RESEARCH ARTICLE

CrossMark

Free choice tasks as random generation tasks: an investigation through working memory manipulations

Christoph Naefgen[1] · Markus Janczyk[1]

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Free choice tasks are tasks in which two or more equally valid response options per stimulus exist from which participants can choose. In investigations of the putative difference between self-generated and externally triggered actions, they are often contrasted with forced choice tasks, in which only one response option is considered correct. Usually, responses in free choice tasks are slower when compared with forced choice task responses, which may point to a qualitative difference in response selection. It was, however, also suggested that free choice tasks are in fact random generation tasks. Here, we tested the prediction that in this case, randomness of the free choice responses depends on working memory (WM) load. In Experiment 1, participants were provided with varying levels of external WM support in the form of displayed previous choices. In Experiment 2, WM load was induced via a concurrent n-back task. The data generally confirm the prediction: in Experiment 1, WM support improved both randomness and speed of responses. In Experiment 2, randomness decreased and responses slowed down with increasing WM load. These results suggest that free choice tasks have much in common with random generation tasks.

Introduction

In everyday life we often have to make choices without having a clear criterion for which option is better: Choosing what to eat when we only care about whether we eat, which set of purpose-appropriate clothes to pick from our wardrobe, from which lane we want to take a shopping cart when they're all equally far away and so on. Despite occasional assertions to the contrary,[1] we make such decisions with ease and swiftly. This type of choice devoid of almost all personal meaning, however, is also often used in laboratories when certain modes of action selection are investigated with so-called free choice tasks.

Free choice tasks

In these tasks, participants are instructed to freely choose one of two (or more) response options that are considered

equally correct. For example, consider a task in which whenever an 'H' is displayed on a screen, participants are supposed to press either a button to their left or a button to their right. Often, the participants are instructed to avoid obvious patterns in their choices (like left-right-left-right, for example) and to give all response options in equal proportions. We will discuss potential issues with this type of instruction in the subsequent "Free choice and random generation tasks" section, after we have introduced the task and important observations in the following. The experiments reported in this paper address critical aspects following from such instructions.

Starting with Berlyne's (1957) study, free choice tasks are often used in contrast with forced choice tasks, in which only one response is considered correct to a stimulus. One almost universal observation in the literature is that free choice response times (RTs) are longer than forced choice RTs (but see, e.g., Wirth et al. 2018 for an exception). This RT difference might be taken to indicate qualitative differences with

✉ Christoph Naefgen
   christoph.naefgen@uni-tuebingen.de

[1] Department of Psychology, Eberhard Karls University
   of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany

---

[1] For example, in the thought experiment of Buridan's ass, a hungry donkey has to choose between two piles of hay, resulting in the donkey's death of hunger because there is no criterion by which to choose a pile (see also Rescher 2005 for more information).

regard to response selection. Accordingly, free choice tasks are often used to operationalize what has been termed *self-generated* (or intentional, internally generated, intention-based, voluntary, goal-directed) action, while forced choice tasks are often used to operationalize *externally triggered* (or stimulus-based) action (e.g., Brass and Haggard 2008; Herwig et al. 2007; Passingham et al. 2010; Keller et al. 2006; Waszak et al. 2005). In support of this, there is some evidence that associations between actions and their effects can only be learned in an intention-based action control mode as operationalized with free choice tasks (Herwig et al. 2007; see also; Gaschler and Nattkemper 2012; Herwig and Waszak 2009, 2012; Pfister et al. 2010). However, Pfister et al. (2011) reported that these associations are also learned in forced choice tasks. In addition, there is ample evidence that action effects play a role even when using forced choice tasks (e.g., Gozli et al. 2016; Huffman et al. 2018; Janczyk et al. 2012a, b, 2014, 2017; Kühn et al. 2009, Exp. 3; Kunde 2001; Kunde et al. 2012; Pfister and Kunde 2013; Wolfensteller and Ruge 2011). In sum, it appears that the majority of evidence argues for the same role of action effects in forced and free choice tasks. This conclusion received additional support from other lines of research. For example, Janczyk et al. (2015a) compared both task types with regard to their susceptibility to dual-task interference. While replicating the RT difference in all experiments, no differences in dual-task costs between free and forced choice tasks were observed, again pointing to similar "action control mechanisms" involved in both tasks. In line with this, the RT difference was attributed to a perceptual source in a further study (Janczyk et al. 2015b). Coming from a different perspective, Bermeitinger and Hackländer (2018) observed that response priming effects induced by motion primes affected both free and forced choice tasks similarly.

If, then, both tasks do not differ regarding their response selection mechanisms, it appears helpful to identify further commonalities. As a step toward this, Naefgen et al. (2017) viewed the RT difference through a sequential sampling lens (e.g., Grice 1968). In such a framework, evidence for or against a response option (or more precisely in the context of that study: the desired goal state, that is, the depressing of a left or right response key) is noisily accumulated over time. Once the total amount of this evidence surpasses one of the thresholds, a response is emitted. This results in three theoretically relevant parameters for a choice type: The speed of evidence accumulation, the thresholds for making a choice, and the time not spent accumulating evidence (such as, e.g., time needed for the motor execution of the choice made). Within this framework, Naefgen et al. then asked whether the RT difference can be attributed to differences in the speed of evidence accumulation or to differences outside the accumulation process. To this end, the amount of catch-trials (e.g., Bausenhart et al. 2010) and time pressure

(e.g., Dror et al. 1999) were used to manipulate decision thresholds. If differences in evidence accumulation were the reason, the RT difference should become smaller the lower the thresholds. As this was not observed, the cause is likely located in a process different from evidence accumulation, that is, in the non-accumulation time. The present study aims to address the nature of this process and focuses on the generation of random responses as one candidate.

Free choice and random generation tasks

Frith (2013) argued that in free choice tasks, "in essence, the experimenter is asking her subjects to try to be unpredictable and random" (p. 291). He based this argument both on psychological evidence that participants associate randomness and the perception of choices as free (Ebert and Wegner 2011) and on neuroimaging evidence that random choice tasks and free choice tasks activate similar brain regions (Jahanshahi et al. 2000; Jenkins et al. 2000). This becomes even more evident when looking at the similarities between the instructions for free choice tasks and random generation tasks. The former appear in three variants: (1) explicit instructions to choose responses at random, (2) instructions similar to random generation instructions (e.g., avoidance of patterns[2]), and (3) instructions emphasizing spontaneity or freedom of choice. Lastly, there are also studies in which no instruction as to the desired patterns was reported. Examples for these categories can be found in Table 1. Please note that this overview is meant as an illustration, and is not exhaustive. One thing illustrated by Table 1 is the prevalence of instructions to avoid patterns in the free choice responses. One reason for such instructions is that, when they are not given, participants sometimes give responses with only one or almost only one of the response options.

