

Learn Behav (2012) 40:367–379  
DOI 10.3758/s13420-011-0058-y

# Varied but not necessarily random: Human performance under variability contingencies is affected by instructions

Alessandra S. Souza · Thaissa Neves R. Pontes ·  
Josele Abreu-Rodrigues

Published online: 4 January 2012  
© Psychonomic Society, Inc. 2011

**Abstract** The goal of the present study was to evaluate the role of verbal stimuli in the production of response variability in humans. College students were distributed into three groups and asked to type three-digit sequences. Participants in the *systematic group* were instructed to produce sequences according to a rule of their choice; those in the *random group* were instructed to produce sequences according to chance; and those in the *control group* were not instructed about how to produce sequences. The experiment employed an ABA design. During the A phases, low-frequent sequences were reinforced (variability contingency), whereas during the B phase, reinforcement was withdrawn (extinction). The results indicated the following: (1) The instructions were efficient at producing systematic and random-like patterns for the systematic and random groups, respectively; in the absence of instructions, a mix of both patterns was observed. (2) Behavior was sensitive to extinction independently of the instructions provided. (3) Systematic patterns favored a more equiprobable distribution of sequences across trials. (4) Reaction times were longer for responding in a systematic than in a random-like fashion. The present findings suggest that

individual differences in meeting variability contingencies may be due, at least partially, to instructional control.

**Keywords** Humans · Variability · Instruction · Systematic behavior · Random behavior · Operant conditioning

The behavior of humans and of nonhuman animals has been demonstrated to be sensitive to operant contingencies that specify different levels of variability, such that variable or repetitive behaviors develop at demand (see Neuringer, 2002, 2004, for reviews). For example, Machado (1989) exposed pigeons to a sequence formation task in which a series of four pecks distributed across two keys were followed by food if they met a given variability criterion. Low, medium, and high variability requirements produced corresponding levels of behavioral variability. A similar pattern was observed with college students when the task was to move a square from the top to the bottom of a pyramid: Path variability increased as a function of the variability requirement (Stokes, 1999). Additionally, other studies have shown that behavioral variability comes under discriminative control both by antecedent (Denney & Neuringer, 1998; Page & Neuringer, 1985; Ward, Kynaston, Bailey, & Odum, 2008) and consequent (Souza & Abreu-Rodrigues, 2010) stimuli. These findings have given support to the notion that variability is an operant dimension of behavior, in the same sense as force, duration, frequency, and so forth.

Behavioral variability may be produced by means of differing behavioral patterns. While nonhuman animals have been shown to meet highly demanding variability contingencies by behaving randomly (Abreu-Rodrigues, Hanna, Cruz, Matos, & Delabrida, 2004; Cohen, Neuringer, & Rhodes, 1990; Machado, 1993; McElroy & Neuringer, 1990; Neuringer, 1991; Page & Neuringer, 1985; Ward, Bailey, & Odum,

---

A. S. Souza (✉)  
Department of Psychology, Cognitive Psychology Unit,  
University of Zurich,  
Binzmühlestrasse 14/22,  
CH-8050 Zürich, Switzerland  
e-mail: a.souza@psychologie.uzh.ch

A. S. Souza  
e-mail: alessandra.psi@gmail.com

T. N. R. Pontes · J. Abreu-Rodrigues  
Universidade de Brasília,  
Brasília, Brazil

J. Abreu-Rodrigues  
e-mail: abreu@unb.br

2006), some studies have pointed out that humans use systematic strategies to produce variation. For example, in a study by Schwartz (1982, Exp. 6), college students were asked to type sequences of eight keypresses that differed from the last 50 emitted sequences (a lag 50 requirement). Performance on this task was very accurate (~100%), and thus highly variable, and all participants reported using systematic rules to produce the sequences. According to Schwartz, the sequence emission patterns matched the self-reported rules (data not provided).

In a more recent experiment, Stokes and Harrison (2002, Exp. 1) exposed college students to two pyramids (i.e., a triangle composed of several squares). A white square was displayed at the top of each pyramid, and the task was to bring this square from the top to the bottom of the pyramid by pressing the left and right keys. This task was accomplished by pressing the keys five (small pyramid) or ten (large pyramid) times. Several possible paths (i.e., combinations of left and right responses) could be used to reach the endpoints on the pyramid, and the larger the pyramid, the larger the number of paths available. In four conditions, the variability requirement was increased from a lag of 0 to 2 to 10, and ultimately to lag 20. The greater the number of available paths and the higher the variability requirement, the higher the levels of sequence variation that were obtained. However, when the performance of the students was compared to that of a random generator, it was observed that the participants developed a systematic pattern with the larger pyramid: That is, they produced varied combinations of left and right presses by aiming at different endpoints, but varied the selected paths in a structured way. Furthermore, when requested to describe their performance, most participants reported using this strategy.

Maes (2003) also observed systematic responding under a variability contingency. College students were asked to type sequences of three digits and received feedback for their performance. Across two phases, different feedback contingencies were employed: variability-dependent reinforcement and no reinforcement (Exp. 1), and variability-dependent and variability-independent reinforcement (Exp. 2). Transitions between these contingencies were programmed according to an AB or a BA between-subjects design. Several measures of behavioral variability were computed: (1) the equiprobability in the emission of all sequence alternatives ( $U$  value), (2) the percentage of sequences that met the variability requirement (MetVar), (3) the sequence distribution, and (4) the autocorrelations. The first three measures allowed for the identification of variation at the level of response units, whereas autocorrelations indicated whether the current sequence was influenced by previous sequences (sequence dependency), and therefore pointed out the use of responding strategies.

Maes (2003) found that variability-dependent reinforcement engendered higher MetVar and  $U$  values and also a more even sequence distribution than did no reinforcement or

variability-independent reinforcement. This increase in variable responding, however, was also accompanied by an increase in the tendency to respond systematically or strategically, such that high autocorrelations were obtained. This effect, nevertheless, varied substantially from participant to participant (being present in about half of the sample) and from time to time (e.g., systematicity increased or decreased across blocks of trials). These systematic strategies and the alternative, more random-like patterns shown by other participants, however, were not correlated with differential rates of reinforcement, meaning that both patterns were effective in meeting the variability contingency. Between-subjects variability was attributed to either memory processes or self-rules. That is, Maes argued that systematicity requires memory of previous emitted sequences and thus is not always observed, because memory capacity varies across individuals. Further, some participants might have followed self-rules stating that the emission of sequences in a fixed order was relevant to the task, while others might have developed rules stating that random performance would be more appropriate.

The studies by Schwartz (1982), Stokes and Harrison (2002), and Maes (2003) demonstrated that under variability contingencies, humans tend to develop a systematic strategy in emitting sequences of responses, although a random-like pattern would also be functional. Moreover, Maes suggested that self-rules induced by the experimenter-given instructions might be one of the factors that alters the probabilities of adopting systematic or random strategies to satisfy a variability contingency. This possibility, however, has not been directly tested so far. Given these considerations, our primary concern was to evaluate the role of verbal stimuli (i.e., experimenter-provided instructions that might induce self-rules) in the production of response variability in humans.

