Journal of Experimental Psychology: General

The Order of Disorder: Deconstructing Visual Disorder and Its Effect on Rule-Breaking

Hiroki P. Kotabe, Omid Kardan, and Marc G. Berman

Online First Publication, October 13, 2016. http://dx.doi.org/10.1037/xge0000240

CITATION

Kotabe, H. P., Kardan, O., & Berman, M. G. (2016, October 13). The Order of Disorder: Deconstructing Visual Disorder and Its Effect on Rule-Breaking. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000240

Journal of Experimental Psychology: General

0096-3445/16/$12.00 http://dx.doi.org/10.1037/xge0000240

The Order of Disorder: Deconstructing Visual Disorder and Its Effect on Rule-Breaking

Hiroki P. Kotabe, Omid Kardan, and Marc G. Berman
University of Chicago

Disorderly environments are linked to disorderly behaviors. Broken windows theory (Wilson & Kelling, 1982), an influential theory of crime and rule-breaking, assumes that scene-level social disorder cues (e.g., litter, graffiti) cause people to reason that they can get away with breaking rules. But what if part of the story is not about such complex social reasoning? Recent research suggests that basic visual disorder cues may be sufficient to encourage complex rule-breaking behavior. To test this hypothesis, we first conducted a set of experiments (Experiments 1–3) in which we identified basic visual disorder cues that generalize across visual stimuli with a variety of semantic content. Our results revealed that spatial features (e.g., nonstraight edges, asymmetry) are more important than color features (e.g., hue, saturation, value) for visual disorder. Exploiting this knowledge, we then reconstructed stimuli contrasted in terms of visual disorder, but absent of scene-level social disorder cues, to test whether visual disorder alone encourages cheating in a second set of experiments (Experiments 4 and 5). In these experiments, manipulating visual disorder increased the likelihood of cheating by up to 35% and the average magnitude of cheating by up to 87%. This work suggests that theories of rule-breaking that assume that *complex social reasoning* (e.g., about norms, policing, poverty) is necessary, should be reconsidered (e.g., Kelling & Coles, 1997; Sampson & Raudenbush, 2004). Furthermore, these experiments show that simple perceptual properties of the environment can affect complex behavior and sheds light on the extent to which our actions are within our control.

Keywords: environmental disorder, visual disorder, social disorder, rule-breaking, broken windows theory

Supplemental materials: http://dx.doi.org/10.1037/xge0000240.supp

We feel, think, and act differently when we are in disorderly environments. Disorderly environments have been linked to detrimental outcomes such as perceived powerlessness (Geis & Ross, 1998), distress (Cutrona, Russell, Hessling, Brown, & Murry, 2000), fear of crime and feeling unsafe (Perkins & Taylor, 1996), depression (Ross, 2000), anxiety and performance-monitoring (Tullett, Kay, & Inzlicht, 2015), and self-regulatory failure (Chae & Zhu, 2014; Vohs, Redden, & Rahinel, 2013; for a review, see Kotabe, 2014). Such psychological effects of disorderly environments likely have downstream effects on complex human behaviors. Research suggests that one domain of complex behaviors that may be affected is rule-breaking. According to a prominent sociological theory of rule-breaking and crime called broken windows theory (BWT; Wilson & Kelling, 1982), even minor signs of disorder can encourage rule-breaking behaviors that snowball into major communal problems such as delinquency and criminality. Careful field experiments have shown that the effect of disorderly environments on rule-breaking not only spreads within a domain (e.g., litter begets more litter), but also between domains (e.g., litter begets theft; Keizer, Lindenberg, & Steg, 2008), further compounding the problem. Additional research employing both field and laboratory experimental methods and large-scale correlational methods has further corroborated that disorderly environments encourage impulsive and disorderly behaviors (Braga & Bond, 2008; Braga et al., 1999; Chae & Zhu, 2014; Linares et al., 2001; Vohs, Redden, et al., 2013).

As for the process by which this occurs, the dominant hypotheses for BWT phenomena posit that complex social reasoning about scene-level social cues indicative of rule-breaking behavior (e.g., litter, graffiti, an abandoned building; scene-level social

---

Hiroki P. Kotabe, Omid Kardan, and Marc G. Berman, Department of Psychology, University of Chicago.

This work was supported by a grant from the TKF Foundation to Marc G. Berman, two grants from the John Templeton Foundation (the University of Chicago Center for Practical Wisdom and the Virtue, Happiness, and Meaning of Life Scholars Group) to Marc G. Berman, and a grant from the National Science Foundation (1632445) to Marc G. Berman, as well as an internal grant from the University of Chicago to Marc G. Berman. We give special thanks to Daniel Casasanto, Reid Hastie, and Wilhelm Hofmann for reviewing this work as members of Hiroki P. Kotabe's dissertation committee. We also thank Sian Beilock, Frank Ibarra, John Jonides, Ethan Kross, Howard Nusbaum, Kate Schertz, and Max Weisbuch (the editor) for their useful feedback on this article. An earlier version of this article was incorporated in Hiroki P. Kotabe's dissertation. In addition, some of the ideas and data in this article were presented in a poster at the 28th Annual Convention of the Association for Psychological Science (APS). Specifically, the gist of the introduction and results from Experiment 4 were presented.

Correspondence concerning this article should be addressed to Hiroki P. Kotabe, Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637. E-mail: hkotabe@uchicago.edu

disorder cues) causes increased rule-breaking (Kelling & Coles, 1997; Sampson & Raudenbush, 2004; Wilson & Kelling, 1982). For example, when seeing such disorder cues people may reason that misconduct is normal, policing is absent, or poverty is prevalent (Sampson & Raudenbush, 2004), all of which may increase the likelihood of breaking rules.

There is a general problem with such perspectives which is that they are based on research that has not clearly defined and assessed "disorder," thus any specific interpretations of the evidence are dubious (Harcourt, 2009). One of the specific issues arising from the general problem is that previous research has confounded social disorder and visual disorder, perhaps because of a lack of theoretical rationale to expect an effect of the latter. We define "visual disorder" as the perception of disorder that is attributable to basic (or low-level) visual features (i.e., spatial and color features; basic visual disorder cues). For example, Wilson and Kelling (1982) used the term "disorder" in reference to both environments varying in disorder due to the presence versus absence of litter or graffiti, and to environments varying in disorder due to the presence of a delinquent versus rule-abiding citizen. In our view, the former environments would vary substantially in basic visual features whereas the latter environments would not (e.g., imagine the same person drunk in public vs. sober), but they do not acknowledge this systematic difference. Sampson and Raudenbush (2004) attempt to separate a "physical disorder" component of environmental disorder, but their operationalization (ratings for three questions: how much of a problem is litter/trash, graffiti, and vacant housing/storefronts [in your neighborhood]?) also does not make it clear to what extent basic visual disorder cues versus scene-level social disorder cues in the environment are relevant. Psychological experiments bearing on BWT (e.g., Chae & Zhu, 2014; Keizer et al., 2008; Vohs, Redden, et al., 2013) also have not attempted to separate the influence of social disorder and visual disorder. Despite this ambiguity, researchers continue to attribute their findings entirely to scene-level social disorder cues while overlooking basic visual disorder cues.

This led us to wonder whether basic visual disorder features have been playing a role all along. Considerable evidence from different areas of scientific inquiry suggests that basic visual processing can affect complex behavior. For example, although it has long been thought that basic visual features alone do not carry semantic[1] information (e.g., Biederman, 1987; David Marr, 1976; Marr & Hildreth, 1980), recent research suggests this is not true (Kotabe, Kardan, & Berman, 2016a; Oliva & Torralba, 2006; Walther, Caddigan, Fei-Fei, & Beck, 2009; see also Kardan, Henderson, Yourganov, & Berman, 2016). The assumption that basic visual features carry semantic information helps to explain how semantic categories of scenes can be decoded from activity in V1 (cortical region that receives visual sensory input from the thalamus) nearly as well as from activity in the parahippocampal place area (cortical region consistently involved in the encoding and recognition of environmental scenes; Walther et al., 2009); semantic categories of scenes can be predicted by holistic spatial properties of the scenes (Oliva & Torralba, 2006); and decision making can occur in visual cortex (Brascamp, Blake, & Knapen, 2015). Furthermore, there are numerous feedforward and feedback projections from V1 to various higher cortical areas involved in complex semantic processing and behavior (Felleman & Van Essen, 1991; Gilbert & Li, 2013; Lamme & Roelfsema, 2000; McIntosh et al., 1994). Such research suggests that basic visual features are not devoid of semantic information, but may be richly endowed with it. For example, straight versus nonstraight edges may carry semantic information about deontic concepts, which is further evidenced by a family of linguistic metaphors as in "he's as straight as an arrow" and "he's bending the rules" (for a review of research on the linkage between spatial representations and abstract concepts, see Casasanto & Bottini, 2014). In turn, the activation of such concepts may have downstream effects on related complex behaviors such as rule-breaking (Bargh, 2006; Molden, 2014). More generally, the research reviewed here suggests that different spatial features may carry varying degrees of semantic information about the disorderliness of an environment. Through associative judgment, there may be subsequent downstream effects on rule-breaking, thus suggesting at least one way in which disorderly environments may encourage rule-breaking through psychological processes beyond those posited by BWT.