While this type of instruction could be argued to constrain the choices that participants can give, this is true of all tasks that could feasibly be observed in an experimental laboratory. However, free choice responses are still less constrained than forced choice responses. While free choice instructions and random generation instructions bear similarities, free choice instructions are used this way in the literature on self-generated action and are, as such, worthy of investigation. The next section will discuss the relationship

---

[2] Indeed, the type of instruction used in free choice contexts bears similarities to a common mathematical definition of randomness derived from Kolmogorov complexity (Martin-Löf 1966). (Over-) Simplified, according to this definition, if a string of information can be described in a more concise manner than if it were simply written out, it is not random. For example, the number 4,294,967,296 can be described much shorter as 2^32. Thus, the number would not be seen as very random.

Table 1 Illustrative examples of different instructions for free choice tasks as well as random generation tasks

Example of…	Inclusion criteria	Example
Explicitly random responses	Explicit mention of randomness as a goal	"The subjects were instructed to choose the order of their movements at random." (Hadland et al. 2001, p. 1105; see also Waszak et al. 2005; Elsner and Hommel 2001)
Similar to random response instructions	Overlap between instructions and definitions of randomness No explicit mention of randomness	"[…] participants were instructed to decide spontaneously to produce one or the other action effect without relying on any specific strategy. They were told to choose each alternative about equally often, but it was stressed that the focus should be on spontaneous decisions rather than on a perfectly even distribution of responses." (Pfister and Kunde 2013, p. 650; see also Linser and Goschke 2007)
Emphasizing spontaneity/freedom of choice	No mention of randomness No particular overlap in instructions with definitions of randomness Mention of spontaneity or freedom of choice as goal	"Participants were instructed to […] decide spontaneously between the two response alternatives in free choice trials" (Pfister et al. 2010, p. 319; see also Herwig et al. 2007)
None reported	No explication of instructions present in the text	"When lights of both colors appeared, either response, but not both, was to be performed." (Berlyne 1957, p. 109)
Random generation task instruction	Instructions explicitly aimed at eliciting theoretically random generation behavior	"It was pointed out explicitly that the sequence would be completely jumbled and should not be likely to contain sequences such as '12345' or '98765'" (Azouvi et al. 1996, p. 320)

between random generation tasks and how they are affected by working memory (WM) manipulations.

Random generation and working memory

Baddeley reported that random generation performance can be influenced by various factors such as time constraints (Baddeley 1962, as cited in, 1966) or concurrently performed tasks (Baddeley 1966), suggesting that the capacity to create random information is limited in some way. As such, it stands to reason that adding a secondary task that involves WM to the random generation task would interfere with the random generation task. For example, Cooper et al. (2012) used a dual-tasking paradigm in which a random digit (1–9) generation task was coupled either with a 2-back task or a go/no-go task. Indeed, performance in the random generation task as measured through RTs and different indices of randomness was worse when combined with the 2-back task.

Additional evidence for a relationship between WM functions and random generation can be derived from principal component analyses. In particular, Miyake et al. (2000) reported correlations between the executive functions of updating and inhibition with measures of randomness (equality of response usage and inhibition of prepotent associates, respectively) as described by Towse and Neil (1998).

In sum, the literature suggests that WM plays a critical role in random generation tasks. The assessment of randomness will be discussed in the next section.

Measuring randomness

A difference between the aforementioned random generation tasks and free choice tasks is that in free choice tasks there are most often only two response options while for the random generation tasks there were usually nine response options. This renders several ways of how randomness of a choice sequence can be measured less informative. For example, it cannot be measured, as it can be with nine digits, whether two subsequent responses have adjacent values.

As there is a plethora of different measures of randomness (Towse and Neil 1998 alone described 14 different measures in their review), it is necessary to choose which one(s) to use. For the purposes of the present paper, randomness will be measured through the local unevenness (LU) measure (see, e.g., Heuer et al. 2005, 2010). While earlier studies used a more general form of LU, the following description is specific to a two-response-options situation with left and right responses.

In essence, the LU is a measure of the deviation of empirical responses from an ideal random distribution of responses, as measured in running windows of predefined sizes. "Running window" here means that a sequence is

```
        Window Size 2                      Window Size 4

LRLLRRLRRRLL                      LRLLRRLRRRLL
LR              0                 LRLL            .25
 RL             0                  RLLR            0
  LL            .5                  LLRR           0
   LR           0                    LRRL          0
    RR          .5                    RRLR         .25
     RL         0                      RLRR        .25
      LR        0                       LRRR       .25
       RR       .5                       RRRL      .25
        RR      .5                        RRLL  0 Mean .156
         RL     0
          LL .5 Mean .227


        Window Size 6                      Window Size 8

LRLLRRLRRRLL                      LRLLRRLRRRLL
LRLLRR          0                 LRLLRRLR        0
 RLLRRL         0                  RLLRRLRR       .125
  LLRRLR        0                   LLRRLRRR      .125
   LRRLRR       .166                 LRRLRRRL     .125
    RRLRRR      .333                   RRLRRRLL .125 Mean .1
     RLRRRL     .166
      LRRRLL  0 Mean .095
```

Fig. 1 Examples of (average) values of local unevenness in an example sequence for window sizes 2, 4, 6, and 8

divided into all possible sequential sub-sequences of a pre-defined length and the formula is applied to all of these sub-sequences. For an illustration of what this looks like, see Fig. 1. The formula for the LU in each segment is as follows:

$$LU_w = \sqrt{\frac{(p_{\text{left}} - 0.5)^2 + (p_{\text{right}} - 0.5)^2}{2}},$$

where p is the ratio of the respective response option given in the respective window. Because in the case of only two options the two ratios are complementary, this formula can be further simplified to:

$$LU_w = \frac{\sqrt{(2 \cdot p_{\text{left}} - 1)^2}}{2}.$$

The range of values for the LU lies between 0 and 0.5, where 0 means that in the given window, the distribution is perfectly in line with the expected ratios (i.e., both choices are represented equally often, that is completely evenly) and 0.5 means that only one of the two choices is present in the given window (i.e., the sequence is as uneven as possible).

To illustrate, Fig. 1 gives an example sequence of choices and the resulting LUs, for four different window sizes of 2, 4, 6, and 8, as well as the mean LU for the sequence.

For an infinitely long random sequence, the expected mean value of the LU is, however, not 0.0, as this would imply that in every single segment the options are represented equally often, without, for example, any run-ons of the same choice. Instead, it is the average of all the potential combinations of the options when taking the order of the options into account. Figure 2 illustrates the potential response option combinations when using a window of the size 4.

This results in an ideal LU of 0.1875, as all these potential sequences have the same chance to appear in a random sequence. The ideal values for the four window sizes mentioned above are 0.25, 0.1875, 0.15625, and 0.1367188 (for window size of 2, 4, 6, and 8, respectively). Mean LUs higher than those ideal values then mean that unbalanced segments were overrepresented in the whole sequence compared to what would be expected in a random sequence. Conversely, mean LUs below those ideal values imply that balanced segments were overrepresented. From this follows that the deviation from these ideal LU values in a sufficiently long sequence can be viewed as a deviation from (ideal) randomness.