Thus, college students were asked to type three-digit sequences. Variability-dependent feedback (Phase A) and no feedback (extinction; Phase B) were scheduled according to an ABA design. We provided three groups of participants with different instructions regarding the response strategy: The *systematic group* was told to emit sequences according to a rule of their choice; the *random group* was told to produce sequences according to chance; and the *control group* did not receive instructions regarding their sequence production. We were specifically interested in the effects of these instructions on the serial organization of the sequences across trials. We expected that participants told to produce sequences systematically and randomly would comply with these instructions, and therefore would show distinguishable response patterns, whereas the control group would display more between-subjects variability. To allow for the identification of systematicity, we computed autocorrelations and first-order differences—both measures allowing the observation of deviations from randomness—and we compared the students' data with those produced by a random

generator. We were also interested in evaluating whether these differential patterns of responding (i.e., systematic and random) would affect the probability of meeting the variability contingency (MetVar), the uncertainty in the emission of response sequences ( $U$  value)—traditional measures of response variability—and also the rate of responding. Previous studies (e.g., Maes, 2003) have not found a correlation between systematic and random response patterns and differential rates of reinforcement or uncertainty in the emission of sequences. We expected to replicate these findings. Regarding the response rate, there is evidence that faster responding produces more deviations from randomness—that is, more repetitions than would be expected from random responding (Baddeley, Emslie, Kolodny, & Duncan, 1998; Neuringer, 1991). Thus, we were interested in examining whether instructions to respond systematically or randomly would alter the rates at which participants typed the sequences.

## Method

### Participants

A group of 36 undergraduate students (25 women, 11 men) from the Universidade de Brasília (Brazil) participated in one 1-h session in return for extra credit in introductory psychology classes. The participants' ages ranged from 18 to 55 years old (mean = 22 years). Participants read and signed an informed consent form. Points earned during the task were converted into chances to win a cash prize (approximately US\$25) at the conclusion of the study.

### Procedure

The experimental task was programmed in Visual Basic 6. Participants were randomly assigned to one of three groups: systematic, random, and control. All participants read the following instructions at the beginning of the experiment. The instructions were written in Portuguese and translate into English as follows:

1. This is a learning experiment. Your task is to type sequences of three digits, using the keyboard keys numbered 1, 2, and 3. Each keypress will produce a yellow circle on the screen. This will allow you to keep track of the number of responses you have already made. When you finish typing a sequence, press the Enter button or the spacebar. There are 27 different possible sequences.
2. For each correct sequence, you will receive 10 points, and for each 100 points earned, you will receive a coupon that is worth a chance to win a prize at the end of the experiment. The computer will show the number

of sequences already typed and the number of coupons you have accumulated. Try to earn as many points as you can. You will be asked to type approximately 900 sequences; therefore, try to work at a steady pace.

For participants in the *systematic group*, the following instruction was added to the general instructions:

3. The best way to earn points is by emitting sequences according to some SYSTEMATICITY. Systematic responses are those that display order and regularity. Therefore, you should produce sequences that meet a certain rule of your choice. For example, you might decide to emit sequences grouped by their first number (e.g., 111, 123, 132, etc.). Again, emit sequences according to a predetermined order.

For participants in the *random group*, the following instruction was added to the general instructions:

3. The best way to earn points is by emitting sequences RANDOMLY. Random responses are those that happen by chance. Therefore, you should produce patterns of sequences without following any rule. For example, you might emit the following sequences (e.g., 123, 311, 311, 222, etc.). Again, do not emit sequences according to a predetermined order.

Participants in the *control group* received only the initial instructions.

In addition to the specific instructions, the systematic and random groups were required to complete a small task that ensured that they understood the concepts of systematicity and randomness adopted in this experiment. They were asked to classify the following patterns as systematic or random:

- a) 111, 112, 113, 121, 122, 123
- b) 313, 111, 321, 233, 122, 223
- c) 111, 222, 333, 111, 222, 333
- d) 232, 113, 321, 333, 231, 112
- e) 331, 332, 333, 221, 222, 223
- f) 131, 222, 123, 123, 223, 112

The correct responses were to identify (a), (c), and (e) as systematic patterns and the remaining patterns as random. The participants were also requested to give their own examples of systematic and random patterns. The experimenter provided feedback for all responses. After that, participants were allowed to start the experiment.

The participants sat in a small room containing a table, a chair, and an IBM computer. At the start of each trial, the monitor showed a black screen with the word “Sequence” at

the top and two counters at the bottom: One counter presented the number of trials already completed (on the left) and the other, the number of coupons accumulated (on the right). The task was to emit three-response sequences by pressing the digits 1, 2, and 3 on the keyboard. As the participant typed the digits, yellow circles were presented, from the left to the right, in a row below the word “Sequence.” If a sequence met the reinforcement criterion, pressing the Enter button or the spacebar changed the screen color to white. This screen showed feedback (“You won 10 points”), a smiley face, and the total number of earned points (hereafter called *reinforcers*) for 1 s. None of the reinforcers were produced for noncriterion sequences. Following the emission of each sequence (or the reinforcer’s delivery, when this was the case), a new trial began.

The experiment followed an ABA design. During the VAR (A) phases, a sequence produced the reinforcers if it satisfied two criteria: (a) the current sequence had to differ from each of the two previous sequences (lag 2 criterion), and (b) the weighted relative frequency of the current sequence had to be less than or equal to a certain threshold (cf. Denney & Neuringer, 1998). The relative frequency was computed by dividing the total number of occurrences of each sequence by the total number of completed sequences (trials). To weight recently emitted sequences more than past sequences, after each reinforcer delivery, the relative frequency of each of the 27 possible sequences was multiplied by a weighting coefficient ( $w = .95$ ) that exponentially decreased the contributions of past sequences. The weighted frequency of the current emitted sequence was compared to the threshold value (set at .05) to determine whether this sequence was to be reinforced. If the weighted relative frequency was less than or equal to the threshold (and the lag 2 requirement was satisfied), the reinforcers were delivered; otherwise, the sequence was considered a noncriterion one, and the reinforcers were not provided. In the beginning of this phase, all of a sequence’s counters were set to zero (thus, the first emission of each sequence was followed by reinforcement). During the EXT (B) phase, none of the emitted sequences produced reinforcers (extinction). In this phase, the emission of each sequence was followed by the next trial. Phase changes were not signaled. Each phase lasted 300 trials, except for the last phase, which ended when the number of reinforcers was equal to that obtained in the first exposure to the VAR contingency or after 300 trials, whichever occurred first.

After completing the experimental task, participants answered a postexperimental questionnaire and were debriefed.

#### Data analysis

The data from each phase were divided into blocks of 50 trials, yielding six blocks per experimental phase, and the data from the first and last blocks of each phase were

considered for the analyses. To evaluate the degree to which participants were responding systematically or in a random-like pattern, two measures of sequence dependency were employed:

*Autocorrelations* To evaluate higher-order patterns in the emission of sequences, lag 1 to 27 autocorrelations were computed (see also Maes, 2003). Lag 1 autocorrelations refer to correlations between the sequence emitted in trial  $n$  and the sequence emitted in trial  $n - 1$ ; lag 2 autocorrelations indicate correlations between trials  $n$  and  $n - 2$ ; and so on.

*First-order difference (FOD)* This measure reflects the arithmetic difference between the response in the current trial and the response in the preceding trial (Towse & Neil, 1998). To compute this measure, all sequences were recoded in an ascending order with numbers from 1 to 27 (e.g., sequence 111 was coded as 1; sequence 112 was coded as 2; sequence 113 as 3; and so on, until sequence 333, which was coded as 27). Thus, if across successive trials the sequences 1, 2, 3, 4, 5, and 2 were emitted, the obtained FOD for the first pair of trials ( $2 - 1$ ) would be equal to +1, and the same value would be obtained for the following three pairs ( $3 - 2$ ;  $4 - 3$ ;  $5 - 4$ ). On the other hand, the emission of the sequence 5 followed by sequence 2 would yield an FOD of  $-3$ . If participants were responding randomly, all types of transitions between sequences would be equally likely, so that all FOD values (ranging from  $-26$  to  $+26$ ) should be equiprobable. However, if participants were responding with some systematicity, some transitions would be more likely than others (such as emitting all sequences in ascending order of magnitude, a pattern that would yield an FOD of +1 on most trials).

In addition to autocorrelations and the FOD, more traditional measures of variability and performance were used:

*Percentages of sequences that met the variability criteria (MetVar)* MetVar was computed according to the following formula: (number of trials on which the variability criteria were met) / (total number of trials).