Moreover, as a separable mechanism, the processing of basic visual features varies in terms of processing difficulty (Field, 1987; Kinchla, 1977; Olshausen & Field, 1996; Witkin & Tenenbaum, 1983), which could have downstream consequences on behavior. A key process in early visual processing that concerns processing difficulty is the organization of visual information into coherent and manageable chunks (Witkin & Tenenbaum, 1983; see also Mahoney, 1987). Relatedly, structural redundancy (e.g., as found in symmetry) may reduce processing difficulty by increasing processing efficiency (Kinchla, 1977; see also Field, 1987; Olshausen & Field, 1996). Such research suggests that processing visually disordered scenes may be more difficult than processing visually ordered scenes, holding all else constant, to the extent that basic visual disorder cues are subjectively less coherent and structurally less redundant. Theories of self-regulation suggest that processing difficulty would fatigue cognitive resources necessary for self-regulation (Hofmann, Friese, Schmeichel, & Baddeley, 2011; Kaplan & Berman, 2010; Kotabe & Hofmann, 2015), which may lead to unregulated rule-breaking behavior (Gino, Schweitzer, Mead, & Ariely, 2011; Shalvi, Eldar, & Bereby-Meyer, 2012). Furthermore, the experience of processing difficulty, or "disfluency," may also be used as a metacognitive cue in judgment, with disfluency activating effortful thinking (Alter, Oppenheimer, Epley, & Eyre, 2007) and generally being interpreted in some negative way through people's naïve theories about its meaning (Alter & Oppenheimer, 2009; Schwarz, 2004). Effortful thinking and regulation of negative thoughts and feelings may further fatigue the capacity to execute self-regulated behavior (Kotabe & Hofmann, 2015; Muraven & Baumeister, 2000).

All in all, it is plausible that even basic visual disorder cues could affect complex human behaviors such as rule-breaking. The goal of this review was not to make strong arguments for one mechanism or another, but rather to support the plausibility of the general hypothesis. Here, we focus on testing the general hypothesis to determine whether there is truth to our concern that social disorder and visual disorder are confounded in previous research

---

[1] By "semantic," we are referring to meaningful information such as that involved in the recognition of objects, places, and general descriptors; thus, any concept is considered semantic. For example, "car"/"building" (objects), "park"/"city" (places), and "urban"/"disorderly" (general descriptors) would all be considered semantic.

bearing on BWT, and whether the latter can have an effect in isolation. To answer these questions, our first goal was to define visual disorder by identifying specific basic visual disorder cues. Our second goal was to examine if exposure to these cues can encourage rule-breaking behavior. Toward the first goal, we took a principled approach to quantifying, extracting, and scrambling objective visual features of various environmental scenes (as recommended by Geisler, 2008) and analyzed hundreds of people's disorder ratings for these scenes as well as a wide variety of stimuli derived from these scenes. Toward the second goal, we conducted behavioral experiments using large and diverse online samples to investigate whether basic visual disorder cues alone could encourage cheating behavior.

Overview of Experiments

Across all of our experiments, we sampled broadly from real-world environments by utilizing 260 images of environmental scenes that ranged from more urban to more natural according to ratings previously collected in the laboratory (Berman et al., 2014; Kardan et al., 2015; see Figure 1 for examples; all images can be downloaded here in original resolution: goo.gl/S8ShgT).[2] Experiments 1–3 aim to quantify visual disorder by identifying some of the basic visual cues that drive disorder judgments. In these experiments, we utilize novel quantitative methods to measure and manipulate statistical regularities in a wide assortment of scenes. Such methods have great potential for advancing research on visual perception (Geisler, 2008). Furthermore, these experiments provide the scientific community with novel methods to manipulate not only visual disorder elements, but more generally, basic spatial and color features in isolation from higher-level scene semantics. In Experiment 1, we had people provide disorder ratings for the scene images in their original form. In Experiments 2 and 3, we had people provide disorder ratings for derived stimuli which contained basic visual features but no scene-level social disorder cues. Analysis was conducted on image-level summary statistics. Experiments 4 and 5 focused on testing the broader hypothesis pertinent to BWT that basic visual disorder cues alone would be sufficient to encourage rule-breaking. In these experiments, we reconstructed visual disorder based on what was learned from deconstructing visual disorder in Experiments 1–3. Analysis was conducted at the level of individual participants.

Experiment 1: Quantifying Visual Disorder Part 1

In this experiment, we had people rate the scene images in terms of disorder. In addition, we quantified spatial- and color-related basic visual features of these images (as in Berman et al., 2014; Kardan et al., 2015) to test the extent to which they predicted disorder judgments. Spatial features included nonstraight edge density, straight edge density, and asymmetry. According to research referenced in the introduction, these spatial features may be particularly important because of their relation to mental metaphors (e.g., manifested in linguistic metaphors such as "bending the rules;" "straight as an arrow;" and "even-keeled"), and due to their relation with processing difficulty (i.e., nonstraight lines and asymmetry have less structural redundancy. Color features were taken from the standard hue-saturation-value representation of the full RGB color space and included mean hue, saturation, and value and standard deviations of hue, saturation, and value as measures of hue diversity, saturation diversity, and value diversity. According to research referenced in the introduction concerning semantics carried by basic spatial features and the role of redundancy in efficient visual processing, we hypothesized that spatial features would influence disorder judgments. In contrast, we were uncertain about any effect of color features because most if not all of the research on visual processing difficulty concerns spatial features and also we do not know of any linguistic metaphors that clearly link color features to deontic concepts. Furthermore, in the experiments of Vohs et al. (2013) and Chae and Zhu (2014), they successfully manipulated environmental disorder by varying the spatial arrangement of the same objects in a room, thus varying spatial features significantly but color features minimally. Again, this substantiates the hypothesis that spatial features would influence disorder judgments, but provides no evidence that colors would have an influence. If spatial features influence visual disorder but color features have minimal influence, it would considerably reduce the dimensions of the visual feature space necessary to manipulate visual disorder.

Method

Participants and design. One-hundred and five U.S.-based adults (51 men; 54 women) were recruited from the online labor market Amazon Mechanical Turk (AMT).[3] Sample size and stopping rule were based on our goal to receive ~20 ratings per image. Ages ranged from 19 to 75 ($M = 36.15$, $SD = 12.07$). Eighty-four participants identified primarily as White/Caucasian, 10 identified as Black/African American, five identified as Hispanic/Latino, four identified as Asian/Asian American, and two identified as Native Hawaiian/Pacific Islander. The median experiment duration was 5 min and 58 s and participants were compensated $0.50 for participating. Informed consent was administered by the Institutional Review Board (IRB) of the University of Chicago.

Procedure. Participants first received instructions that they would be presented a series of 50 images of various environmental scenes and that they were to rate each scene in terms of how disorderly or orderly it looked. Here, and in Experiments 2 and 3, we did not explicitly define disorder to participants because our goal was to evaluate systematic relationships between basic visual features and people's subjective impressions of disorder.

Next, they were taken to the image rating task (IRT). Scene images (all 4:3 ratio) were presented on a plain white background in a 600 × 450 pixel frame. Below the image frame, text was presented that asked "How disorderly or orderly is this environ-

---

[2] Relevant to the ecological validity of scene images, it was shown that walking in urban versus natural environments has similar effects on directed-attention performance as viewing images of urban versus natural environments (Berman, Jonides, & Kaplan, 2008). Furthermore, scene images are used extensively in research on the perception of naturalistic environments.

[3] The participant count here and elsewhere does not include those who did not complete the essential procedural components (i.e., they quit before reaching the demographics form). On AMT, workers on occasion do not complete a study (e.g., they may only get to the consent form), but this does not seem to be a major issue because the data are nevertheless demonstrably of high quality (see supplementary material, Section 1). In case of concerns about attrition rates, see the Table S1 for attrition rates of each experiment.

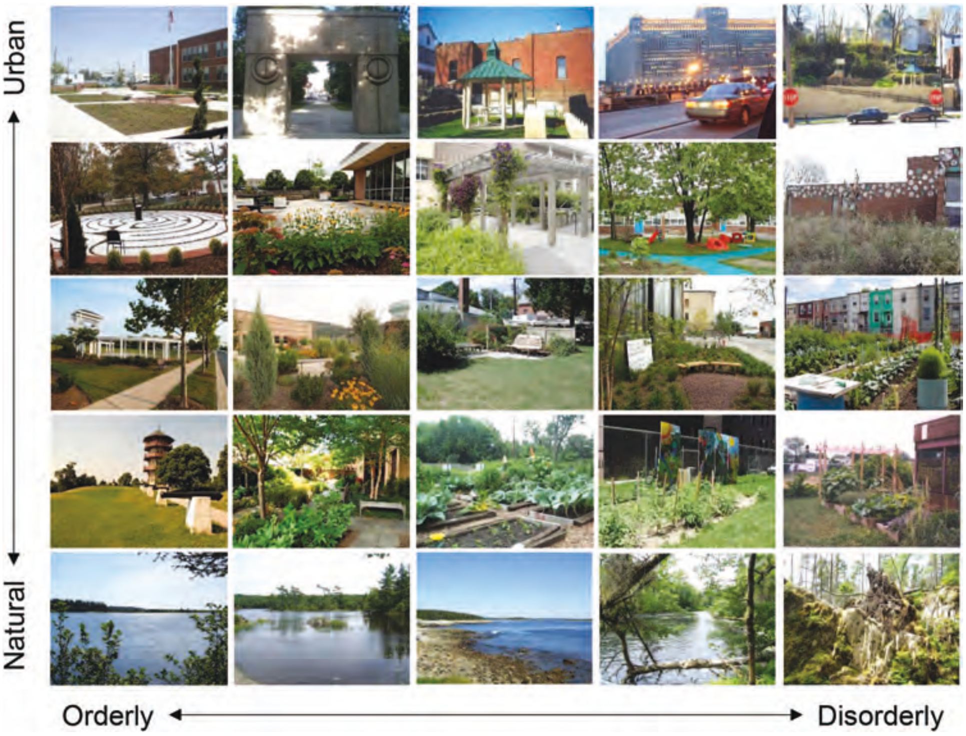

Figure 1. Matrix of examples of the 260 scene images utilized for this study. Each row represents a quintile of naturalness/urbanness and each column represents a quintile of order/disorder. See the online article for the color version of this figure.

ment?" And below that was a 7-point semantic differential scale anchored by the options *very disorderly* and *very orderly*. Each participant was randomly presented 50 of the 260 scene images. The randomization scheme had two layers. First, we randomly selected 10 images from each quintile of urbanness/naturalness to ensure that each participant rated a wide sample of scene types. Second, we presented these 50 images in random order. Immediately after making a rating, they would automatically proceed to the next image until all 50 images were rated. Here and in Experiments 2 and 3, presentation time was not fixed in order to assess spontaneous disorder ratings.