LLLL (LU=.5)
LLLR (LU=.25)
LLRL (LU=.25)
LLRR (LU= 0)
LRLL (LU=.25)
LRLR (LU= 0)
LRRL (LU= 0)
LRRR (LU=.25)
RLLL (LU=.25)
RLLR (LU= 0)
RLRL (LU= 0)
RLRR (LU=.25)
RRLL (LU= 0)
RRLR (LU=.25)
RRRL (LU=.25)
RRRR (LU=.5)

Fig. 2 All sequences that can occur for window size 4 and the respective value of local unevenness (LU). The resulting ideal LU value is then 0.1875

The present study

Our prediction is that, if free choice tasks are random generation tasks, WM manipulations should influence randomness (and also response speed) accordingly. We chose a complementary approach of both lowering and increasing WM load. WM support should then increase randomness (and LUs should be closer to ideally random LUs) and decrease RTs, while experimentally induced WM load should have the opposite effects. To achieve a decrease and an increase in WM load we (1) either displayed varying amounts of previous choices to reduce the need for participants to remember their choices (Experiment 1), or (2) introduced a concurrent n-back task of varying difficulty (Experiment 2). We then measured the (non-)randomness of the responses in a free choice task via the distance to the ideal LU and the speed of the responses. While analyses of LU are the theoretically most important ones, we also included the analysis of RTs to exclude any kinds of potential trade-offs. For example, it might be the case that participants change from a focus on more random responses to a focus on faster responses (similar to speed-accuracy trade-offs, where faster responses come with committing more errors). Thus, additionally analyzing RTs makes it possible to rule out such phenomena.

Experiment 1

Experiment 1 used a paradigm in which the participants gave free choice responses while receiving different levels of WM support in the form of arrows that display previous choices (for a similar approach, see Hadland et al. 2001). We used WM support because one potential way WM influences the ease with which participants generate random responses is by providing information (i.e., previous responses) that is used to decide which response would look more 'random' if chosen next. We predict that with growing WM support the distance from ideal LU will decrease and the RTs will shorten.

Methods

Participants

Thirty people from the Tübingen area participated for monetary compensation (Mean age = 23 years, 26 female, 4 male). All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection.

Apparatus and stimuli

Stimulus presentation and response collection happened on a PC connected to a 17-in. CRT monitor. Stimuli were a fixation circle in the middle of the screen as well as arrows, appearing within the fixation circle and, depending on block type, above it. Stimuli were white, presented against a black background. The manual responses were given with the left and right Ctrl keys on a QWERTZ keyboard.

Tasks and procedure

The task was to freely choose one of the two response options. The fixation circle was always visible during blocks slightly below the middle of the screen. After a response, an arrow indicating which response was given in the current trial appeared for 50 ms in the fixation circle. During these 50 ms, no new response could be given. In the two block types with WM support, the same arrow then appeared above the fixation circle, shifting all other already displayed
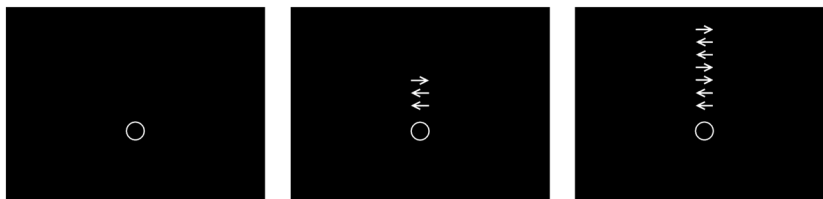
Fig. 3 Examples of the different WM support conditions. In the left panel, no WM support is given, in the middle panel three previous choices are displayed, and in the right panel seven previous choices are displayed. Not visible here are the arrows that appear within the circle for 50 ms after a response

arrows one slot upwards and, once three/seven responses were already given, displacing the oldest arrow at the top of the screen. This results in up to three or seven arrows indicating previous choices that are displayed above the fixation circle, as is illustrated in Fig. 3. The 50 ms in which no new response could be given were the only inter-trial interval. There was no time limit for responses.

Responses were collected in blocks of 500 trials with every participant performing all three block types twice, that is, in a total of six blocks. The order of the first three blocks was counterbalanced and the second set of three blocks was ordered in the reverse of the first three blocks. Participants were informed before each block how many of their previous choices would be displayed in this block.

Participants were instructed to give about equal amounts of left and right responses and to avoid patterns (e.g., alternating left and right responses or repeating sequences). There was one test session per participant which lasted about 45 min.

Design and analyses

The dependent variables were the distances from the ideally random LU (LUD) and the RTs. The independent variable was the level of WM support (0 vs. 3 vs. 7). For analyses of LUDs, however, we also analyzed four different window sizes (2 vs. 4 vs. 6 vs. 8). Accordingly, two main analyses were performed: LUDs were analyzed with a 3×4 analysis of variance (ANOVA) with WM support and window size as repeated-measures. RTs were analyzed with an ANOVA with WM support as a repeated-measure. Because we predicted decreasing RTs and LUD approaching zero with increasing WM support, we calculated Helmert contrasts on WM support (Contrast 1: no support vs. three and seven previous displayed choices; Contrast 2: three vs. seven displayed previous choices). In case of interactions between window size and the Helmert contrast, separate Helmert contrasts for each window size were calculated and are reported in the "Appendix" section.

LUDs were calculated on the whole data set once sufficient responses were given for the respective window size.

For the subsequent analyses, trials were excluded as outliers if their RTs deviated more than 2.5 SDs from the respective cell mean (calculated separately for each participant).

Results

The LUDs and average RTs (1.79% outliers) are visualized in Fig. 4 and are summarized in Table 2. For LUDs, Contrast 1 was significant and indicated a difference between conditions with and without memory support, $t(29) = 3.79$, $p = .001$, without interacting with window size, $t(29) = 1.70$, $p = .100$. However, there was no significant difference between the two memory support conditions according to Contrast 2, $t(29) = 0.36$, $p = .551$. While this contrast interacted with window size, $t(29) = 2.68$, $p = .012$, when tested separately, all contrasts were not significant, all ps ≥ 0.217 (for more details, please see the "Appendix" section).

Responses were significantly slower in the condition without WM support compared with the two other conditions, Contrast 1: $t(29) = 2.63$, $p = .013$, but there was no significant difference between the two WM support conditions, Contrast 2: $t(29) = -0.14$, $p = .886$.

Discussion

In sum, response patterns were more random and RTs shortened with the presence of WM support. No such difference was detectable between the different levels of WM support. These results can be taken as first evidence that WM plays a similar role in free choice tasks as it does for random generation tasks.