*Overall index of sequence uncertainty (U value)*  $U$  values were obtained according to the following equation (Miller & Frick, 1949):

$$\frac{-\sum\{RF_i \times [\log(RF_i)/\log(2)]\}}{[\log(27)/\log(2)]},$$

where RF is the relative frequency of the sequence  $i$  (for  $i = 1$  to  $n$ ), and  $n$  is the number of all possible sequences (27). If each of the 27 possible sequences were emitted equally often, the  $U$  value would be equal to 1; if only one sequence were emitted, the  $U$  value would be equal to 0.



**Mean reaction times (RTs)** RTs were computed as the time (in seconds) to complete a sequence on each trial. RTs were computed as a measure of response rate.

All of these measures were evaluated separately for each of the three experimental groups. Moreover, autocorrelations, FOD, MetVar, and  $U$  values were also calculated on the basis of the data produced by a random generator. To simulate the emission of random sequences, the random number generator of MATLAB was used. We produced 12 simulations of the random selection of 300 integers with values ranging from 1 to 27 (simulating the production of the 27 possible sequences in the present experiment). These data served as a baseline to compare the levels of randomness achieved by the experimental groups on each measure of variability.

## Results

The results were tested for significance with a mixed repeated measures analysis of variance (ANOVA) having Block (first and last) and Phase (VAR, EXT, and VAR) as within-subjects factors, and Group (systematic, random, and control) as a between-subjects factor. In some of the analyses performed, the sphericity assumption was violated, and therefore Greenhouse–Geisser-adjusted degrees of freedom are reported (recognizable by the noninteger values). To correct for nonsphericity, the Bonferroni adjustment for multiple comparisons was also employed. Partial eta-squared ( $\eta_p^2$ ) results are provided for the significant effects.

### Response strategy

To quantify the extent to which participants were using systematic and random strategies to produce their sequences, we employed two measures: autocorrelations and FOD.

**Autocorrelations** Strategic responses can be evaluated by higher-order measures such as autocorrelations. Random responding should produce very low levels of autocorrelations, since the response emitted on the current trial should be

completely independent of the previously emitted responses. To allow for the evaluation of the degree of randomness achieved by our participants, we computed autocorrelations for the simulated data. The patterns of autocorrelations obtained in the first and last blocks of 50 trials for four of our simulations are depicted in Fig. 1. As would be expected from random responding, the autocorrelations were low for all lags and for the first and last blocks.

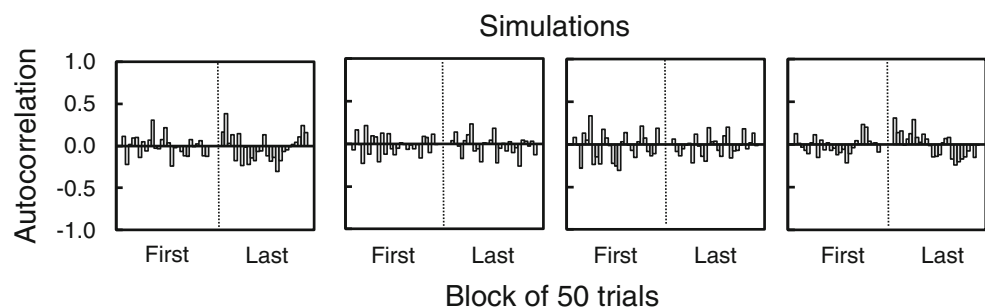
Figure 2 shows the autocorrelations in the first and last blocks of each experimental phase, computed for each participant of the systematic group. Participants S1 to S5 displayed high autocorrelations in all phases, whereas participants S6 to S10 showed high autocorrelations in some blocks but not in others. Nevertheless, for the latter participants, autocorrelations tended to be higher during the variability (Var) phases than during extinction (Ext). Lastly, participants S11 and S12 displayed low autocorrelations across most blocks. Given that these participants also reported not using a consistent systematic rule (see Appendix A), their data were excluded from the subsequent analyses.

Figure 3 presents the corresponding autocorrelation analysis for the participants in the random group. Most participants (R1 to R10) showed low autocorrelations in all blocks, with the exception of participant R8, who tended to show intermediate autocorrelations in the first block of the first Var phase and in the last block of reconditioning. Participants R11 and R12, on the other hand, showed high autocorrelations across most blocks, and they additionally reported using a systematic strategy to produce the sequences (see Appendix A). The data of these 2 participants were also excluded from the subsequent analyses.

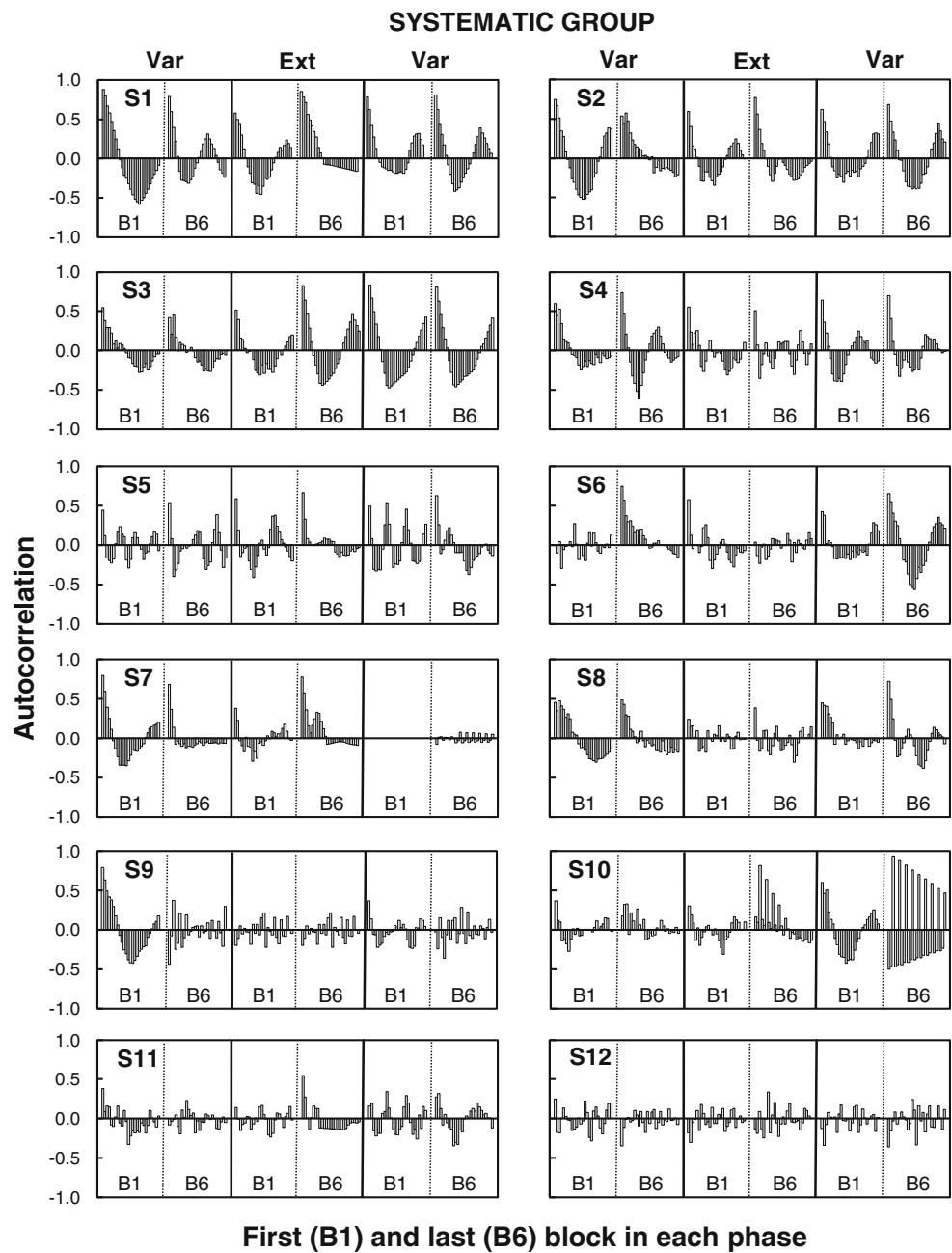
Figure 4 shows the autocorrelations computed for the control group. Participants C1 to C6 displayed low to intermediate autocorrelations in most blocks, whereas participants C7 to C12 displayed intermediate to high autocorrelations in most blocks.