Quantifying spatial and color features. We utilized MATLAB's Image Processing Toolbox to quantify three basic spatial features and six basic color features to statistically estimate how much perceived disorder could be explained by to objective spatial and color features of the scene. The spatial features we quantified were nonstraight edge density (a measure of how many nonstraight edges are in the scene image), straight edge density (a measure of how many straight edges are in the scene image), and vertical reflectional asymmetry ("asymmetry" for short; a measure of how well the left and right halves of the image mirror each other). The resulting color features, based on the standard hue-saturation-value (HSV) model, were mean hue (a measure of the average color appearance of a scene), mean saturation (a measure of how intense or pure the colors of the scene are on average), and mean value (a measure of the average luminance of a scene). We also calculated standard deviations of those color measures as measures of hue diversity, saturation diversity, and value diversity. Straight edge density and nonstraight edge density and saturation, value, *SD* saturation, and *SD* value were all quantified from their respective maps created as in (Berman et al., 2014; Kardan et al., 2015). Because hue of a pixel is an angular value, mean and *SD* hue were calculated using circular statistics (Circular Statistics Toolbox for MATLAB, Berens, 2009). Asymmetry was quantified by summing up the dot product of the left and mirrored-right half of the edge map of images. These sums were then normalized to [0 1] range by being divided by the total number of nonzero pixels in the edge map of the corresponding image (total edge space). See Table S2 in supplemental materials for a correlation matrix of all of these visual features.

Results

Because we aggregated individual raters to estimate the overall disorder of each scene, it was important to estimate interrater reliability. We estimated interrater reliability with Shrout and Fleiss's (1979) Case 2 intraclass correlation formula which utilizes a two-way random effects model in which image and rater are both modeled as random effects. The model yielded $rel = .89$, $p < .001$, 95% CI [.87, .91], which falls in the conventionally "excellent" range (Cicchetti, 1994).

We conducted a multiple linear regression analysis in which mean disorder ratings were regressed on all of the individual spatial and color features. About a fifth of the variance in mean disorder ratings was explained by these visual features, $R^2_{\mathrm{adj}} = .17$. Nonstraight edge density had the largest effect (see Table 1). A

Table 1
Basic Visual Features Predicting Disorder Ratings in Experiment 1

Predictor	β	*SE*	*p*	η_p^2
Spatial				
Nonstraight edge density	.74	.11	<.001	.150
Straight-edge density	.17	.08	.006	.019
Asymmetry	.21	.09	.004	.021
Color				
Hue	−.12	.07	.024	.013
Saturation	−.20	.08	.002	.024
Value	.04	.06	.587	.002
SD hue	.16	.08	.089	.019
SD saturation	.06	.08	.458	.002
SD value	.07	.06	.456	.004

linear contrast comparing beta coefficients indicated that the average effect of the spatial features was significantly larger than the average effect of the color features, $F = 11.46$, $p = .001$. Although hue and saturation were significant predictors whereas value was not, linear contrasts indicated that their effect sizes did not significantly differ from one another. Because we had no hypothesis about differences between individual color predictors, we do not interpret these results.

To compare variance in disorder ratings explained by spatial versus color features, we also conducted two multiple linear regression analyses separately regressing mean disorder ratings on the spatial features versus the color features (see supplementary materials for results). Adjusting for the number of predictors, the spatial features explained over 10 times as much variance as the color features, $R^2_{\text{adj}} = .10$ versus $R^2_{\text{adj}} < .01$.

These results suggest that spatial features—particularly nonstraight edge density—are indeed more important than color features for visual disorder. However, because the scene images contained not only various basic visual disorder cues but also various possible scene-level social disorder cues, we did not have experimental control over whether the latter influenced the results. In the following experiments, we extracted and scrambled basic visual features from the scene images to remove scene-level social disorder cues. Thus, in the following experiments the possibility of confounding basic visual disorder cues and scene-level social disorder cues was substantially reduced.

Experiment 2: Quantifying Visual Disorder Part 2

We separately extracted and scrambled the edge features and the color features from the scene images to remove scene-level social disorder cues while preserving basic visual features of the scenes (see Figure 2b–c). People were randomly assigned to rate these derived scrambled-edge or scrambled-color stimuli in terms of disorder as in Experiment 1. With these ratings, we could statistically estimate the extent to which perceived disorder at the scene level was a function of visually disordered edges versus visually disordered colors. Based on the results of Experiment 1, we predicted that scene-level disorder would be better predicted by visually disordered edges than visually disordered colors.

Method

Participants and design. One-hundred and 91 U.S.-based adults (108 men, 82 women, one other) were recruited from AMT and participated in this two-condition (stimuli: scrambled-edge stimuli vs. scrambled-color stimuli) between-subjects experiment. Sample size and stopping rule were based on our goal to receive ~20 ratings per image. Ages ranged from 18 to 64 ($M = 32.16$, $SD = 11.09$). One-hundred and 59 participants identified primarily as White/Caucasian, 11 identified as Asian/Asian American, 10 identified as Black/African American, eight identified as Hispanic/Latino, and three identified as other. The median experiment duration was 4 min and 16 s and participants were compensated $0.50. Informed consent was administered by the IRB of the University of Chicago.

Creating scrambled-edges and scrambled-color stimuli. For the scrambled-edge stimuli, we devised a novel method to remove scene-level social disorder cues while preserving edge formations from the original scene images as much as possible (see Figure 3 for an illustration of the processes of this method). First, we created an edge map from the original scene images, created as in (Berman et al., 2014; Kardan et al., 2015; Figure 3, Process 1). Next, the edge map of the target image was randomly rotated either 90 or 270 degrees and overlaid on the 180-degrees rotated edge map (Figure 3, Process 2), creating a stimulus comprising twice as many edges (but same straight and nonstraight edge ratios) as the scene image. A mask matrix was then constructed to be the same

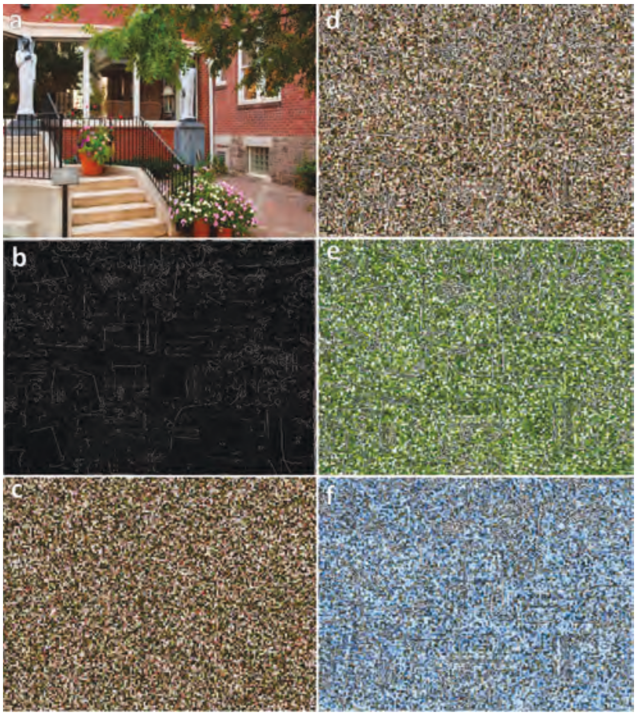

Figure 2. One sample scene image and its derived stimuli. (a) The original image used in Experiment 1; (b) its scrambled-edge stimulus and (c) scrambled-color stimulus used in Experiment 2; and (d) its color-congruent stimulus, (e) color-incongruent stimulus, and (f) control stimulus used in Experiment 3. All images can be downloaded here in original resolution: goo.gl/S8ShgT.

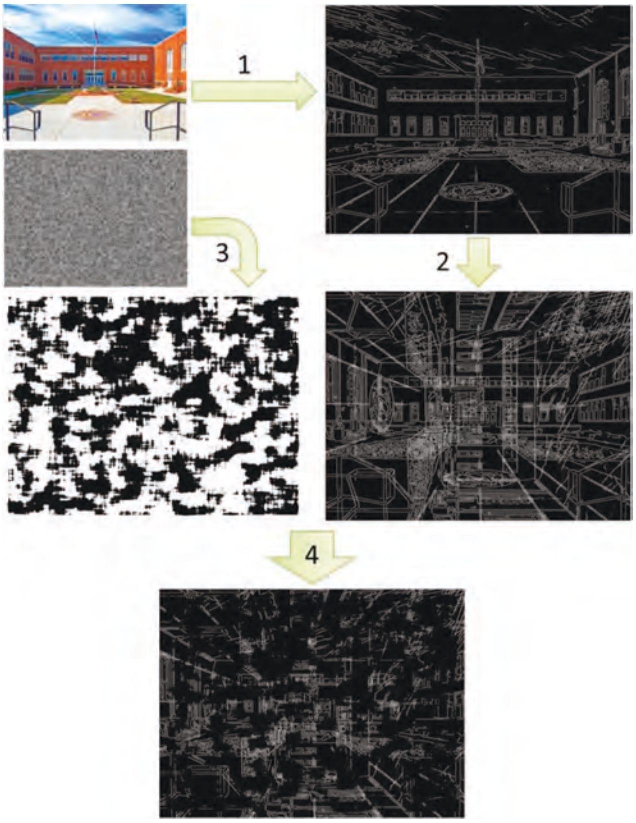

Figure 3. Illustration of the edge extraction and scrambling process. Process 1: Created an edge map from the original image. Process 2: The edge map was randomly rotated either 90 or 270 degrees and overlaid on the 180-degrees rotated edge map creating a stimulus comprising twice as many edges (but same straight and nonstraight edge ratios) as the scene image. Process 3: Started with a mask matrix constructed to be the same size as the scene images (600 × 800) with its elements randomly assigned between zero and one. This matrix was then convolved with a median filter sized 30 × 40 pixels. In this way, patches of 1s and 0s were made randomly and placed at random locations across the mask with random sizes equal to or greater than 30 × 40 pixels, with half of every mask having, on average, half a surface of 1s and half a surface of 0s. Process 4: This mask was then multiplied (dot product) by the stimulus from Process 2 so that half of its edges were removed at random, resulting in the final stimulus (bottom). See the online article for the color version of this figure.

size as the scene images (600 × 800) with its elements randomly assigned between zero and one. This matrix was then convolved with a median filter sized 30 × 40 pixels (Figure 3, Process 3). In this way, patches of 1s and 0s were made randomly and placed at random locations across the mask with random sizes equal to or greater than 30 × 40 pixels, with half of every mask having, on average, half a surface of 1s and half a surface of 0s. This mask was then multiplied (dot product) by the stimuli from Process 2 so that half of its edges were removed at random (Figure 3, Process 4). The resulting stimulus had, on average, the same amount of edges with similar edge types as the original scene image from which it was derived.