There is one potential confound in this particular experimental design: The presence of the arrows employed as WM support can be interpreted as a type of action effect (or action outcome), which conceivably differs between the no-support and the two support conditions. Furthermore, the last presented arrow was always *spatially compatible* with the selected response. Importantly, RTs are shorter when the responses produce compatible action effects compared with incompatible ones (Kunde 2001; see also; Janczyk and Lerche 2018; Janczyk et al. 2017; Koch and Kunde 2002).
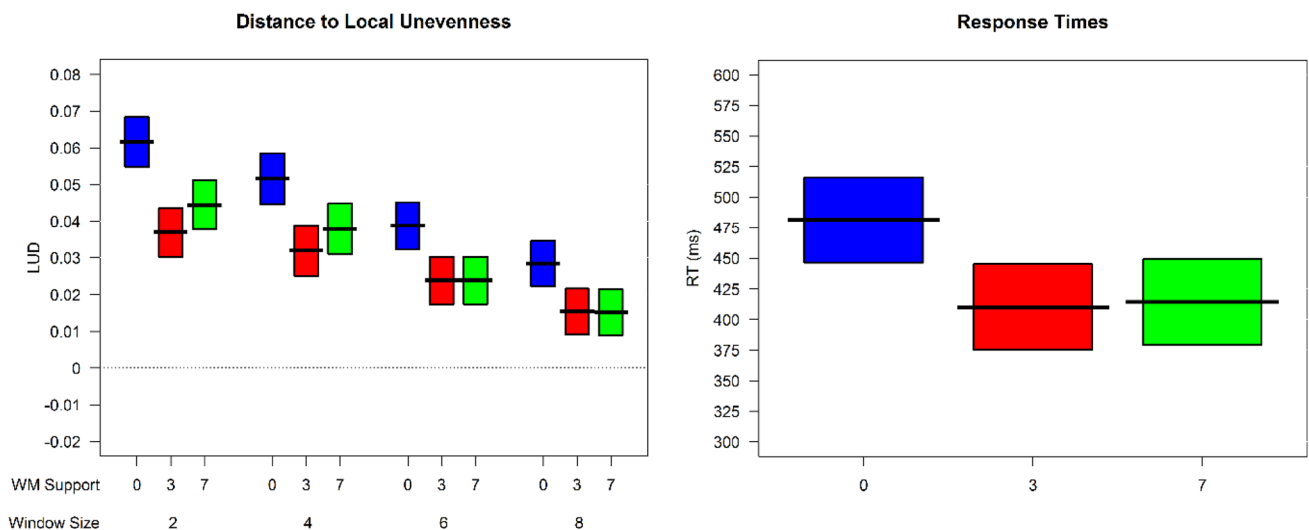
Fig. 4 Mean distances to ideal local unevenness (LUDs) (for the window sizes 2, 4, 6, and 8) and response times (RTs) in Experiment 1 for each level of working memory (WM) support. Error bars are 95% within confidence intervals (separate for all window sizes in case of the LUDs) (Loftus and Masson 1994)

Table 2 Means (and SD) of RTs in ms and LUDs for Experiment 1 for each WM support condition

	WM support: number of displayed choices		
Dependent variable	0	3	7
LUDs	0.045 (0.220)	0.027 (0.213)	0.030 (0.215)
RTs	481 (289)	410 (208)	414 (266)

At first glance, this might have contributed to the shorter RTs in the two WM support conditions. However, we believe that this argument does not pose serious problems for several reasons. First, it is important to note that in *all* conditions an immediate and compatible arrow appeared in the center of the fixation circle. Second, in the two WM support conditions, always multiple arrows were present on the screen. Thus, there would most of the time (unless the participants repeated responses multiple times) be a mixture of compatible and incompatible action effects be present what would weaken a potential impact on RTs. Third, the RT difference we observed (roughly 70 ms) is larger than the usual effects of action effect compatibility (e.g., between 20 and 50 ms in Kunde 2001). Hence, if this confound played a role in the RT results, it likely would account only for a part of the difference. Lastly, and potentially most important, it is not clear how the theoretically more important LUD results would be affected by compatible or incompatible action effects.

A further objection might be that the presence of the previous choices on the screen turned the free choice task into a "cue-dependent task". Of course, we cannot exclude that participants' used different strategies between conditions. It is the case, though, that the information about the previous choices were actually always available to the participants in form of a memory trace. The presence of the WM support arrows merely made it more accessible.

To attain more and converging evidence from a different kind of experimental manipulation, we experimentally increased WM load through an *n*-back task in Experiment 2.[3]

Experiment 2

In Experiment 2, we paired a free choice task with a WM-intensive task to induce WM load. Specifically, we alternated a free choice task with an *n*-back task for this purpose (Kirchner 1958). In all *n*-back conditions, participants had to react only under specific circumstances: For 0-back, whenever a stimulus (colored circles that were displayed left/right and above/below center on the screen) with a pre-specified color or location appeared, and for 1-, 2-, and 3-back whenever the stimulus color or location in a given trial matched that *n* trials ago. The two relevant stimulus features (color vs. location) were chosen to generalize the results and counteract potential modality-specific influences. Furthermore, this experiment completely avoids the potential confound of compatible

---

[3] Another experiment was performed in which the same type of WM support was given except the previous 0, 1, 2, and 3 choices were displayed and a block in which three symbols unrelated to the task were shown instead of previous choices. As the results were largely compatible with the other results, the experiment is not reported here.

Fig. 5 **a** Example of a sequence of displayed stimuli and fixation crosses on the screen. **b** Example of a sequence of *n*-back stimuli (free choice stimuli not displayed). In the color-based 2-back condition, only panel (4) would require a response, while in the location-based 2-back condition, panel (3) would require a response
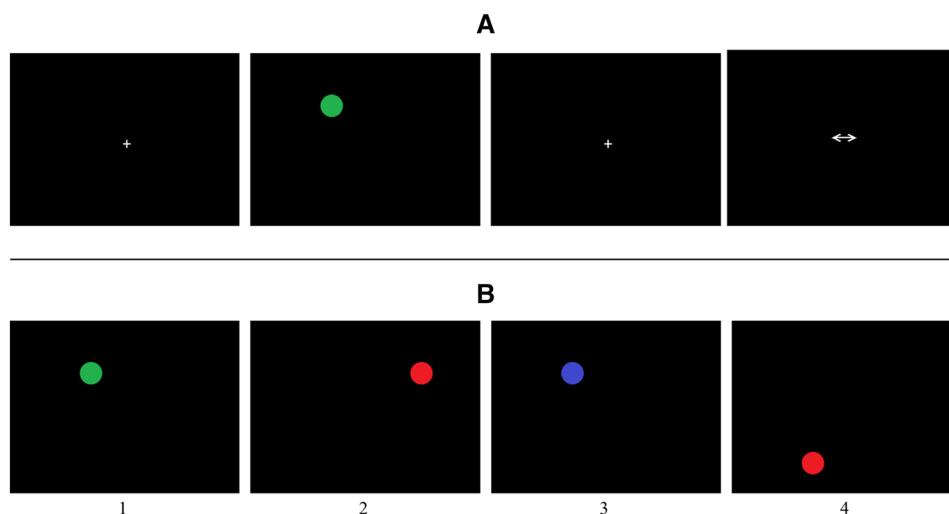
action effects from Experiment 1. Conversely to the previous experiment, we predict that with an increasing WM load, the LUDs should deviate more from zero and the RTs should increase.