**FOD** If participants were behaving randomly, transitions between any given pair of sequences should occur with the same likelihood, which would mean that any value of FOD (ranging from  $-26$  to  $+26$ ) would be equally probable. To allow us to compare the behavior of the different experimental groups with that in the computer-generated random data, we plotted

**Fig. 1** Lag 1 (leftmost bar of each block) to lag 27 (rightmost bar) autocorrelations in the first and last blocks of 50 trials (out of 300 trials) obtained in four simulations of random responding



**Fig. 2** Lag 1 (leftmost bar of each block) to lag 27 (rightmost bar) autocorrelations in the first and last blocks of each phase for each participant of the systematic group



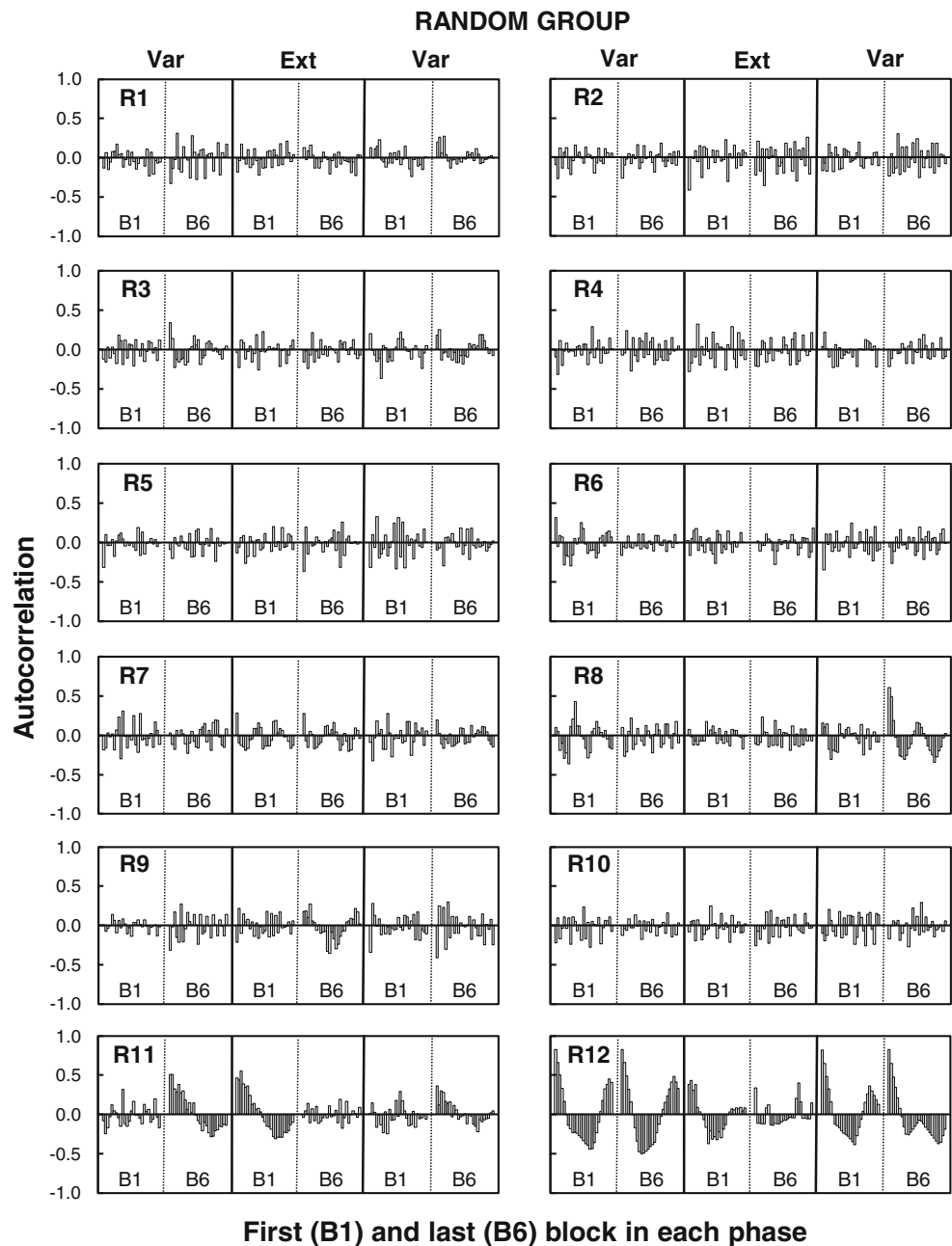
the cumulative frequency (as percentages) of each value of FOD obtained in the first and last blocks of each phase for the experimental groups and the FOD values obtained from the simulated data. Figure 5 shows the results for the first block of the initial VAR phase for the experimental groups and for the data from the random simulation. Similar patterns were obtained in the last block of that phase and in both blocks of the subsequent phases, and therefore the data are not shown.

As can be seen in Fig. 5, the systematic group displayed a pattern of responses that resembled a step function: The most frequent values of FOD were +1 and -1, thus indicating that participants were producing sequences in ascending

order (e.g., sequences 1, 2, 3, 4, 5, which would yield an FOD of +1 for all pairs) or descending order (e.g., 5, 4, 3, 2, 1, yielding an FOD of -1). For the random group, the curves were a roughly constant function very similar to the pattern expected in the case of random responding (panel in the bottom right), thus meaning that any emitted sequence was equally likely to be followed by any other sequence; the control group, on the other hand, showed more within-subjects variability, with some participants displaying strategic-like behavior (step-like curves) and some displaying more random-like behavior (constant curves).

To summarize the results of the different groups across phases, we arranged the absolute frequency of each FOD

**Fig. 3** Lag 1 (leftmost bar of each block) to lag 27 (rightmost bar) autocorrelations in the first and last blocks of each phase for each participant of the random group