For the scrambled-color stimuli, we randomly repositioned windows of 5 × 5 pixels from the image. The window size was selected so that (a) scene-level social disorder cues would become nondiscernable, and (b) the color textures of the scene would be preserved. For example, using a 1 × 1 pixel window size resulted in stimuli in which less frequent colors were so scattered that they became invisible to the eye whereas using a 10 × 10 pixel window kept some of the objects or parts of the scene identifiable, thus possibly preserving scene-level social disorder cues.

Procedure. The procedure was the same as in Experiment 1 except that participants in Experiment 2 were randomly assigned to rate a random 50 of the 260 scrambled-edge stimuli or a random 50 of the 260 scrambled-color stimuli—10 randomly selected from each urbanness/naturalness quintile (based on urbanness/naturalness ratings for the original scene images).

Results

We estimated interrater reliability as in Experiment 1. For the scrambled-edge stimuli, the intraclass correlation model yielded $rel = .90$, $p < .001$, 95% CI [.88, .92], and for the scrambled-color stimuli, the intraclass correlation model yielded $rel = .66$, $p < .001$, 95% CI [.60, .71]. The former reliability estimate falls within the conventionally "excellent" range and the latter reliability estimate falls within the conventionally "good" range (Cicchetti, 1994).

We conducted two regression models in which we separately regressed disorder ratings for the scene images (collected in Experiment 1) on disorder ratings for the scrambled-edge versus scrambled-color stimuli (collected in Experiment 2; see Table S3 in supplemental materials for a correlation matrix of disorder ratings for these three sets of stimuli). Disorder ratings for the scrambled-edge stimuli significantly predicted disorder ratings for the scene images, $\beta = 0.38$, $p < .001$ ($R^2_{\text{adj}} = .144$). In contrast, disorder ratings for the scrambled-color stimuli did *not* significantly predict disorder ratings for the scene images, $\beta = 0.02$, $p = .731$ ($R^2_{\text{adj}} = .00$). The adjusted R^2s in this study were similar to the adjusted R^2s when separately regressing disorder ratings on the spatial features ($R^2_{\text{adj}} = .124$) versus the color features ($R^2_{\text{adj}} = .038$) in Experiment 1, supporting the validity of our edge and color extraction methods.

These results further corroborate that visual disorder is more a function of spatial features than color features. However, in the stimuli used in this experiment, edges and colors were isolated from one another so we could not speak to the possibility of an interaction. In the next experiment, we created stimuli that simultaneously contained edge and color features, and experimentally manipulated the color features to further test for an effect of color features on disorder judgments.

Experiment 3: Quantifying Visual Disorder Part 3

We manipulated the color features while holding the edge features constant (see Figure 2d–f): In the "color-congruent" condition, scrambled edges from disorderly (orderly) scenes were paired with scrambled colors from disorderly (orderly) scenes (i.e., edges and colors from the same scene). In the "color-incongruent" condition, scrambled edges from disorderly (orderly) scenes were paired with scrambled colors from orderly (disorderly) scenes (i.e., edges and colors from different scenes). In the control condition, scrambled edges were paired with scrambled colors from randomly selected scenes (i.e., edges and colors from different scenes). People were assigned to rate stimuli from all three conditions in terms of disorder. These conditions allowed us to experimentally test whether edge features are more heavily weighted into disorder judgments than are color features when both types of features are simultaneously present. If edge features are more important than color features in determining visual disorder, the color manipulation should have a relatively small effect. If color features are more important than edge features, the color manipulation should have a relatively large effect.

Method

Participants and design. Two-hundred and 22 U.S.-based adults (111 men, 111 women) were recruited from AMT and participated in this three-condition (stimuli: control vs. color-congruent stimuli vs. color-incongruent stimuli) within-subjects experiment. Sample size and stopping rule were based on our goal to receive ~20 ratings per image. Ages ranged from 19 to 76 ($M = 35.41$, $SD = 11.31$). One-hundred and 78 participants identified primarily as White/Caucasian, 17 identified as Asian/Asian American, 13 identified as Black/African American, 11 identified as Hispanic/Latino, two identified as multiple ethnicities, and one identified as Native American. The median experiment duration was 5 min and 58 s and participants were compensated $0.75. Informed consent was administered by the IRB of the University of Chicago.

Creating color-congruent and color-incongruent (and control) stimuli. To create the "scrambled-edge-on-scrambled-color" stimuli we overlaid the previously made scrambled-edge stimuli on the scrambled-color stimuli. First, the scrambled-edge images were sorted in order of visual disorder, that is, image 1, 2,. . ., k . . ., 260, with image 1 being the most visually ordered and image 260 being the most visually disordered, according to ratings from Experiment 2. Then the scrambled-edge stimulus made from k_{th} image was overlaid on the scrambled-color stimulus that was (a) made from the same image (color-congruent stimuli); (b) made from the $(260 – k)_{th}$ image (color-incongruent stimuli); or (c) made from j_{th} image where j is a random number between one and 260 without resampling (control stimuli).

Because in the resulting stimuli some of the scrambled edges were not discernable from the background scrambled colors, we made one pixel surrounding the edges (which are white) black to preserve the contrast and consistency of the edges. We note that although this did inevitably remove some color information, the proportion of remaining pixels belonging to color content (~82% on average) was still much larger than the proportion of remaining pixels belonging to edge content (~18% on average; see Figure 2d–f), making our test of the hypothesis that edge features are more important than color features for visual disorder, in favor of the color features nonetheless.

Procedure. The procedure was the same as in Experiment 2 except for the following differences: Instead of a between-subjects design, participants in Experiment 3 were assigned to all three visual conditions within-subjects. They rated a total of 75 images—25 of the 260 new color-congruent images, 25 of the 260 new color-incongruent images, and 25 of the 260 new control images. The randomization scheme was similar to that used in Experiments 1 and 2: First, we randomly selected five images from each disorder quintile (based on disorder ratings for the original scene images), and repeated this for each of the three sets of color stimuli, resulting in 25 images from each set. Second, we presented these 75 images in random order.

Results

We estimated interrater reliability as in the previous experiments. For the color-congruent stimuli, the intraclass correlation model yielded $rel = .71$, $p < .001$, 95% CI [.65, .75]; for the color-incongruent stimuli, the intraclass correlation model yielded $rel = .73$, $p < .001$, 95% CI [.69, .78]; and for the control stimuli, the intraclass correlation model yielded $rel = .68$, $p < .001$, 95% CI [.62, .73] (all within the high end of the conventionally "good" range). The reliability estimates for these stimuli fell between the reliability estimates for the scrambled-edge and scrambled-color stimuli from which they were derived.

The analysis here differs from the analyses in the previous experiments because in this experiment we manipulated edge and color features within-subjects. This allowed us to test which of these manipulations had a stronger effect on disorder judgments in two ways. First, to determine whether edge features were more important than color features for visual disorder, we compared the disorder ratings for the three new sets of experimental stimuli with the disorder ratings for the scene images (collected in Experiment 1; see Table S4 in supplemental materials for a correlation matrix of disorder ratings for these four sets of stimuli). If edge features were more important than color features for disorder judgments at the scene level, we would expect similar correlations between disorder ratings for each set of color-manipulated stimuli and disorder ratings for the original scene images (i.e., the color manipulation would have little effect). This was indeed the case (see Figure 4). Williams' (1959) t tests confirmed that none of the

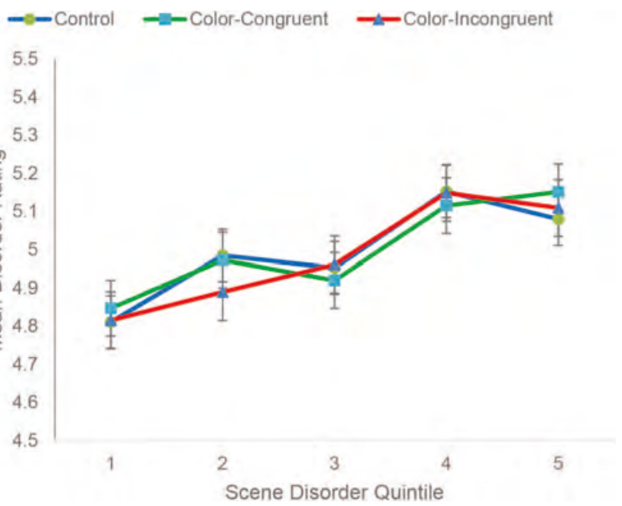

Figure 4. Mean disorder ratings by scene disorder quintile in Experiment 3. The x-axis indicates the quintiles of scene disorder ratings (collected in Experiment 1) on which the color manipulation in Experiment 3 was based. The overlapping lines indicates that manipulating color features had relatively little effect on disorder ratings compared to manipulating edge features. Error bars indicate mean ± SEM. See the online article for the color version of this figure.

pairwise dependent correlations significantly differed. Second, a repeated-measures GLM indicated that there was no significant interactive effect of stimuli set (color-congruent vs. color-incongruent vs. control) and scene-disorder quintile (collected in Experiment 1) on disorder ratings for the color-manipulated stimuli, again suggesting that the color manipulation had little effect on disorder judgments.