Methods

Participants

Thirty-two people from the Tübingen area participated for monetary compensation or course credit (Mean age = 24 years, 22 female, 10 male). All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection.

Apparatus and stimuli

Stimulus presentation and response collection happened on a PC connected to a 17-inch CRT monitor. Stimuli were a white fixation cross in the middle of the screen, circles that could be red, green, blue, and yellow and that could appear in the top left, top right, bottom left, and bottom right location of the screen as well as a white double-headed arrow. Stimuli were presented against a black background. The responses were given with the left and right Ctrl keys on a QWERTZ keyboard (free choice task) and foot pedals placed under the feet of the participants (*n*-back task).

Tasks and procedure

Participants performed two tasks in alternation (for an illustration, see Fig. 5). In the *free choice task*, they were to freely choose one of the two manual response options in

response to the appearance of the double-headed arrow. In the *n-back task* they were to compare the current stimulus with a specific one or one that occurred *n*-trials back. A trial (with both tasks) started with the appearance of a fixation cross for 250 ms, followed by a blank screen for 250 ms, followed by the appearance of an *n*-back stimulus for up to 1500 ms or until a response was given, followed by another fixation cross and blank screen, which in turn was followed by the double-headed arrow appearing for 1500 ms or until a response was given. After this, the inter-trial interval was 250 ms.

n-back level was manipulated block-wise. Every level appeared a total of four times, with the task either requiring attention to the color or the location of the stimulus. A participant performed all levels of the color *n*-back blocks first or all of the location *n*-back tasks first, followed by the other block-type. The order of the block-types was balanced according to a Latin square for the first half and then mirrored for the second half of the experiment, in which the order of the color and location base of the task from the first half was repeated.

Participants performed in 16 blocks (four n-back levels, two modalities, repeated twice) of 61 responses each for *n*-back conditions 1, 2, and 3, and 60 responses each for the 0-back condition. They were informed before each block which criterion needed to be fulfilled for the *n*-back task in order to press the foot pedal and which foot pedal to use. Half of the participants used the left foot pedal in the first half of the experiment and the right foot pedal in the second half and vice versa for the other half of the participants. The criterion was fulfilled when either a specific color or a specific location appeared for the 0-back task, or when the color/location in the current trial matched the color/location 1, 2, or 3 trials before the current one. The course of a trial as well as an example for the *n*-back task is illustrated in Fig. 5.

Participants were instructed to give about equal amounts of right and left responses and to avoid patterns (e.g., alternating left and right responses or repeating sequences) in the free choice task. There was one test session per participant which lasted about 45 min. In cases where the distribution of free choice responses skewed too far in one direction (> 80%) data of the participant was discarded and new data collected (1 case).

Design and analyses

As in Experiment 1, the two dependent measures were the LUDs and the RTs. The independent variables were the WM load condition (0 vs. 1 vs. 2 vs. 3 in the n-back task). For analyses of LUDs we again analyzed four different window sizes (2 vs. 4 vs. 6 vs. 8). Accordingly, two main analyses were performed: LUDs were analyzed with a 4×4 ANOVA with WM load and window size as repeated-measures. RTs (and error rates for the n-back task) were analyzed with an ANOVA with WM load as a repeated-measure. Because we predicted LUDs increasingly deviating from zero with increasing WM load and increasing RTs (and error rates in the n-back task), we calculated Helmert contrasts on WM load (Contrast 1: 0-back vs. higher difficulties; Contrast 2: 1-back vs. higher difficulties; Contrast 3: 2-back vs. 3-back). In case of an interaction between window size and the Helmert contrast, separate Helmert contrasts for each window size were calculated and are reported in the "Appendix" section.

LUDs were calculated on the whole data set once sufficient responses were given for the respective window size. For the subsequent analyses, trials were excluded as outliers if their RTs deviated more than 2.5 SDs from the respective cell mean (calculated separately for each participant).

Results

In a preliminary analysis, we included the relevant stimulus feature (location vs. color; 2.86% outliers based on free choice RTs). With LUDs as the dependent variable, this additional variable did not yield a significant main effect, $F(1, 31) = 3.94$, $p = .056$, $\eta_p{}^2 = 0.11$, and the three-way interaction WM load $\times$ window size $\times$ stimulus feature was also not significant, $F(9, 279) = 0.96$, $p = .421$, $\eta_p{}^2 = 0.03$, $\varepsilon = 0.36$ (Greenhouse–Geisser estimate). With RTs as the dependent variable, also no significant main effect was observed, $F(1, 31) = 3.75$, $p = .062$, $\eta_p{}^2 = 0.11$, and the interaction WM load $\times$ stimulus feature was also not significant, $F(3, 93) = 2.74$, $p = .063$, $\eta_p{}^2 = 0.08$, $\varepsilon = 0.78$. To simplify the main analyses, we thus dropped this variable from further analyses.

Manipulation check: performance in the n-back task

We excluded 2.35% of trials as outliers (based on only the n-back task), and descriptive statistics are provided in Table 3. Contrast 1 yielded no significant result for the RTs, $t(31) = 1.91$, $p = .065$, but did yield a significant result for the error rates, $t(31) = 12.60$, $p < .001$. Contrast 2 was significant for both the RTs, $t(31) = 2.35$, $p = .025$, and the error rates, $t(31) = 13.51$, $p < .001$. Contrast 3 also was significant for both the RTs, $t(31) = 2.19$, $p = .036$, and the error rates, $t(31) = 13.10$, $p < .001$. Thus, the n-back task induced a load as expected.

Free choice task

Average LUDs and RTs (2.81% outliers) for the free choice task are visualized in Fig. 6 and are summarized in Table 3. LUDs increased with WM load for all window sizes and all contrasts were significant and in the same direction. Contrast 1 was significant, $t(31) = 3.42$, $p = .002$, but it interacted with window size, $t(31) = 3.36$, $p = .002$. Contrast 2 was also significant, $t(31) = 3.15$, $p = .004$, and it interacted with window size, $t(31) = 2.41$, $p = .022$. Finally, Contrast 3 was significant, $t(31) = 3.835$, $p = .001$, but did not interact with window size, $t(31) = 1.65$, $p = .110$. Note, however, that the descriptive pattern was the same for all window sizes despite the interactions (for more details on separate analyses per window size, please see the "Appendix" section). For RTs, Contrast 1 was not significant, $t(31) = 1.68$, $p = .104$, but RTs increased for the following levels and both Contrast 2, $t(31) = 4.94$, $p < .001$, and Contrast 3, $t(31) = 2.33$, $p = .026$, were significant.