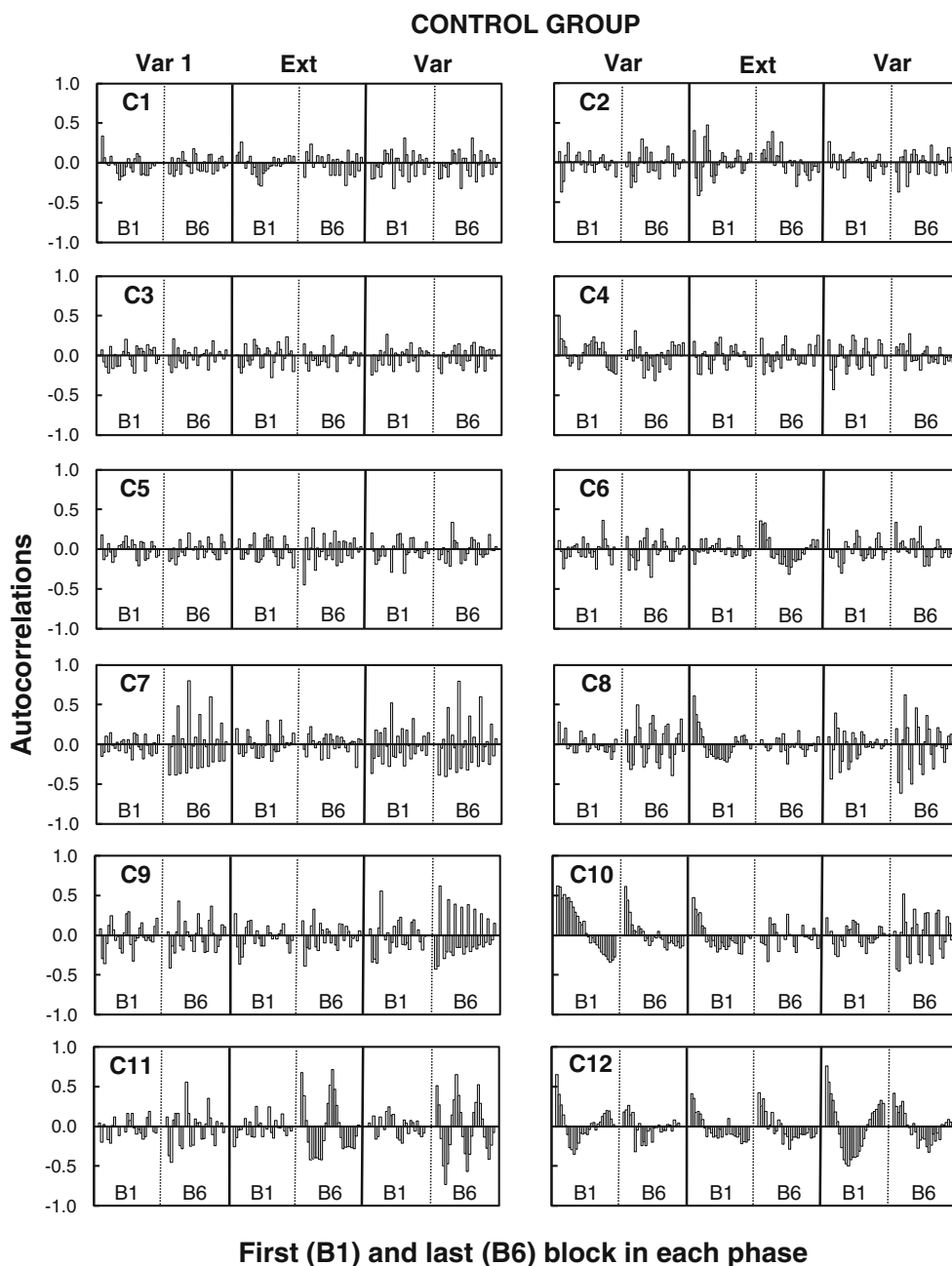


value and the absolute values of the autocorrelations for each participant in a descending order. Then we drew the maximum obtained value for each measure, which we termed *MaxFOD* and *MaxAut* (see also Maes, 2003), respectively. The larger the *MaxFOD* and *MaxAut*, the more the pattern deviated from random responding, or alternatively, the more it could be characterized as systematic. Figure 6 presents the average *MaxFOD* and *MaxAut* for the three experimental groups in the first and last blocks of each phase. The dashed lines indicate the predicted random performance drawn from our simulations.

For *MaxFOD*, a significant main effect of phase was found,  $F(2, 58) = 5.53$ ,  $p = .006$ ,  $\eta_p^2 = .16$ , and repeated contrasts

indicated that the first VAR phase was not significantly different from the EXT phase ( $F < 1$ ), but the EXT phase differed from the subsequent VAR phase,  $F(1, 29) = 8.38$ ,  $p = .007$ ,  $\eta_p^2 = .22$ . The main effect of group also reached significance,  $F(2, 29) = 21.65$ ,  $p < .001$ ,  $\eta_p^2 = .60$ , and repeated contrasts indicated that the systematic group was different from the random group ( $p < .001$ ) and the random group was different from the control group ( $p = .033$ ). Both effects interacted significantly [Phase  $\times$  Group:  $F(4, 58) = 3.1$ ,  $p = .022$ ,  $\eta_p^2 = .18$ ], probably because the transition to extinction produced a decrease in FOD for the systematic group, a slight increase for the control group, but no change for the random group. To test this possibility more closely, we

**Fig. 4** Lag 1 (leftmost bar of each block) to lag 27 (rightmost bar) autocorrelations in the first and last blocks of each phase for each participant of the control group

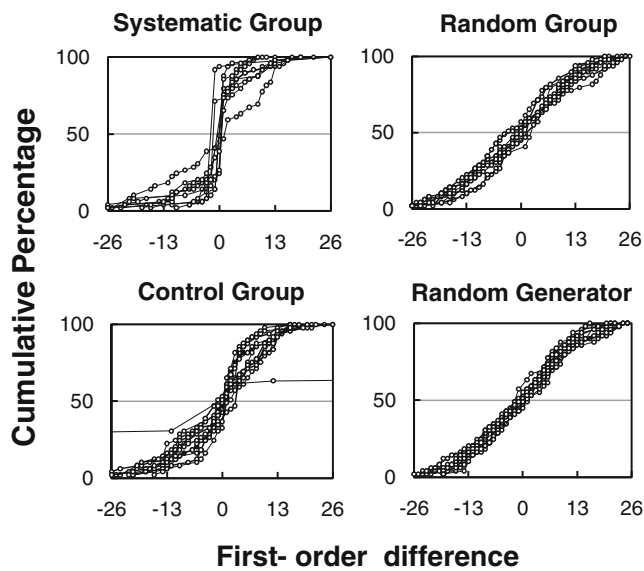


**First (B1) and last (B6) block in each phase**

computed the change in FOD with the transition from the first VAR to the EXT phase [i.e., (first block of EXT) / (last block of VAR)], which yielded values of 0.75, 1.07, and 1.21 for the systematic, random, and control groups, respectively. These values were submitted to a one-way ANOVA that yielded only a marginally significant effect of group,  $F(2, 31) = 2.67, p = .086$ . Pairwise comparisons suggested that the systematic and control groups slightly differed from each other ( $p = .091$ ), but the systematic and random groups did not differ, nor the random and control groups ( $p > .410$ ). Finally, the main effect of block ( $F < 2.1, p > .15$ ) and the two-way (Block  $\times$  Phase, Block  $\times$  Group) and three-way (Block  $\times$  Phase  $\times$  Group) interactions were not significant ( $F_s < 2, p_s > .15$ ).

To test whether responding during the VAR phases differed from random responding, we compared the performance of the experimental groups across the VAR phases with the data from the random generator (12 simulations) by means of a mixed repeated measures ANOVA with three factors (Block, Phase, and Group). Given that we were specifically interested in the between-subjects factor (Group), we focused on the main effect of this variable and on the results of simple contrasts with the performance of the random generator as the reference category. This analysis yielded a significant effect of group,  $F(3, 40) = 40.64, p < .001, \eta_p^2 = .75$ , and simple contrasts showed that the systematic ( $p < .001$ ) and control ( $p = .007$ ) groups, but

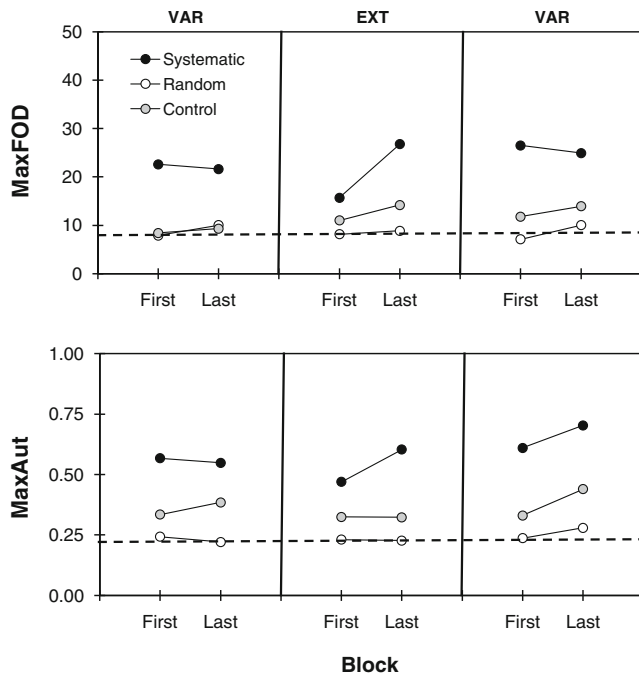




**Fig. 5** Cumulative percentages for each value of first-order difference (ranging from -26 to +26) in the first block of the VAR phase for individual participants (represented by different lines) in the systematic, random, and control groups, and for the data obtained from 12 simulations of random responding

not the random group ( $p = .257$ ), significantly departed from the random-generator data.