These results further demonstrate that edge features are more important than color features for visual disorder. The three experiments so far each shed light on some key quantitative visual features that define visual disorder. These results not only help us define visual disorder, but they also give us something tangible to work with to manipulate visual disorder, which was pivotal for the following experiments. In the following experiments, we switched focus to conduct the first investigation into whether being exposed to basic visual disorder cues alone (as defined in Experiments 1–3) is sufficient to encourage rule-breaking, despite the absence of scene-level social disorder cues and complex social reasoning. These experiments bear on the original question we had of whether the confounding of social disorder and visual disorder in previous research bearing on BWT is problematic. They would also serve as a test of the theoretical rationale we presented in the introduction, namely, that visual disorder could encourage rule-breaking behavior through various plausible mechanisms including differential processing difficulty/effort and associative/metaphorical thinking.

Experiment 4: The Effect of Visual Disorder on Rule-Breaking Part 1

For the following two experiments, we focused on one major form of rule-breaking—cheating. Cheating is a domain particularly suitable for the study of rule-breaking more generally (see Ariely, 2013) in that cheating situations possess the typical characteristics of a rule-breaking situation in which one wants to do X but a proscriptive norms are telling them not to do X. When faced with this motivational conflict between enacting a short-term, selfish desire versus the proscriptively normative behavior, one must use self-control to pursue the virtuous path (Gino et al., 2011). Relatedly, some scholars argue that the absence of self-control is the single most important factor in producing criminal behavior (Gottfredson & Hirschi, 1990).

To study cheating, we adapted an experimental procedure developed by Mazar, Amir, and Ariely (2008). This procedure involves taking a challenging incentivized test, and then later grading oneself, which provides an opportunity to cheat to varying degrees. Immediately before grading themselves, participants were exposed to either visually disordered stimuli or visually ordered stimuli for five minutes. In the self-grading phase, participants were told they would receive bonus money for each question they reported as correct. We predicted that participants exposed to the visually disordered stimuli would cheat more than participants exposed to the visually ordered stimuli.

Method

Participants and design. Four-hundred and five U.S.-based adults (206 women, 198 men, one unreported) were recruited from AMT and participated in this two-condition (visually disordered stimuli vs. visually ordered stimuli) between-subjects experiments. Sample size and stopping rule were based on our goal to obtain a large sample size of ~200 per condition to increase power to detect an effect of unknown size. Ages ranged from 19 to 69 ($M = 35.34$, $SD = 11.05$). Three-hundred and 14 participants identified primarily as White/Caucasian, 29 as Hispanic/Latino, 25 as Black/African American, 22 as Asian/Asian American, 12 as multiple ethnicities, and three as Native American. The median experiment duration was 14 min and 33 s and participants were compensated $0.50 for participating plus a bonus of up to $2.00 (i.e., $0.20 per correct answer). Informed consent was administered by the IRB of the University of Chicago.

Creating and pretesting visually disordered and visually ordered stimuli. We exploited what we learned in Experiments 1–3 by creating a 2 (symmetry vs. asymmetry) × 2 (visually ordered edges vs. visually disordered edges) set of stimuli using the 50 most orderly and 50 most disorderly scrambled-edge stimuli from Experiment 2. Although asymmetry was a weaker predictor of disorder judgments than was nonstraight edge density in Experiment 1, this may have been due to the absence of high symmetry in our sample of scenes. Like we did for colors in Experiment 3, we conducted an experiment in which we manipulated (a)symmetry to test its effect on disorder judgments because symmetry is intuitively close to the definition of order. This experiment is reported as a pretest for Experiment 4 below. The results indicate that the effect of symmetry on disorder judgments is not negligible, and may be underestimated in Experiment 1. Symmetry is also relevant to the fluency mechanism because the redundant information in a symmetrical pattern should make it easier to process (Kinchla, 1977; Reber, Schwarz, & Winkielman, 2004).

The (a)symmetrical stimuli were created based on the same method we described in Experiment 2 (see Figure S1 in supplemental materials). This method starts with creating the edge maps (Process 1, same as Process 1 in Figure 3). Next, we created two random masks of patches each having, on average, half a surface of 1s and half a surface of 0s (Process 2, same as Process 3 in Figure 3). The edge map was then multiplied (dot product) with each of the two masks (Process 3, same as Process 4 in Figure 3) and the resulting pictures were overlaid, one being flipped on the *x*-axis before being overlaid on the other (Process 4). The important manipulation here is that if we begin the whole process with identical 0s and 1s matrices, we will get identical patches (outputs of Process 2), resulting in the two identical stimuli being outputted by Process 3. Hence, the flip and overlay (Process 4) will result in a symmetrical stimulus with on average, the same amount of edges and similar edge types as the original image. On the other hand, if we start with two different random 0s and 1s matrices, we get different patches, resulting in the two different stimuli being outputted by Process 3. Hence, the flip and overlay (Process 4) will result in an asymmetrical stimulus with, on average, the same amount of edges and similar edge types as the original image. This method of producing (a)symmetrical stimuli was applied to both the 50 most orderly and 50 most scrambled-edge stimuli from Experiment 2 to create the full 2 (symmetry vs. asymmetry) × 2 (visually ordered edges vs. visually disordered edges) set of stimuli for pretesting.

Pretesting. The purpose of this pretest was to test whether visually disordered edges and asymmetry in the new stimuli would independently increase disorder judgments. In this pretest, we conducted a 2 (symmetry vs. asymmetry) × 2 (visually ordered edges vs. visually disordered edges) within-subjects experiment on AMT with 222 participants. Participants were randomly presented 40 images total (10 from each cell; see Figure S2 in supplemental materials for examples). They rated each image in terms of disorder as in the Experiments 1–3.

Manipulations of asymmetry and disorderly edges were dummy coded. A multiple linear regression model with disorder ratings regressed on asymmetry, disorder, and their interaction ($R^2_{\text{adj}} = .854$) revealed that asymmetry, B $= 1.28$, $SE = .064$, $p < .001$, $\eta^2_{\text{p}} = .437$, and disorderly edges, B $= 0.79$, $SE = .064$, $p < .001$, $\eta^2_{\text{p}} = .671$, independently increased disorder ratings (see Figure S3 in supplemental materials), as predicted. Their interaction was not significant, B $= 0.06$, $SE = .090$, $p = .520$, $\eta^2_{\text{p}} = .002$.

Manipulating visual disorder and visual order. The 30 most (dis)orderly stimuli from the pretest were used for our manipulation of visual disorder versus visual order in Experiments 4 and 5 (see Figure 5 for examples). We also used the next five most (dis)orderly stimuli from the pretest for the filler task in Experiments 4 and 5.

Procedure. Participants first received a consent document which disguised the purpose of the experiment by describing it as about the "interplay between visual perception and cognitive performance." Participants then were given a brief introduction to the IRT as in Experiments 1–3. Next, they performed a filler task in which they were presented the 10 filler stimuli for 10 s each (1 min and 40 s of total exposure) and were asked to rate disorder the same way as in the Experiments 1–3. The filler task had two purposes: (a) getting participants acquainted with the IRT before implementing the manipulation, and (b) masking the purpose of the study by displaying images both before testing and before self-grading. In the next part of the study, we adapted the procedure that Mazar et al. (2008) developed to study cheating behavior. This procedure involves taking an incentivized test, then grading oneself on this test, which gives participants the opportunity to cheat. First, in the test phase, participants attempted a task in which they were given two minutes to search for pairs of numbers that add to 10 within 4 × 3 matrices composed of numbers between 0 and 10 with two decimal digits ("Matrices Test"). There were 10 matrices with each containing one solution. Participants were told that they would receive a \$0.20 bonus (~8 min of work at the median reservation wage on AMT, Horton, Rand, & Zeckhauser, 2011) for each matrix they solved correctly. After 2 min, they were automatically taken to the next part of the study in which we implemented our manipulation. Participants were randomly assigned to view and rate either the 30 visually disordered stimuli or the 30 visually ordered stimuli in terms of disorder. We added the simple rating task to keep participants engaged. These images (all 4:3 ratio) were presented on a plain white background in a 720 × 540 pixel frame. Each image was presented for 10 s (5 min total exposure). Next, they moved on to the self-grading phase of the experiment. Participants were then instructed to grade themselves on the Matrices Test they had performed earlier. They were reminded that they would be paid \$0.20 for each question that they had solved correctly, and that we would take their word for it (i.e., participants could report getting more correct than they actually did). Each matrix from earlier was presented with the correct solution marked and their answers from before were presented just below each matrix, thus no recall was involved. For each matrix, they were asked to simply respond "Yes" or "No" to the prompt, "Did you get it right?" (see screenshot in supplemental materials). After grading themselves, all participants completed the state PANAS Scale (Watson, Clark, & Tellegen, 1988) and a demographics survey before being debriefed. Statistically adjusting for state positive and negative affect did not change the pattern of results in Experiments 4 and 5, so it is not discussed further.

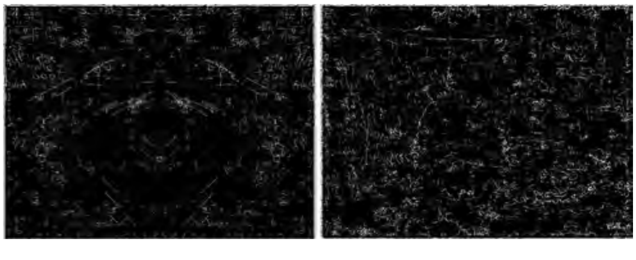

Figure 5. Examples of visual order (left) and visual disorder (right) stimuli. These were constructed based on the results of Experiments 1–3, and used in Experiment 4 and 5.

Results

Manipulation check. An independent-samples t test confirmed that the manipulation had a significant effect on disorder ratings, $t(403) = 8.17$, $p < .001$, $d = 0.82$, with the visually disordered stimuli ($M = 5.20$, $SD = 0.94$) receiving higher disorder ratings than the visually ordered stimuli ($M = 4.53$, $SD = 0.67$).