Discussion

In summary, for the critical analyses, all contrasts were in the predicted direction and significant except for the

Table 3 Means (and *SD*s) of RTs in ms, LUDs, and percentage of errors (PE) for Experiment 2 for each WM load condition

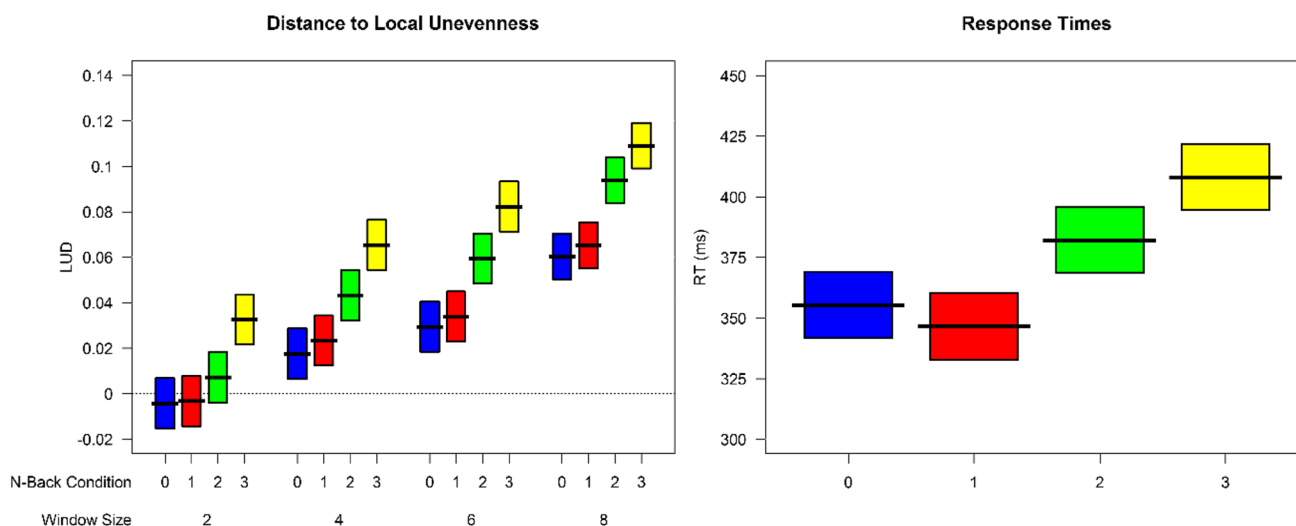|		n-back condition			
Task	Dependent variable	0	1	2	3
n-back	RT	560 (90)	600 (112)	685 (112)	753 (113)
	PE	4.60 (3.15)	6.62 (6.12)	13.69 (7.68)	26.64 (8.44)
Free choice	LUD	0.025 (0.197)	0.027 (0.199)	0.048 (0.206)	0.069 (0.204)
	RT	340 (94)	329 (93)	358 (98)	386 (118)

Fig. 6 Mean distances from ideal local unevenness (LUDs) (for the window sizes 2, 4, 6, and 8) and response times (RTs) in the free choice task in Experiment 2 for each level of working memory (WM) load. Error bars are 95% within confidence intervals (separate for all window sizes in case of the LUDs) (Loftus and Masson 1994)

difference in RTs and LUDs for the window size 2 for the contrast between the 0-back condition and higher *n*-back conditions. Thus, in line with our predictions, randomness (and RTs) in the free choice task decreased with increasing WM load. These results again suggest that free choice tasks are similar to random generation tasks.

General discussion

In this study, our participants performed free choices combined with either a WM support manipulation (Exp. 1) or a WM load manipulation (Exp. 2). This was done to investigate whether the impact of these manipulations on the patterns in free choices is the same as the impact of such manipulations on random generation tasks. To support WM, we displayed the previous choices that the participants made. To increase WM load, we used a concurrent *n*-back task.

Summary of results

In both experiments, the direction of the results was consistent with the idea that free choice tasks are related to random generation tasks: overwhelmingly, lack of WM support as well as higher WM load led to responses in which the LUDs were farther away from what would be expected in a random sequence of choices, which were also slower. More specifically, the absolute values of the LUDs moved in a more positive direction with less WM support/more WM load, suggesting that the proportion of sequences that are less balanced increased (e.g., L-L-R-L, R-R-R-R for a window size of 4).

Theoretical implications

In light of these results, we tentatively suggest that free choice tasks (as they are used in contemporary research) are at the very least related, if not outright identical, to random generation tasks, giving support to ideas expressed, for example, by Frith (2013) or Schüür and Haggard (2011): that free choice tasks are not what they are often thought to be. Frith claimed that free choice tasks are essentially random generation tasks, while Schüür and Haggard claimed that free choice tasks are either underdetermined or determined by uncontrolled internal cues like the preceding choices. Both ideas are compatible with our results: Hindering the maintenance of a memory trace of previous choices leads to responses that are 'less random' and also slower.

This potentially has wide-reaching implications for the literature on self-generated action. Assume that free choice tasks are in fact random generation tasks (as our results suggest). At least two cases can be distinguished: First, if one commits to the idea that self-generated-ness of actions must exclude all aspects of random generation, our results imply that free choice tasks do not operationalize self-generated action. Second, and in contrast, if one assumes that randomness is an inherent component of self-generated actions, then unfortunately the role of free choice tasks is even more unclear, because even without any extra cognitive load it is unclear whether the resulting sequence of actions is truly random.[4] This assumption seems not be universal among

---

[4] In fact, LUD was different from zero in most of the conditions of our experiments.

researchers though. Passingham et al. ([2010], p. 18), for example, mention as a condition for self-generated action that "One action can serve as a cue for the next action". In other words, one action is not independent from previous actions.

Limitations

An intrinsic limitation for every investigation into randomness of responses is the requirement to choose which measure of randomness to use. This choice effectively determines which kinds of patterns can be detected. It is always possible though that participants chose a different non-random production strategy for choosing responses that the researchers in question did not take into account. In our case, we chose a measure of (non-)randomness that essentially measures the proportion of different levels of balancedness in the response strings and whether they skew more towards balanced or unbalanced strings. Two weaknesses of this measure are that it cannot detect the order of responses within one window nor their identity. The string L-L-R-L looks, from a LU perspective, the same as the strings R-R-L-R and L-R-L-L: all result in LU = 0.25. However, this limitation would only pose serious problems if we had observed no differences in randomness between our different conditions.

Another issue is that WM manipulations affect the RTs of tasks involving higher cognitive processes of any kind. This makes a pure RT analysis not diagnostic with regard to whether a task is a random generation task. However, there is no reason to assume that the detected randomness of a task that is not a random generation task would suffer from WM load or benefit from WM support. This supports the interpretation of the present results as indicative of free choice tasks being random generation tasks. While this interpretation relies on drawing an analogy between free choice tasks and random generation tasks, we can at present only speculate about the specific mechanisms behind our results. One example of a plausible candidate mechanism known from the random generation literature, is the inhibition of prepotent associates (e.g., Towse and Neil [1998]). Easier monitoring of ongoing choices could make it easier to identify and suppress these stereotypical responses (e.g., fewer repetitions than would be appropriate for a random sequence).

