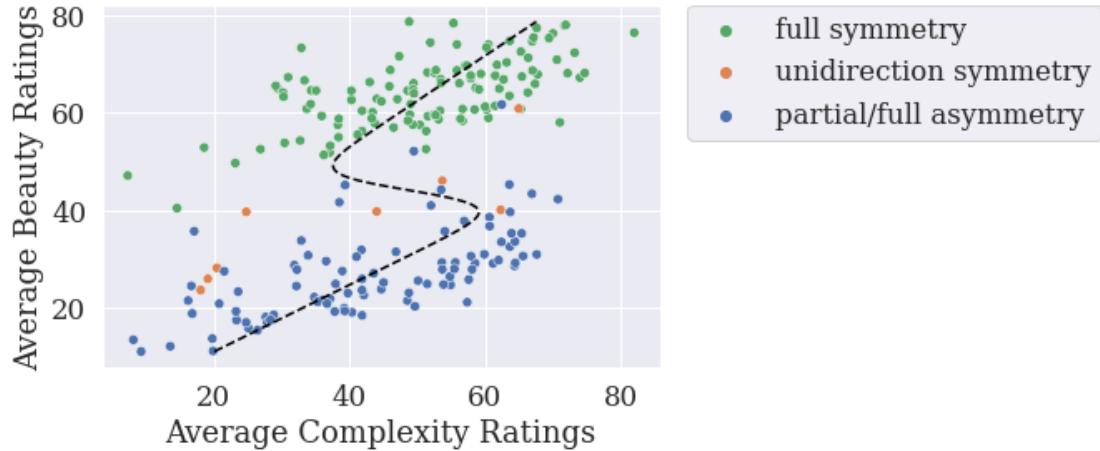
1602 As seen above, the plot of average complexity ratings versus average beauty ratings reflects a
1603 bimodal distribution. We therefore used the gmr package in Python to perform Gaussian
1604 Mixture Regression (GMR) with 2 modes on the average ratings. In GMR, first the joint
1605 distribution $p(x, y)$ is learnt, where in this case x refers to average beauty ratings and y refers to
1606 average complexity ratings. Then, the conditional $p(y | x)$ is computed to make predictions.
1607 GMR with 2 components is able to fit the data well (Figure AIII.6) with an RMSE of 12.0 on
1608 a held-out test set containing 40 data points.

1609

1610 **Figure AIII.6**

1611 *GMM fit on average ratings. The dashed line represents the model fit by plotting the prediction*
1612 *of average complexity ratings against a continuous range of average beauty ratings.*

1613



1614

1615

1616 8. Check for Individual Differences

1617

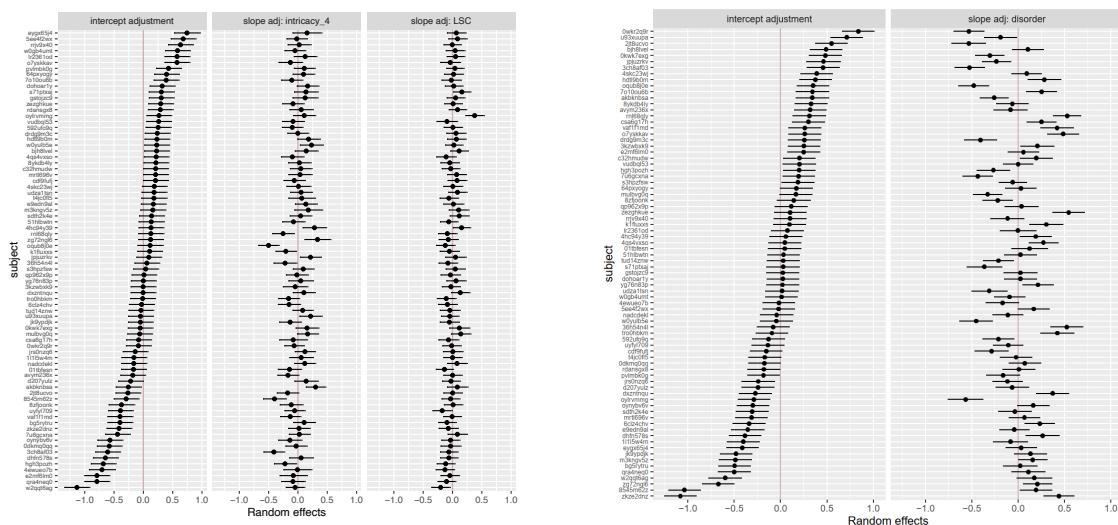
1618 Figure AIII.7 displays a plot of random effects for the best performing model of complexity
1619 ratings (A) and beauty ratings (B). The plots show that random intercepts for participant have
1620 a higher variance than random slopes for either intricacy or disorder. The variation in random
1621 intercepts of participants show that while some people rate complexity to be high on average,
1622 some others rate them to be low, and similarly for beauty. Moreover, the random slopes indicate
1623 that people perceive the impact of LSC, intricacy, and disorder on complexity and beauty
1624 respectively to different degrees. More thorough analysis of individual differences is intended
1625 to be a part of our future work.

1626

1627 **Figure AIII.7**

1628 *Plot of random effects from our best performing (A) complexity model and (B) beauty model*

1629



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

(B) Random intercept of participant and
random slope of disorder

1630

1631 **References**

1632

1633 Finkel, R. A., & Bentley, J. L. (1974). Quad trees a data structure for retrieval on composite
1634 keys. *Acta informatica*, 4(1), 1-9.

1635

1636 Packard, N. H., & Wolfram, S. (1985). Two-dimensional cellular automata. *Journal of
1637 Statistical physics*, 38(5), 901-946.

1638

1639 Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of modern
1640 physics*, 55(3), 601.