Cheating analysis: Actual performance versus reported performance. First we examined actual performance versus reported performance in the visual-order versus visual-disorder condition. Six participants (1.5% of the sample) were excluded from the cheating analysis for performing perfectly on the Matrices Test because it would be impossible for them to cheat. Actual performance and reported performance were imperfectly correlated at $r = .54$ indicating that the procedure encouraged people to cheat. We utilized the *lme4* package in R to conduct a linear mixed-effects model with performance on the Matrices Test predicted by visual condition, actual versus reported, and their interaction as fixed factors and a random intercept for each participant. Degrees of freedom was estimated with Satterthwaite's approximation. This model revealed a significant main effect of actual versus reported, $t(399.00) = 11.10$, $p < .001$, with participants across visual conditions reporting 55% higher performance ($M = 4.60$, $SD = 3.27$) than their actual performance ($M = 2.97$, $SD = 2.14$) on the Matrices Test. Importantly, there was a significant interaction between actual versus reported and visual condition, $t(399.00) = 3.59$, $p < .001$, with participants in the visual-disorder condition reporting 70% higher performance than their actual performance and participants in the visual-order condition reporting 39% higher performance than their actual performance (see Figure 6a). The simple effect of visual condition within reported performance was also significant, $t(639.53) = 4.13$, $p < .001$. A follow-up test of multivariate simple effects of actual versus reported performance within the visual-disorder and visual-order conditions revealed that the effect size in the visual-disorder condition, $\eta^2_{\text{p}} = .232$, was nearly three times larger than the effect size

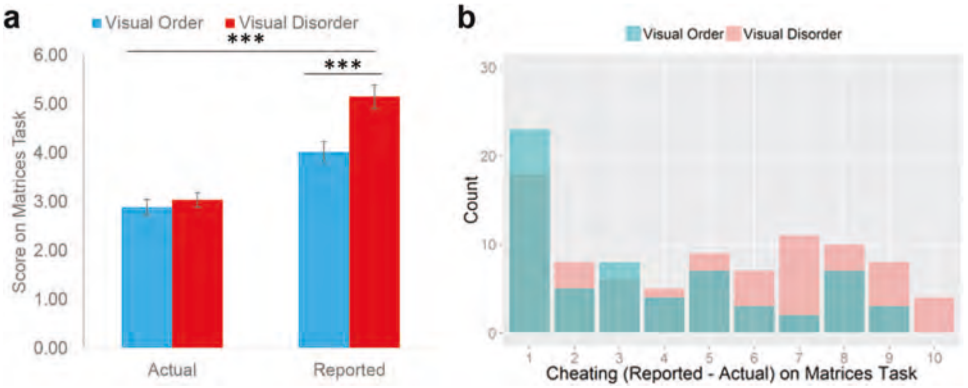

Figure 6. (a) Actual versus reported performance by condition in Experiment 4. Error bars indicate mean ± SEM. (b) Overlapping bar chart presenting magnitude of cheating (reported performance—actual performance) by visual condition in Experiment 4 (darker color indicates overlap). *** $p < .001$. See the online article for the color version of this figure.

in the visual-order condition, $\eta_p^2 = .078$. These results indicate that those in the visual-disorder condition cheated more than those in the visual-order condition.

Cheating analysis: Likelihood of cheating. Second, we tested whether the likelihood of cheating differed between the visual order and visual disorder conditions. A chi-squared test of independence conducted on a condition-by-cheating (yes/no) contingency table was significant, $\chi^2(1, N = 399) = 5.27$, $p = .022$, $\varphi = 0.115$, OR = 1.62, with 43% of participants cheating in the visual-disorder condition and 32% of participants cheating in the visual-order condition (35% relative increase, adjusted residual = 2.30).

Cheating analysis: Magnitude of cheating. Third, we compared the visual-order group and the visual disorder group on a measure of absolute cheating magnitude (reported performance—actual performance). An independent samples t test revealed a significant effect of visual disorder on magnitude of cheating, $t(397) = 3.60$, $p < .001$, $d = 0.36$, with those in the visual-disorder condition ($M = 2.12$, $SD = 2.12$) cheating by 87% larger relative magnitude than those in the visual-order condition ($M = 1.13$, $SD = 2.26$; see Figure 6b).[4]

Prior to this experiment, we conducted an identical experiment with reduced cheating incentives ($0.10 per correct matrix rather than $0.20), which yielded results trending in the predicted direction but not reaching significance (see supplementary materials for results). However, a meta-analysis of the combined data from this preliminary experiment and Experiment 4 yielded significant effects of the manipulation of visual disorder on both cheating likelihood, $\chi^2(1, N = 800) = 8.04$, $p = .005$, $\varphi = 0.100$, OR = 1.51, and cheating magnitude, $t(798) = 2.78$, $p = .006$, $d = 0.20$, with visual disorder increasing relative cheating likelihood by 29% on average and relative cheating magnitude by 41% on average.

Experiment 5: The Effect of Visual Disorder on Rule-Breaking Part 2

One concern with Experiment 4 was that having people rate disorder may have driven the observed cheating effects, perhaps by inadvertently causing them to think about scene-level social disorder cues. Alternatively, we may have merely increase the salience of basic visual disorder cues as was our intention. In case of the former, we conducted Experiment 5, which was identical to Experiment 4, except that we had people rate *preference* instead of disorder during both the training phase and the manipulation phase. Thus, there was not a single explicit mention of "order" or "disorder" in this experiment. This should alleviate any concern about having people rate (dis)order in Experiment 4. Because rating preference presumably would reduce the salience of visual disorder, we expected the effect of visual disorder on cheating to be attenuated compared to in Experiment 4.

Method

Participants and design. Three-hundred and 94 U.S.-based adults (202 men, 189 women, three other) were recruited from AMT and participated in this two-condition (visually disordered stimuli vs. visually ordered stimuli) between-subjects experiments. Sample size and stopping rule were the same as in Experiment 4. Ages ranged from 19 to 76 ($M = 34.25$, $SD = 11.07$). Three-hundred and 27 participants identified primarily as White/Caucasian, 27 as Asian/Asian American, 20 as Black/African American, nine as Hispanic/Latino, four as multiple ethnicities, three as Native Hawaiian, and two as Native American. The median experiment duration was 13 min and 35 s and participants were compensated $0.50 for participating plus a bonus of up to $2.00 (same as in Experiment 4). Informed consent was administered by the IRB of the University of Chicago.

Results

Preference ratings. An independent-samples t test revealed that the visual disorder manipulation did not have a significant

---

[4] These magnitude estimates take into account noncheaters in both conditions. If we estimate magnitude among only the cheaters, those in the visual-disorder condition ($n = 86$, $M = 4.98$, $SD = 2.99$) cheated by 38% larger relative magnitude than those in the visual-order condition ($n = 62$, $M = 3.60$, $SD = 2.71$).

effect on aesthetic preference ratings, $t(392) = 0.13$, $p = .897$, with the visually disordered stimuli ($M = 3.31$, $SD = 0.90$) receiving virtually the same aesthetic preference ratings on average as the visually ordered stimuli ($M = 3.32$, $SD = 1.09$). Although this result seems inconsistent with the disfluency mechanism suggested in the introduction (see Reber et al., 2004), we note that in our other work, we have found that disorder judgments inversely correlated with aesthetic preference at $r = -.64$, $p < .001$ (Kotabe, Kardan, & Berman, 2016b). In this work, for the visually disordered-stimuli, disorder judgments from Experiment 4 inversely correlated with aesthetic preference ratings at $r = -.70$, $p < .001$ and for the visually ordered stimuli, at $r = -.37$, $p = .047$. These inverse correlations are consistent with a disfluency mechanism.

Cheating analysis: Actual performance versus reported performance. Cheating was assessed the same way as in Experiment 4. Five participants (1.3% of the sample) were excluded from the cheating analysis for performing perfectly on the Matrices Test because it would be impossible for them to cheat. Actual performance and reported performance were imperfectly correlated at $r = .52$ indicating that the procedure encouraged people to cheat. As in Experiment 4, we conducted a linear mixed-effects model with performance on the Matrices Test predicted by visual condition, actual versus reported, and their interaction as fixed factors and a random intercept for each participant. This model revealed a significant main effect of actual versus reported, $t(394.00) = 11.10$, $p < .001$, with participants across visual conditions reporting 53% higher performance ($M = 4.54$, $SD = 3.26$) than their actual performance ($M = 2.97$, $SD = 2.06$) on the Matrices Test. Importantly, there was again a significant interaction between actual versus reported and visual condition, $t(394.00) = 2.08$, $p = .038$, with participants in the visual-disorder condition reporting 63% higher performance than their actual performance and participants in the visual-order condition reporting 43% higher performance than their actual performance (see Figure 7a). The simple effect of visual condition within reported performance was also significant, $t(627.46) = 2.29$, $p = .023$. A follow-up test of multivariate simple effects of actual versus reported performance within the visual-disorder and visual-order conditions revealed that the effect size in the visual-disorder condition, $\eta_p^2 = .183$, was nearly twice as large as the effect size in the visual-order condition, $\eta_p^2 = .094$. These results corroborate that visual disorder encourages cheating, and this effect is not due to rating disorder.

Cheating analysis: Likelihood of cheating. A chi-square test of independence conducted on a condition-by-cheating (yes/no) contingency table was not significant, $\chi^2(1, N = 389) = 1.29$, $p = .257$, $\varphi = 0.058$, OR $= 1.27$, however, there was a descriptive difference in the predicted direction, with 37% of participants cheating in the visual-disorder condition and 32% of participants cheating in the visual-order condition (17% relative increase). To compare this result with that observed in Experiment 4, we took the difference of the natural logarithm of the ORs (odds ratios), δ, and calculated SE of δ with $\sqrt{SE(\ln(\mathrm{OR}_1))^2 + SE(\ln(\mathrm{OR}_2))^2}$. We then obtained z with $\delta/SE(\delta)$. The result of the chi-squared test in Experiment 5 did not significantly differ from the result of the chi-squared test in Experiment 4, $\delta = -0.24$, $z = -0.79$, $p = .428$.