Conclusion

We investigated whether LU, as a measure of randomness, based on responses from free choice tasks and RTs in this task is affected by WM support and load in a similar way as random generation tasks are. In short, we observed that they are and conclude that free choice tasks are related to or identical to random generation tasks. This potentially casts doubt on some types of investigations into self-generated action. The present study also provides evidence that random (response) generation is one of the processes that contribute to the mean RT difference

between free and forced choice tasks, a difference that was tentatively attributed to the non-accumulation time by Naefgen et al. ([2017]). It is an open question whether this is the full extent of what makes up this difference or if there are other, additional processes that differentiate free and forced choice tasks.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Appendix

For completeness, we report the Helmert contrasts separately for each window size in this "Appendix" section.

Experiment 1

The descriptive results of the following analyses are summarized in Table 4.

Contrast 1:

- Window Size 2: $t(29) = 4.44$, $p < .001$
- Window Size 4: $t(29) = 3.48$, $p = .002$
- Window Size 6: $t(29) = 3.11$, $p = .004$
- Window Size 8: $t(29) = 2.82$, $p = .009$

Contrast 2:

- Window Size 2: $t(29) = 1.26$, $p = .217$
- Window Size 4: $t(29) = 0.96$, $p = .344$
- Window Size 6: $t(29) = 0.00$, $p = .997$
- Window Size 8: $t(29) = 0.04$, $p = .963$

Experiment 2

The descriptive results of the following analyses are summarized in Table 5.

Table 4 Means (and *SD*s) of distances from ideal local unevenness (LUDs) for window sizes 2, 4, 6, and 8 for Experiment 1 for each working memory (WM) support condition

	WM support: number of displayed choices		
Window size	0	3	7
2	0.062 (0.306)	0.037 (0.066)	0.044 (0.067)
4	0.052 (0.213)	0.032 (0.057)	0.038 (0.059)
6	0.039 (0.174)	0.024 (0.053)	0.024 (0.056)
8	0.028 (0.152)	0.015 (0.051)	0.015 (0.052)

Table 5 Means (and *SD*s) of distances from ideal local unevenness (LUDs) for window sizes 2, 4, 6, and 8 for Experiment 2 for each working memory (WM) load condition

	n-back			
Window size	0	1	2	3
2	− 0.003 (0.116)	− 0.004 (0.123)	0.007 (0.131)	0.031 (0.123)
4	0.018 (0.124)	0.023 (0.126)	0.043 (0.135)	0.064 (0.128)
6	0.030 (0.127)	0.034 (0.130)	0.059 (0.138)	0.081 (130)
8	0.060 (0.118)	0.065 (0.119)	0.094 (0.125)	0.109 (0.121)

Contrast 1:

- Window Size 2: $t(31) = 1.99$, $p = .056$
- Window Size 4: $t(31) = 3.35$, $p = .002$
- Window Size 6: $t(31) = 3.77$, $p = .001$
- Window Size 8: $t(31) = 4.23$, $p < .001$

Contrast 2:

- Window Size 2: $t(31) = 2.18$, $p = .017$
- Window Size 4: $t(31) = 2.88$, $p = .007$
- Window Size 6: $t(31) = 3.51$, $p = .001$
- Window Size 8: $t(31) = 3.85$, $p = .001$

Contrast 3:

- Window Size 2: $t(31) = 3.95$, $p < .001$
- Window Size 4: $t(31) = 3.84$, $p = .001$
- Window Size 6: $t(31) = 3.92$, $p < .001$
- Window Size 8: $t(31) = 2.81$, $p = .008$

References

Azouvi P, Jokic C, Der Linden MV et al (1996) Working memory and supervisory control after severe closed-head injury. A study of dual task performance and random generation. J Clin Exp Neuropsychol 18:317–337. https://doi.org/10.1080/01688639608408990

Baddeley AD (1962). Some factors influencing the generation of random letter sequences. Med Res Council Appl Psychol Unit Rep. 422/62

Baddeley AD (1966) The capacity for generating information by randomization. Q J Exp Psychol 18:119–129. https://doi.org/10.1080/14640746608400019

Bausenhart KM, Rolke B, Seibold VC, Ulrich R (2010) Temporal preparation influences the dynamics of information processing: evidence for early onset of information accumulation. Vis Res 50:1025–1034. https://doi.org/10.1016/j.visres.2010.03.011

Berlyne DE (1957) Conflict and choice time. Br J Psychol 48:106–118. https://doi.org/10.1111/j.2044-8295.1957.tb00606.x

Bermeitinger C, Hackländer RP (2018) Response priming with motion primes: negative compatibility or congruency effects, even in free-choice trials. Cogn Process. https://doi.org/10.1007/s10339-018-0858-5

Brass M, Haggard P (2008) The what, when, whether model of intentional action. Neurosci 14:319–325. https://doi.org/10.1177/1073858408317417

Cooper RP, Karolina W, Davelaar EJ (2012) Differential contributions of set-shifting and monitoring to dual-task interference. Q J Exp Psychol 63: 587–612

Dror IE, Basola B, Busemeyer JR (1999) Decision making under time pressure: An independent test of sequential sampling models. Mem Cognit 27:713–725. https://doi.org/10.3758/BF03211564

Ebert JP, Wegner DM (2011) Mistaking randomness for free will. Conscious Cogn 20:965–971. https://doi.org/10.1016/j.concog.2010.12.012

Elsner B, Hommel B (2001) Effect anticipation and action control. J Exp Psychol Hum Percept Perform 27:229–240. https://doi.org/10.1037/0096-1523.27.1.229

Frith C (2013) The psychology of volition. Exp Brain Res 229:289–299. https://doi.org/10.1007/s00221-013-3407-6

Gaschler R, Nattkemper D (2012) Instructed task demands and utilization of action effect anticipation. Front Psychol 3:. https://doi.org/10.3389/fpsyg.2012.00578

Gozli DG, Huffman G, Pratt J (2016) Acting and anticipating: Impact of outcome-compatible distractor depends on response selection efficiency. J Exp Psychol Hum Percept Perform 42:1601–1614. https://doi.org/10.1037/xhp0000238

Grice GR (1968) Stimulus intensity and response evocation. Psychol Rev 75:359–373

Hadland KA, Rushworth MFS, Passingham RE et al (2001) Interference with performance of a response selection task that has no working memory component: an rTMS comparison of the dorsolateral prefrontal and medial frontal cortex. J Cogn Neurosci 13:1097–1108. https://doi.org/10.1162/089892901753294392