Regarding MaxAut, the ANOVA yielded a significant effect of phase,  $F(2, 58) = 4.77, p = .012, \eta_p^2 = .14$ . Repeated



**Fig. 6** Averages of the maximum absolute frequency of first-order difference (MaxFOD) and of the maximum value of autocorrelations (MaxAut) obtained in the first and last 50-trial blocks of each phase for the systematic, random, and control groups. The dashed lines depict the values expected with random responding

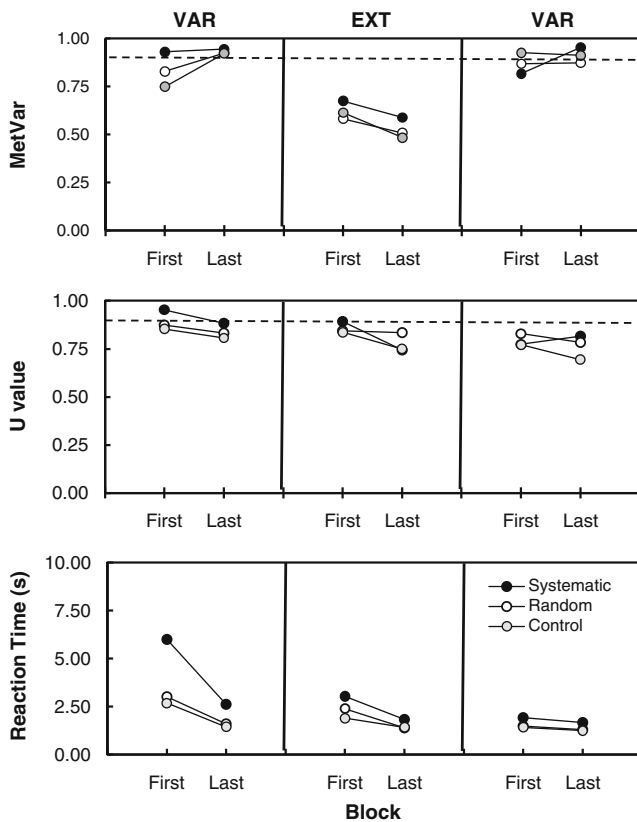
contrasts showed that the first VAR phase was not different from the EXT phase ( $F < 1$ ), but the EXT phase differed from the subsequent VAR phase,  $F(1, 29) = 11.44, p = .002, \eta_p^2 = .28$ . The main effect of block was only marginally significant,  $F(1, 29) = 3.72, p = .064, \eta_p^2 = .11$ , probably because the tendency of MaxAut to increase across blocks is only visible during the EXT phase and the subsequent VAR phase. The main effect of group also reached significance,  $F(2, 29) = 37.71, p < .001, \eta_p^2 = .72$ , and pairwise comparisons showed that the systematic group had a higher MaxAut than the random group ( $p < .001$ ) and that the random group had a lower MaxAut value than the control group ( $p = .039$ ). The two- and three-way interactions were not significant ( $F < 1.5, p > .25$ ). Finally, comparison of the performance of the three groups during the VAR phases against the data of the random generator yielded a significant effect of group,  $F(3, 40) = 43.27, p < .001, \eta_p^2 = .76$ , and simple contrasts showed that the systematic and control groups significantly departed from the random-generator data ( $p < .001$ ), but the random group did not ( $p = .682$ ).

Figure 7 shows plots of MetVar (top panels),  $U$  value (middle panels), and RT (bottom panels) in the first and last blocks of each phase per group. Dashed lines indicate the performance predicted by random responding (only for the two variability measures).

MetVar

Inspection of Fig. 7 shows that MetVar tended to increase from the first to the last block during the first VAR phase, reaching the level predicted by random responding. This trend was true for all groups except the systematic group, whose performance was always above random. During the EXT phase, MetVar was reduced for all groups, whereas reexposure to the VAR phase was followed by recovery of the preextinction MetVar levels. These effects are confirmed by the results of the ANOVA, which yielded a significant main effect of phase,  $F(2, 58) = 34.80, p < .001, \eta_p^2 = .55$ . Repeated contrasts indicated that responding differed significantly in the first VAR and the EXT phase, as well as in the EXT and the second VAR phase,  $F_s(1, 29) = 58.02$  and  $41.05, p_s < .001, \eta_p^2_s = .67$  and  $.59$ . The main effect of phase was modulated by an interaction with block,  $F(2, 58) = 81.52, p < .001, \eta_p^2 = .74$ , and a marginally significant interaction with group,  $F(2, 58) = 2.38, p = .062, \eta_p^2 = .14$ . Repeated contrasts showed that the effect of block differed in the VAR and EXT phases: While during the VAR phases the MetVar value increased across blocks, during the EXT phase the tendency was to decrease:  $F_s(2, 29) = 24.85$  and  $87.9, p_s < .001, \eta_p^2_s = .46$  and  $.75$ , for comparisons between the first VAR and EXT phases and between the EXT and second VAR phases, respectively.

Regarding the Phase  $\times$  Group interaction, repeated contrasts indicated that the groups differed in the transition from



**Fig. 7** Average proportions of sequences meeting the variability criteria (MetVar), sequence uncertainty ( $U$  value), and reaction times for the systematic, random, and control groups in the first and last blocks of each experimental phase. The dashed lines depict the values expected with random responding

the VAR to the EXT phase,  $F(2, 29) = 3.62, p = .039, \eta_p^2 = .20$ . We ran an ANOVA (with Block and Group as factors) on the data from the first VAR phase, and another on the data from the EXT phase. In both phases, the effect of block was significant [VAR,  $F(1, 29) = 10.36, p = .003, \eta_p^2 = .26$ ; EXT,  $F(1, 29) = 4.72, p = .038, \eta_p^2 = .14$ ]. However, only during the VAR phase did the effect of group reach significance,  $F(2, 29) = 3.95, p = .030, \eta_p^2 = .21$ , and the Block  $\times$  Group interaction was marginally significant,  $F(2, 29) = 2.53, p = .071, \eta_p^2 = .15$ . Simple contrasts with the systematic group as the reference category indicated that this group differed significantly from the control group ( $p = .009$ ), but not from the random group ( $p = .12$ ). The effects of block and group and the Block  $\times$  Group and Block  $\times$  Phase  $\times$  Group interactions were not significant ( $F < 1.5, p > .20$ ).

To test whether responding during the VAR phases differed from random responding, we again compared the performance of the experimental groups across the VAR phases with the data from the random generator. The main effect of group was still not significant ( $F < 1, p > .35$ ), thus suggesting that all groups reached the level predicted by random responding during the variability contingency.

## $U$ value

The  $U$  values were relatively high (above .75) during all phases. The ANOVA showed main effects of phase,  $F(1.6, 47.6) = 8.60, p = .001, \eta_p^2 = .23$ , and block,  $F(1, 29) = 15.22, p = .001, \eta_p^2 = .34$ . Repeated contrasts indicated that  $U$  values were higher during the first VAR phase than during the EXT phase,  $F(1, 29) = 10.28, p = .003, \eta_p^2 = .26$ , but were similar across the EXT and second VAR phases ( $F < 1$ ). In all phases,  $U$  values were smaller in the last block than in the first block. The main effect of group and the Phase  $\times$  Group, Block  $\times$  Group, and Block  $\times$  Phase  $\times$  Group interactions were not significant ( $F_s < 1.5, p_s > .25$ ).