Cheating analysis: Magnitude of cheating. An independent samples *t* test revealed a significant effect of visual disorder on magnitude of cheating as in Experiment 4, $t(387) = 2.08$, $p = .038$, $d = 0.21$, with those in the visual-disorder group ($M = 1.86$, $SD = 3.00$) cheating by 46% larger relative magnitude than those in the visual-order condition ($M = 1.27$, $SD = 2.57$; see Figure 7b).[5] To compare this result with that observed in Experiment 4, we first obtained *r*s from each *t* test, then compared *r*s with the *r.test* function in R which compares Fisher *r*-to-*z* transformed correlations, which revealed that the result of the *t* test in Experiment 5 did not significantly differ from the result of the *t* test in Experiment 4, $z = -1.04$, $p = .297$.

Considering the results from Experiments 4 and 5 together, we conclude that visual disorder is indeed sufficient to encourage rule-breaking behavior. When cheating incentives were sufficiently large and visual disorder was salient (Experiment 4), the effect of visual disorder on cheating was largest. When the salience of visual disorder was reduced (Experiment 5), the effect of visual disorder on cheating was still marked but weaker. We note that the effect of visual disorder on both cheating likelihood and cheating magnitude did not significantly differ between Experiments 4 and 5, thus differences in effect size may be entirely due to chance. If the effect in Experiment 5 were actually weaker, it would suggest that, although basic visual disorder cues alone may encourage rule-breaking, there could be some top-down processes at work. However, this is not complex social reasoning of the kind put forward by BWT researchers, rather it may have to do with priming visual disorder and its associations.

General Discussion

This study set out to answer two major questions. First, what are some of the key basic visual features that define visual disorder? Second, are these basic visual disorder cues alone sufficient to encourage rule-breaking despite the absence of scene-level social disorder cues and complex social reasoning? Our first set of experiments (Experiments 1–3) and our pretest for Experiments 4 and 5 showed that nonstraight edge density and asymmetry are key components of visual disorder. More generally, these experiments suggest that spatial features are more important than color features for visual disorder. Such insights into the building blocks of visual disorder are important if we are to make significant advancements in our understanding of phenomena relevant to BWT and more broadly how visual processing can affect complex human behavior. Our second set of experiments (Experiments 4 and 5) demonstrated that exposure to basic visual disorder cues alone is sufficient to encourage rule-breaking behavior. One broad implication of this work is that established theories of rule-breaking that assume that complex reasoning about scene-level social disorder cues fully mediates BWT phenomena, should be reconsidered (e.g., Kelling & Coles, 1997; Sampson & Raudenbush, 2004).

To elaborate on Experiments 1–3, the stronger effect of spatial features than color features on disorder judgments may reflect that

---

[5] These magnitude estimates take into account noncheaters in both conditions. If we estimate magnitude among only the cheaters, those in the visual-disorder condition ($n = 73$, $M = 4.97$, $SD = 2.92$) cheated by 25% larger relative magnitude than those in the visual-order condition ($n = 62$, $M = 3.98$, $SD = 3.15$).

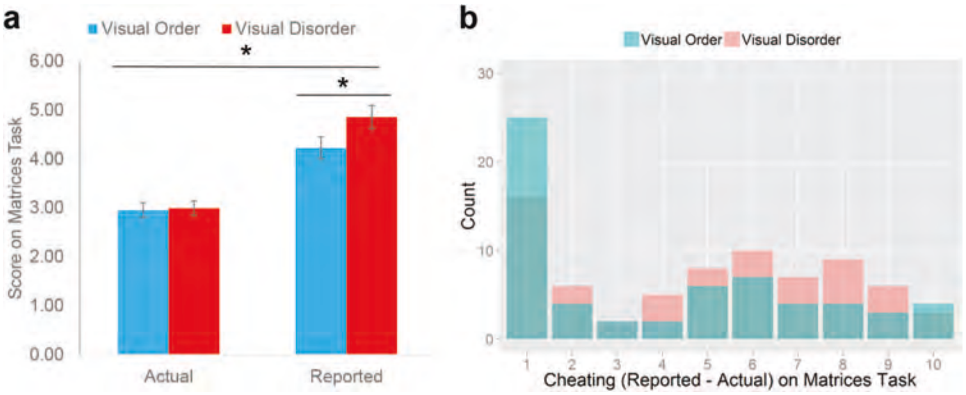

Figure 7. (a) Actual versus reported performance by condition in Experiment 5. Error bars indicate mean ± SEM. (b) Overlapping bar chart presenting magnitude of cheating (reported performance—actual performance) by visual condition in Experiments 6 (darker color indicates overlap). * $p < .05$. See the online article for the color version of this figure.

the concept of "disorder," both in its physical and social senses, is represented spatially in the mind. The first definition of order in the Oxford Dictionary is "the arrangement or disposition of *people* or *things* in relation to each other according to a particular sequence, pattern, or method" [emphasis added]. Subdefinitions of this primary definition refer to both physical and social forms of disorder. This blurring of boundaries between physical and social disorder is further evidenced by the fact that people use spatial metaphors to refer to deontic concepts. Such evidence leads us to speculate that the learned concept of "disorder" is intimately linked to spatial thinking.

Although Experiments 1–3 start to answer the broader question of what spatial features factor into visual disorder, they surely do not exhaustively answer this question. It would be interesting to model other possible basic visual disorder cues such as variability in edge orientation, or the depth perception cues. In particular, although curved edges were seen as disorderly in our experiments, intuition says that several curved edges arranged in parallel to one another would be more orderly than curved edges arranged in a haphazard way. Our results linking symmetry and order support this intuition but not directly. In addition, although colors were *relatively* unimportant compared with edges for visual disorder in our experiments, there was some variance in disorder ratings for the scrambled-color stimuli, which suggests that colors are not *absolutely* unimportant (after all, colors can have a sequence or pattern). Perhaps low saturation increases perceptual disorder, as suggested by the regression results of Experiment 1. We encourage researchers to derive other spatial and color metrics and stimuli to attempt to quantify visual disorder.

To elaborate on Experiments 4 and 5, the results suggest that the "cheating effect" of visual disorder actually has two aspects. Visual disorder not only increased the amount by which cheaters cheat (*cheating magnitude*), it also encouraged people who normally would not cheat to cheat (*cheating likelihood*). That is, cheaters were nudged toward cheating more and noncheaters were nudged toward cheating at all. A meta-analysis of the combined data from Experiment 4 and 5 yielded significant effects of the manipulation of visual disorder on both cheating likelihood, $\chi^2(1, N = 788) = 5.95$, $p = .015$, $\varphi = 0.087$, OR = 1.44, and cheating magnitude, $t(786) = 4.01$, $p < .001$, $d = 0.29$, with visual disorder increasing relative cheating likelihood by 26% on average and relative cheating magnitude by 66% on average. This sort of cheating could have major economic and societal consequences. Imagine if the amount by which people underreported their taxes increased by just 1%—billions of dollars would be lost.

One of the big questions remaining is, how exactly is this happening? We speculate about two classes of mechanisms—the first reflecting an information processing perspective and the second reflecting a priming or spreading activation perspective. From the first perspective, the visually disordered stimuli were less redundant (i.e., fewer spatially predictable patterns) and conveyed more information than the visually ordered stimuli. These aspects of visual disorder could make viewing visually disordered stimuli more cognitively demanding than viewing visually ordered stimuli (see Field, 1987; Kinchla, 1977; Olshausen & Field, 1996; Witkin & Tenenbaum, 1983), which could have two consequences leading to the same outcome—decreased self-regulatory capacity. First, cognitive demand could directly fatigue cognitive resources involved in self-regulation (Hofmann et al., 2011; Kaplan & Berman, 2010; Kotabe & Hofmann, 2015). Second, cognitive demand could lead to the sense of perceptual disfluency, which in turn may be used as a metacognitive cue in judgment. Assuming disfluency activates effortful thinking (Alter et al., 2007) and is generally interpreted in negative ways (Alter & Oppenheimer, 2009), this could lead to the effortful regulation of negative thoughts and feelings which would further tax self-regulatory resources (Kotabe & Hofmann, 2015; Muraven & Baumeister, 2000). Holding all else constant, one would become more likely to enact tempting rule-breaking behaviors such as cheating for money (Kotabe & Hofmann, 2015).

From the second perspective, prolonged exposure to visual disorder may activate a mindset that things are random and uncontrollable, which may reduce the motivation to self-control (Kotabe, 2014; see also Tullett et al., 2015). Similarly, it may activate mental metaphors, which are manifested in a family of linguistic metaphors linking spatial features to deontic concepts such as in "the straight-edge lifestyle" and "the crooked politician" (Casasanto & Bottini, 2014). Consistent with this perspective, there is a rich literature on feedforward and feedback projections from basic visual cortical areas to various higher cortical areas involved in complex semantic processing and behavior (Felleman & Van Essen, 1991; Gilbert & Li, 2013; Lamme & Roelfsema, 2000; McIntosh et al., 1994). If basic spatial features of visual disorder are endowed with such semantic content, there could be downstream effects on conceptually related complex behaviors such as rule-breaking (Bargh, 2006; Molden, 2014). It is possible that both information-processing and priming/spreading-activation mechanisms are at work in producing the cheating effects observed and could interact with each other. Regardless, these possible mechanisms paint a completely different picture from current explanations for BWT phenomena. Thus, they point to a vast and unattended area of research, which we encourage researchers to venture into.

Another important remaining question concerns what we mean, specifically, when we say that visual disorder "encourages" cheating, or in other words, increases the likelihood of enacting a temptation to cheat. According to integrative self-control theory (Kotabe & Hofmann, 2015), there would be several components at work in a rule-breaking situation involving a desire that conflicts with a proscriptive norm. Each component could be involved in increasing the likelihood of enacting the tempting rule-breaking behavior. For example, exposure to visual disorder could fatigue cognitive-control capacity via information "overload," it could decrease cognitive-control motivation via activating a certain mindset, or it could fuel desire via some priming or spreading activation process. And to complicate things, these are likely not independent processes. For example, increasing desire strength may activate or inhibit control goals (Fishbach, Friedman, & Kruglanski, 2003) and fatiguing control capacity may increase desire strength (Vohs et al., 2013; Wagner, Altman, Boswell, Kelley, & Heatherton, 2013). Teasing apart these processes is a major challenge (Kotabe & Hofmann, in press), and thus much work remains.