Herwig A, Waszak F (2009) Short article: intention and attention in ideomotor learning. Q J Exp Psychol 62:219–227. https://doi.org/10.1080/17470210802373290

Herwig A, Waszak F (2012) Action-effect bindings and ideomotor learning in intention- and stimulus-based actions. Front Psychol 3. https://doi.org/10.3389/fpsyg.2012.00444

Herwig A, Prinz W, Waszak F (2007) Two modes of sensorimotor integration in intention-based and stimulus-based actions. Q J Exp Psychol 60:1540–1554. https://doi.org/10.1080/17470210601119134

Heuer H, Kohlisch O, Klein W (2005) The effects of total sleep deprivation on the generation of random sequences of key-presses, numbers and nouns. Q J Exp Psychol Sect A 58:275–307. https://doi.org/10.1080/02724980343000855

Heuer H, Janczyk M, Kunde W (2010) Random noun generation in younger and older adults. Q J Exp Psychol 63:465–478. https://doi.org/10.1080/17470210902974138

Huffman G, Gozli DG, Hommel B, Pratt J (2018) Response preparation, response selection difficulty, and response-outcome learning. Psychol Res 1–11. https://doi.org/10.1007/s00426-018-0989-4

Jahanshahi M, Dirnberger G, Fuller R, Frith CD (2000) The role of the dorsolateral prefrontal cortex in random number generation: a study with positron emission tomography. NeuroImage 12:713–725. https://doi.org/10.1006/nimg.2000.0647

Janczyk M, Lerche V (2018) A diffusion model analysis of the response-effect compatibility effect. J Exp Psychol Gen. https://doi.org/10.1037/xge0000430

Janczyk M, Pfister R, Kunde W (2012a) On the persistence of tool-based compatibility effects. Z Für Psychol 220:16–22. https://doi.org/10.1027/2151-2604/a000086

Janczyk M, Pfister R, Crognale MA, Kunde W (2012b) Effective rotations: Action effects determine the interplay of mental and manual rotations. J Exp Psychol Gen 141:489–501. https://doi.org/10.1037/a0026997

Janczyk M, Pfister R, Hommel B, Kunde W (2014) Who is talking in backward crosstalk? Disentangling response- from goal-conflict in dual-task performance. Cognition 132:30–43. https://doi.org/10.1016/j.cognition.2014.03.001

Janczyk M, Nolden S, Jolicoeur P (2015a) No differences in dual-task costs between forced- and free-choice tasks. Psychol Res 79:463–477. https://doi.org/10.1007/s00426-014-0580-6

Janczyk M, Dambacher M, Bieleke M, Gollwitzer PM (2015b) The benefit of no choice: goal-directed plans enhance perceptual processing. Psychol Res 79:206–220. https://doi.org/10.1007/s00426-014-0549-5

Janczyk M, Durst M, Ulrich R (2017) Action selection by temporally distal goal states. Psychon Bull Rev 24:467–473. https://doi.org/10.3758/s13423-016-1096-4

Jenkins IH, Jahanshahi M, Jueptner M et al (2000) Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. Brain J Neurol 123:1216–1228

Keller PE, Wascher E, Prinz W et al (2006) Differences between intention-based and stimulus-based actions. J Psychophysiol 20:9–20. https://doi.org/10.1027/0269-8803.20.1.9

Kirchner WK (1958) Age differences in short-term retention of rapidly changing information. J Exp Psychol 55:352–358. https://doi.org/10.1037/h0043688

Koch I, Kunde W (2002) Verbal response-effect compatibility. Mem Cogn 30:1297–1303. https://doi.org/10.3758/BF03213411

Kühn S, Elsner B, Prinz W, Brass M (2009) Busy doing nothing: Evidence for nonaction-effect binding. Psychon Bull Rev 16:542–549. https://doi.org/10.3758/PBR.16.3.542

Kunde W (2001) Response-effect compatibility in manual choice reaction tasks. J Exp Psychol Hum Percept Perform 27:387–394. https://doi.org/10.1037/0096-1523.27.2.387

Kunde W, Pfister R, Janczyk M (2012) The locus of tool-transformation costs. J Exp Psychol Hum Percept Perform 38:703–714. https://doi.org/10.1037/a0026315

Linser K, Goschke T (2007) Unconscious modulation of the conscious experience of voluntary control. Cognition 104:459–475. https://doi.org/10.1016/j.cognition.2006.07.009

Loftus GR, Masson ME (1994) Using confidence intervals in within-subject designs. Psychon Bull Rev 1:476–490. https://doi.org/10.3758/BF03210951

Martin-Löf P (1966) The definition of random sequences. Inf Control 9:602–619. https://doi.org/10.1016/S0019-9958(66)80018-9

Miyake A, Friedman NP, Emerson MJ et al (2000) The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cognit Psychol 41:49–100. https://doi.org/10.1006/cogp.1999.0734

Naefgen C, Dambacher M, Janczyk M (2017) Why free choices take longer than forced choices: evidence from response threshold manipulations. Psychol Res 1–14. https://doi.org/10.1007/s00426-017-0887-1

Passingham RE, Bengtsson SL, Lau HC (2010) Medial frontal cortex: from self-generated action to reflection on one's own performance. Trends Cogn Sci 14:16–21. https://doi.org/10.1016/j.tics.2009.11.001

Pfister R, Kunde W (2013) Dissecting the response in response–effect compatibility. Exp Brain Res 224:647–655. https://doi.org/10.1007/s00221-012-3343-x

Pfister R, Kiesel A, Melcher T (2010) Adaptive control of ideomotor effect anticipations. Acta Psychol (Amst) 135:316–322. https://doi.org/10.1016/j.actpsy.2010.08.006

Pfister R, Kiesel A, Hoffmann J (2011) Learning at any rate: action–effect learning for stimulus-based actions. Psychol Res 75:61–65. https://doi.org/10.1007/s00426-010-0288-1

Rescher N (2005) Cosmos and Logos: Studies in Greek Philosophy. Ontos Verlag

Schüür F, Haggard P (2011) What are self-generated actions? Conscious Cogn 20:1697–1704. https://doi.org/10.1016/j.concog.2011.09.006

Towse JN, Neil D (1998) Analyzing human random generation behavior: a review of methods used and a computer program for describing performance. Behav Res Methods Instrum Comput 30:583–591. https://doi.org/10.3758/BF03209475

Waszak F, Wascher E, Keller P et al (2005) Intention-based and stimulus-based mechanisms in action selection. Exp Brain Res 162:346–356. https://doi.org/10.1007/s00221-004-2183-8

Wirth R, Janczyk M, Kunde W (2018) Effect monitoring in dual-task performance. J Exp Psychol Learn Mem Cogn 44:553–571. https://doi.org/10.1037/xlm0000474

Wolfensteller U, Ruge H (2011) On the timescale of stimulus-based action–effect learning. Q J Exp Psychol 64:1273–1289. https://doi.org/10.1080/17470218.2010.546417