As can be seen in Fig. 7, the obtained  $U$  values tended to be smaller than the ones predicted by random responding across all phases. To test this assumption, we compared the  $U$  values obtained by the experimental groups during the VAR phases with the ones obtained in the simulations of random responding. After entering the random generator as one of the groups, the main effect of group reached significance,  $F(3, 40) = 3.93, p = .015, \eta_p^2 = .23$ . Simple contrasts indicated that the performance of the control group was significantly different from the random-generator data ( $p = .002$ ), whereas for the random group this effect was marginally significant ( $p = .061$ ). The systematic group did not differ from the random generator ( $p = .234$ ).

## Reaction times

For RTs, the ANOVA yielded main effects of block,  $F(1, 29) = 64.64, p < .001, \eta_p^2 = .69$ , and phase,  $F(2, 58) = 37.70, p < .001, \eta_p^2 = .57$ . Repeated contrasts showed significant differences between the first VAR and the EXT phase, as well as between the EXT and the second VAR phase:  $F_s(1, 29) = 21.94$  and  $28.26, p_s < .001, \eta_p^2_s = .43$  and  $.50$ , respectively. The effects of block and phase interacted significantly,  $F(2, 58) = 19.42, p < .001, \eta_p^2 = .40$ , because the reduction in RTs across blocks was greater in the first phase than in the second,  $F(1, 29) = 9.65, p = .004, \eta_p^2 = .25$ , and greater in the second than in the third,  $F(1, 29) = 16.28, p < .001, \eta_p^2 = .36$ .

The main effect of group was also significant,  $F(2, 29) = 10.44, p < .001, \eta_p^2 = .42$ , given that the systematic group was slower than the other groups ( $p = .002$ ), but the random and control groups did not differ ( $p = .482$ ). The Group factor also entered in significant interactions with block,  $F(2, 29) = 4.71, p = .017, \eta_p^2 = .25$ , and phase,  $F(4, 58) = 7.35, p < .001, \eta_p^2 = .34$ , and a triple interaction with block and phase was also present,  $F(4, 58) = 3.08, p = .023, \eta_p^2 = .17$ . All of these interactions reflected relatively higher gains in speed across blocks and phases for the systematic group as compared to the other two groups.

## Discussion

Our findings support Maes's (2003) suggestion that, under an unconstrained variability task—that is, when either random or systematic patterns are equally effective for producing reinforcers—the resulting pattern may be influenced by the instructions (see also Souza, Abreu-Rodrigues, & Baumann, 2010). Our instruction manipulation was successful in establishing differential response patterns across groups: Participants in the systematic and random groups produced systematic and random-like performance—as assessed by the FOD and autocorrelation analyses—therefore consistently deviating from (systematic group) or approximating (random group) the performance predicted by random responding. On the other hand, in the absence of instructions (control group), between-subjects variability was observed: Some participants behaved in a random-like fashion, whereas others employed more systematic strategies. This was also the case for most participants in Maes's study. Considering that (1) several studies have shown that verbal stimuli provided by the experimenter (instructions) and verbal stimuli provided by the individual (self-instructions) tend to produce similar behavioral effects (Baumann, Abreu-Rodrigues, & Souza, 2009; Matthews, Catania, & Shimoff, 1985; Rosenfarb, Newland, Brannon, & Howey, 1992; Torgrud & Holborn, 1990), and (2) individuals tend to formulate self-instructions while performing a task (see, e.g., Skinner, 1969), it seems plausible to suggest that self-instructions might be one of the reasons why participants in our control group, and in other studies with no explicit instructions about randomness, differed in their likelihood of approaching a variability task in a systematic or random fashion.

Interestingly, the present study showed that instructions to perform in a random-like fashion can be effectively used to reduce higher-order dependencies under variability contingencies, therefore promoting more randomness in human behavior. This result contrasts with several reports in the literature showing that such instructions have not gained control over behavior (Baddeley 1966; Baddeley et al., 1998; Bar-Hillel & Wagenaar, 1991; Rapoport & Budescu, 1997; Wagenaar, 1972), with the exception of those situations in which feedback concerning the degree of randomness was also given (Neuringer, 1986). On the contrary, in the present study, participants showed random patterns even though there was no feedback on randomness. This inconsistency could be related to the fact that in the present study variability-dependent feedback was provided. Thus, although reinforcement was not directly contingent on the degree of randomness, rule following was eventually reinforced.

Regarding the other behavioral measures evaluated in the present study, systematic responding tended to increase the likelihood of meeting the variability contingency in the first block of training, as compared to a random strategy;

however, this difference disappeared as training progressed. During extinction, the probability of meeting the MetVar decreased substantially for all groups, and reexposure to the variability contingency was followed by an abrupt increase in the measure, thus showing sensitivity to the operant contingency established between low-frequent sequences and positive feedback. These effects, however, were not affected by the systematic or random strategies underlying the production of the sequences.

Systematic responding was also associated with higher  $U$  values, probably because the most frequent form of systematicity employed (i.e., emitting sequences in ascending or descending order) promoted an even distribution of sequences. The  $U$  values for the systematic group tended to be indistinguishable from the ones predicted by random responding. For the random and control groups, however,  $U$  values were usually below the level predicted by random responding, suggesting that participants tended to be biased toward emitting some sequences. These results indicate that although higher-order dependencies (e.g., FOD and autocorrelations) were reduced by the instruction to perform randomly, participants failed to emit all sequences with equal probability. This finding also shows that, depending on the measurement under consideration, higher variability can be ascribed to participants showing higher-order dependencies than to participants using more random strategies to produce sequences.

Finally, the analysis of RTs showed that responding was slower in the first phase than in subsequent ones, and responding became faster the more trials were completed. Furthermore, producing sequences according to a systematic rule consumed more time than did employing random (random group) or no specific strategies (control group), especially in the first phase. This result suggests that following a well-defined instruction in generating sequences involved the greatest difficulty, because it required encoding the previously emitted sequence and retrieving the next sequence to meet the instruction. The finding that random responding was related to faster reaction times than was systematic responding is inconsistent with previous reports. Some studies have shown that reducing the interval between two responses leads to more repetitions, whereas random-like responding improves with the imposition of long delays between responses (Baddeley et al., 1998; Neuringer, 1991). The imposition of these delays has usually been assumed to impair memory for the previous responses, therefore encouraging the emergence of response patterns that are not correlated with the previous emitted response (one of the criteria for randomness). It is important, however, to highlight that when participants are asked to behave randomly, this instruction might reduce the likelihood of strategies such as actively remembering the last emitted response and trying to generate the next response by means of some rule. Nevertheless, when no instructions are provided, participants might use the delay

to prepare a specific response. Consequently, it is possible to assume that the imposition of delays between responses might have different effects, depending on the instructions (or possible self-instructions): Long delays might increase randomness when participants are instructed to behave in a random-like fashion, whereas it might decrease randomness if the instruction is to behave systematically.

In conclusion, the present findings show that instructions can have large effects on the way that participants try to satisfy a variability contingency. Although traditional measures of behavioral variability (MetVar and *U* value) may not be affected by the different strategies employed by the participants, these strategies can be revealed by measures that consider high-order dependencies—such as autocorrelations and first-order differences—and by comparing the obtained data with those produced by a random generator. Our findings also illustrate that variability and randomness are not synonyms: Variability is related to how differently one behaves, whereas randomness is related to the predictability of this behavior (Stokes & Harrison, 2002). These concepts might overlap under some conditions, but nevertheless, one cannot assume randomness from variation—because variability can also be produced in a systematic fashion—or, conversely, variation from a random process—because random processes can also lead to repetitions.