To conclude, research on environmental disorder has tended to focus on its consequences (e.g., Braga & Bond, 2008; Keizer et al., 2008; Kelling & Coles, 1997), yet little is known about what makes an environment disorderly in the first place. As this work demonstrates, basic visual features shape judgments of disorder, and spatial features associated with deontic concepts and processing difficulty may play a particularly important role. Deconstructing disorder and orderly environments into their lower and higher level features may help us understand how disorderly environments affect us in ways harmful to ourselves and to society. In addition, taking this approach may inform the design of both real and virtual environments. Considering the observed effect of visual disorder on rule-breaking behavior, and the evidence that rule-breaking behaviors spread (Keizer et al., 2008), we should take (imparting) visual disorder in our environments seriously.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235. http://dx.doi.org/10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136,* 569–576. http://dx.doi.org/10.1037/0096-3445.136.4.569

Ariely, D. (2013). *The honest truth about dishonesty: How we lie to everyone–especially ourselves* (Reprint ed.). New York, NY: Harper Perennial.

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36,* 147–168. http://dx.doi.org/10.1002/ejsp.336

Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software, 31,* 1–21. http://dx.doi.org/10.18637/jss.v031.i10

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., . . . Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLoS ONE, 9,* e114572. http://dx.doi.org/10.1371/journal.pone.0114572

Berman, M. G., Jonides, J., & Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science, 19,* 1207–1212. http://dx.doi.org/10.1111/j.1467-9280.2008.02225.x

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–117. http://dx.doi.org/10.1037/0033-295X.94.2.115

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology, 46,* 577–607. http://dx.doi.org/10.1111/j.1745-9125.2008.00124.x

Braga, A. A., Weisburd, D. L., Waring, E. J., Mazerolle, L. G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology, 37,* 541–580. http://dx.doi.org/10.1111/j.1745-9125.1999.tb00496.x

Brascamp, J., Blake, R., & Knapen, T. (2015). Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nature Neuroscience, 18,* 1672–1678. http://dx.doi.org/10.1038/nn.4130

Casasanto, D., & Bottini, R. (2014). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews: Cognitive Science, 5,* 139–149. http://dx.doi.org/10.1002/wcs.1271

Chae, B. G., & Zhu, R. J. (2014). Environmental disorder leads to self-regulatory failure. *The Journal of Consumer Research, 40,* 1203–1218. http://dx.doi.org/10.1086/674547

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290. http://dx.doi.org/10.1037/1040-3590.6.4.284

Cutrona, C. E., Russell, D. W., Hessling, R. M., Brown, P. A., & Murry, V. (2000). Direct and moderating effects of community context on the psychological well-being of African American women. *Journal of Personality and Social Psychology, 79,* 1088–1101. http://dx.doi.org/10.1037/0022-3514.79.6.1088

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47. http://dx.doi.org/10.1093/cercor/1.1.1

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4,* 2379–2394. http://dx.doi.org/10.1364/JOSAA.4.002379

Fishbach, A., Friedman, R. S., & Kruglanski, A. W. (2003). Leading us not unto temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology, 84,* 296–309. http://dx.doi.org/10.1037/0022-3514.84.2.296

Geis, K. J., & Ross, C. E. (1998). A new look at urban alienation: The effect of neighborhood disorder on perceived powerlessness. *Social Psychology Quarterly, 61,* 232–246. http://dx.doi.org/10.2307/2787110

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59,* 167–192. http://dx.doi.org/10.1146/annurev.psych.58.110405.085632

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14,* 350–363. http://dx.doi.org/10.1038/nrn3476

Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes, 115,* 191–203. http://dx.doi.org/10.1016/j.obhdp.2011.03.001

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Palo Alto, CA: Stanford University Press. Retrieved from http://psycnet.apa.org/PsycINFO/1990-97753-000

Harcourt, B. E. (2009). *Illusion of order: The false promise of broken windows policing*. Cambridge, MA: Harvard University Press.

Hofmann, W., Friese, M., Schmeichel, B. J., & Baddeley, A. D. (2011). Working memory and self-regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (Vol. 2, pp. 204–225). New York, NY: Guilford Press.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14,* 399–425. http://dx.doi.org/10.1007/s10683-011-9273-9

Kaplan, S., & Berman, M. G. (2010). Directed attention as a common resource for executive functioning and self-regulation. *Perspectives on Psychological Science, 5,* 43–57. http://dx.doi.org/10.1177/1745691609356784

Kardan, O., Demiralp, E., Hout, M. C., Hunter, M. R., Karimi, H., Hanayik, T., . . . Berman, M. G. (2015). Is the preference of natural versus man-made scenes driven by bottom-up processing of the visual features of nature? *Frontiers in Psychology, 6,* 471. http://dx.doi.org/10.3389/fpsyg.2015.00471

Kardan, O., Henderson, J. M., Yourganov, G., & Berman, M. G. (2016). Observers' cognitive states modulate how visual inputs relate to gaze control. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1429–1442. http://dx.doi.org/10.1037/xhp0000224

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322,* 1681–1685. http://dx.doi.org/10.1126/science.1161405

Kelling, G. L., & Coles, C. M. (1997). *Fixing broken windows: Restoring order and reducing crime in our communities*. New York, NY: Touchstone.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Perception & Psychophysics, 22,* 19–30. http://dx.doi.org/10.3758/BF03206076

Kotabe, H. P. (2014). The world is random: A cognitive perspective on perceived disorder. *Frontiers in Psychology, 5,* 606. http://dx.doi.org/10.3389/fpsyg.2014.00606

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science, 10,* 618–638. http://dx.doi.org/10.1177/1745691615593382

Kotabe, H. P., & Hofmann, W. (in press). How depletion operates in an integrative theory of self-control. *Self-Regulation and Ego Control*. http://dx.doi.org/10.1016/B978-0-12-801850-7.00019-6

Kotabe, H. P., Kardan, O., & Berman, M. G. (2016a). *Can the high-level semantics of a scene be preserved in the low-level visual features of that scene?* Manuscript submitted for publication.

Kotabe, H. P., Kardan, O., & Berman, M. G. (2016b). *The nature-disorder paradox: How nature scenes are aesthetically pleasing despite being disorderly*. Manuscript submitted for publication.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579. http://dx.doi.org/10.1016/S0166-2236(00)01657-X

Linares, L. O., Heeren, T., Bronfman, E., Zuckerman, B., Augustyn, M., & Tronick, E. (2001). A mediational model for the impact of exposure to community violence on early child behavior problems. *Child Development, 72,* 639–652. http://dx.doi.org/10.1111/1467-8624.00302

Mahoney, J. V. (1987). *Image chunking: Defining spatial building blocks for scene analysis*. Fort Belvoir, VA: DTIC.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 275,* 483–519. http://dx.doi.org/10.1098/rstb.1976.0090

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 207,* 187–217. http://dx.doi.org/10.1098/rspb.1980.0020

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45,* 633–644. http://dx.doi.org/10.1509/jmkr.45.6.633

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. *The Journal of Neuroscience, 14,* 655–666.

Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? An overview and understanding. *Social Cognition, 32,* 243–249.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126,* 247–259. http://dx.doi.org/10.1037/0033-2909.126.2.247

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36. http://dx.doi.org/10.1016/S0079-6123(06)55002-2

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607–609. http://dx.doi.org/10.1038/381607a0

Perkins, D. D., & Taylor, R. B. (1996). Ecological assessments of community disorder: Their relationship to fear of crime and theoretical implications. *American Journal of Community Psychology, 24,* 63–107. http://dx.doi.org/10.1007/BF02511883

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8,* 364–382. http://dx.doi.org/10.1207/s15327957pspr0804_3

Ross, C. E. (2000). Neighborhood disadvantage and adult depression. *Journal of Health and Social Behavior, 41,* 177–187. http://dx.doi.org/10.2307/2676304

Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of "broken windows." *Social Psychology Quarterly, 67,* 319–342. http://dx.doi.org/10.1177/019027250406700401

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348. http://dx.doi.org/10.1207/s15327663jcp1404_2

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23,* 1264–1270. http://dx.doi.org/10.1177/0956797612443835

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Tullett, A. M., Kay, A. C., & Inzlicht, M. (2015). Randomness increases self-reported anxiety and neurophysiological correlates of performance monitoring. *Social Cognitive and Affective Neuroscience, 10,* 628–635. http://dx.doi.org/10.1093/scan/nsu097

Vohs, K. D., Baumeister, R. F., Mead, N. L., Hofmann, W., Ramanathan, S., & Schmeichel, B. J. (2013). *Engaging in self-control heightens urges and feelings*. Manuscript submitted for publication.

Vohs, K. D., Redden, J. P., & Rahinel, R. (2013). Physical order produces healthy choices, generosity, and conventionality, whereas disorder produces creativity. *Psychological Science, 24,* 1860–1867. http://dx.doi.org/10.1177/0956797613480186

Wagner, D. D., Altman, M., Boswell, R. G., Kelley, W. M., & Heatherton, T. F. (2013). Self-regulatory depletion enhances neural responses to rewards and impairs top-down control. *Psychological Science, 24,* 2262–2271. http://dx.doi.org/10.1177/0956797613492985

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience, 29,* 10573–10581. http://dx.doi.org/10.1523/JNEUROSCI.0559-09.2009

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society Series B, Methodological, 21,* 396–399.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic Monthly, 249,* 29–38.

Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. *Human and Machine Vision, 1983,* 481–543. http://dx.doi.org/10.1016/B978-0-12-084320-6.50022-0

Received December 23, 2015
Revision received August 30, 2016
Accepted September 7, 2016 ■