## Appendix A

Participants' responses to the question *Did you use any rule to emit the sequences? Which one?* in the postexperimental questionnaire

### Part. Systematic Group

- S1 Emitted numbers in ascending and descending order.
- S2 I tried, but as time went by, I got confused and I couldn't tell in which sequence I had stopped. I opted to emit sequences in ascending order.
- S3 Yes, in ascending order from 111 to 333.
- S4 I started with the easy sequences, and then I tried several ones until I got points.
- S5 Only the rule of putting the numbers in numerical order.
- S6 Yes. Several different rules.
- S7 I used mainly two rules. I varied the last digit of the sequences, and I emitted sequences that started with a 3 when I got no points.
- S8 Yes. From the smallest to the largest, I sorted them in ascending order.
- S9 I tried to organize myself in terms of the change in the digits. I tried to find new sequences. Sometimes I changed only the first, sometimes only the last digit.
- S10 Yes. I alternated the digit slots.
- S11 Yes. I tried to start with a digit and vary the others. Later I would change the digits.

- S12 I didn't use a constant rule. I repeatedly typed the digits, using three fingers, and varying the order.

### Random Group

- R1 No.
- R2 Yes. Never repeat a sequence twice.
- R3 I tried to memorize the sequences I typed to vary them.
- R4 No.
- R5 My rule was not to repeat the sequences, but there wasn't any algorithm.
- R6 No.
- R7 Yes. Not to use the previous rule.
- R8 Yes. Start with one and try all sequences with one, then with two and three.
- R9 No. I used random sequences.
- R10 I thought of several sequences, and then I typed one I haven't used recently.
- R11 When I was very bored I used the factorial rule, but most of the time I did it randomly.
- R12 Yes. Systematic sequences.

### Control Group

- C1 No. I repeated the same sequence in intervals of three sequences.
- C2 Not to repeat the just-emitted sequence.
- C3 I started with the basic sequences and then I tried to remember one that I hadn't used.
- C4 No. I just tried to be quick.
- C5 First, I emitted all possible sequences. Then I made pairs like 123 and 321, but there were moments in which I just pressed the keys.
- C6 Every 10th sequence, I paid attention not to repeat the sequence, and most of the time I followed a specific order.
- C7 Yes. 111, 222, 333, 123, 321, 112, 113, 221, 223, 312.
- C8 Most of the time it was random because a sequence was only repeated in an interval of three sequences.
- C9 Logical sequences so that I wouldn't forget any sequence.
- C10 Not to repeat a sequence very soon. Then I made cycles of five triplets to be quicker.
- C11 Yes. I pressed the first key three times, then two times and another one, than one time and another one, and so on.
- C12 I used the ascending order starting at 111 (the smallest) to 333 (the largest).

## References

- Abreu-Rodrigues, J., Hanna, E. S., Cruz, A. P. M., Matos, R., & Delabrida, Z. (2004). Differential effects of midazolam and pentylentetrazole on behavioral repetition and variation. *Behavioural Pharmacology*, *15*, 535–543.
- Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, *18*, 119–129. doi:10.1080/14640746608400019
- Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, *51A*, 819–852. doi:10.1080/027249898391413



- Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*, 428–454. doi:10.1016/0196-8858(91)90029-1
- Baumann, A. A., Abreu-Rodrigues, J., & Souza, A. S. (2009). Rules and self-rules: Effects of variation upon behavioral sensitivity to change. *Psychological Record*, *59*, 641–670.
- Cohen, L., Neuringer, A., & Rhodes, D. (1990). Effects of ethanol on reinforced variations and repetitions by rats under a multiple schedule. *Journal of Experimental Analysis of Behavior*, *54*, 1–12. doi:10.1901/jeab.1990.54-1
- Denney, J., & Neuringer, A. (1998). Behavioral variability is controlled by discriminative stimuli. *Animal Learning & Behavior*, *26*, 154–162. doi:10.3758/BF03199208
- Machado, A. (1989). Operant conditioning of behavioral variability using a percentile reinforcement schedule. *Journal of Experimental Analysis of Behavior*, *52*, 155–166. doi:10.1901/jeab.1989.52-155
- Machado, A. (1993). Learning variable and stereotypical sequences of responses: Some data and a new model. *Behavioural Processes*, *30*, 103–129. doi:10.1016/0376-6357(93)90002-9
- Maes, J. H. R. (2003). Response stability and variability induced in humans by different feedback contingencies. *Learning & Behavior*, *31*, 332–348.
- Matthews, B. A., Catania, A. C., & Shimoff, E. (1985). Effects of uninstructed verbal behavior on nonverbal responding: Contingency descriptions versus performance descriptions. *Journal of Experimental Analysis of Behavior*, *43*, 155–164.
- McElroy, E., & Neuringer, A. (1990). Effects of alcohol on reinforced repetitions and reinforced variations in rats. *Psychopharmacology*, *102*, 49–55. doi:10.1007/BF02245743
- Miller, G. A., & Frick, F. C. (1949). Statistical behavioristics and sequences of responses. *Psychological Review*, *56*, 311–324. doi:10.1037/h0060413
- Neuringer, A. (1986). Can people behave “randomly?” The role of feedback. *Journal of Experimental Psychology: General*, *115*, 62–75. doi:10.1037/0096-3445.115.1.62
- Neuringer, A. (1991). Operant variability and repetition as functions of interresponse time. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 3–12. doi:10.1037/0097-7403.17.1.3
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, *9*, 672–705.
- Neuringer, A. (2004). Reinforced variability in animals and people. *American Psychologist*, *59*, 891–906. doi:10.1037/0003-066x.59.9.891
- Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 429–452. doi:10.1037/0097-7403.11.3.429
- Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review*, *104*, 603–617. doi:10.1037/0033-295X.104.3.603
- Rosenfarb, I. S., Newland, M. C., Brannon, S. E., & Howey, D. S. (1992). Effects of self-generated rules on the development of schedule-controlled behavior. *Journal of Experimental Analysis of Behavior*, *58*, 107–121. doi:10.1901/jeab.1992.58-107
- Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General*, *111*, 23–59. doi:10.1037/0096-3445.111.1.23
- Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. New York, NY: Appleton-Century-Crofts.
- Souza, A. S., & Abreu-Rodrigues, J. (2010). Discriminative proprieties of vary and repeat contingencies. *Behavioural Processes*, *85*, 116–125. doi:10.1016/j.beproc.2010.06.018
- Souza, A. S., Abreu-Rodrigues, J., & Baumann, A. A. (2010). History effects on induced and operant variability. *Learning & Behavior*, *38*, 426–437. doi:10.3758/LB.38.4.426
- Stokes, P. D. (1999). Learned variability levels: Implications for creativity. *Creativity Research Journal*, *12*, 37–45.
- Stokes, P. D., & Harrison, H. M. (2002). Constraints have different concurrent effects and aftereffects on variability. *Journal of Experimental Psychology: General*, *131*, 552–566. doi:10.1037/0096-3445.131.4.552
- Torgrud, L. J., & Holborn, S. W. (1990). The effects of verbal performance descriptions on nonverbal operant responding. *Journal of Experimental Analysis of Behavior*, *54*, 273–291. doi:10.1901/jeab.1990.54-273
- Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers*, *30*, 583–591. doi:10.3758/BF03209475
- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, *77*, 65–72. doi:10.1037/h0032060
- Ward, R. D., Bailey, E. M., & Odum, A. L. (2006). Effects of d-amphetamine and ethanol on variable and repetitive key-peck sequences in pigeons. *Journal of Experimental Analysis of Behavior*, *86*, 285–305. doi:10.1901/jeab.2006.17-06
- Ward, R. D., Kynaston, A. D., Bailey, E. M., & Odum, A. L. (2008). Discriminative control of variability: Effects of successive stimulus reversals. *Behavioural Processes*, *78*, 17–24. doi:10.1016/j.beproc.2007.11.007