# The similarities and differences of free and forced choice tasks

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Christoph Naefgen

aus Neuss

Tübingen

2018

Tag der mündlichen Qualifikation: 09.02.2019
Dekan: Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter: Prof. Dr. Markus Janczyk
2. Berichterstatter: Prof. Dr. Hartmut Leuthold

Index

Index ........ 3

Abstract ........ 4

1 What are free choice tasks? ........ 7

1.1 What do Free Choice Tasks look like? ........ 8

1.2 Which related tasks are not meant here? ........ 12

2 For what are free choice tasks used? ........ 15

2.1 Self-generated and externally triggered actions. ........ 15

2.2 Conflict tasks and tasks with ambiguous stimulus-response links. ........ 19

3 How similar are free and forced choice tasks? ........ 24

3.1 How are free and forced choice tasks dissimilar? ........ 24

3.2 How are free and forced choice tasks similar? ........ 26

3.3 Evaluation of similarities. ........ 29

4 Research questions ........ 30

5 Smaller backward crosstalk effect for free choice tasks are not the result of immediate conflict adaptation ........ 31

6 Free choices compared to forced choices: Just underdetermined or is there an additional process? (Why free choices take longer than forced choices: evidence from response threshold manipulations) ........ 36

7 Is the additional process a random generation task? (Free Choice Tasks as Random Generation Tasks: An Investigation through Working Memory Manipulations) ........ 42

8 Discussion ........ 48

8.1 Central results and conclusions ........ 48

8.2 Implications and outlook ........ 50

8.3 Summary ........ 57

References ........ 58

Appendix A – Study 1 ........ 73

Appendix B – Study 2 ........ 104

Appendix C – Study 3 ........ 143

# Abstract

Free choice tasks are tasks in which more than one response is considered correct, while in forced choice tasks only one response is considered correct. They are often used in conjunction to investigate differences between self-generated (free choice) and externally triggered (forced choice) actions. The general purpose of the present work was to investigate what free choice tasks are, both in themselves and in contrast with forced choice tasks. This was investigated over the course of three studies. Study 1 was a follow-up study to Naefgen, Caissie, and Janczyk (2017), which in turn was an investigation of the mechanisms behind the backward crosstalk effect (BCE). The BCE is an interference effect that appears in dual-tasking situations and refers to the phenomenon that response times in the first task are influenced by whether the two tasks are compatible or incompatible on (theoretically) any dimension. Naefgen, Caissie et al. investigated the role of stimulus-response links in the BCE, finding reduced BCEs when one of the tasks was a free choice task. The alternative explanation for these results that Study 1 investigated was that this reduction was due to conflict adaptation in response to the presence of free choice tasks. As the BCE in Study 1 was reduced neither in trials following free choice Task 1 trials nor with higher proportions of free choice Task 1 trials in a block, the alternative explanation was rejected. In Study 2, the common observation that free choice tasks have slower responses than forced choice tasks was investigated. Within a sequential sampling framework, in which evidence is noisily accumulated towards decision thresholds, the crossing of which causes a response to be emitted, the mean response time difference was sought to be attributed to either differences in the speed of evidence accumulation or differences in the time of non-accumulation time. This was done by manipulating the decision thresholds with proportions of catch trials in Experiment 1 and time pressure in Experiments 2 and 3. If the difference is due to different evidence accumulation speeds, the response time difference should change. As it did not, the difference was attributed to a difference in non-accumulation time, possibly suggesting that free choice tasks involve an additional process. In Study 3, the question whether free choice tasks are random generation tasks was investigated. This was done by manipulating the working memory load and observing whether the randomness of the choices changes in a manner consistent with random generation tasks. As both a manipulation supporting the working memory and one adding working memory load had effects consistent with random generation tasks, it was concluded that free choice tasks are at least similar to random generation tasks. The implications of the results of all three studies for free choice tasks and their uses are discussed in the General Discussion.

# Enclosed Publications

**Journal Publications**

Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: evidence from response threshold manipulations. *Psychological Research*, *82*(6), 1039–1052. doi:10.1007/s00426-017-0887-1

Naefgen, C., & Janczyk, M. (2018a). Free choice tasks as random generation tasks: an investigation through working memory manipulations. *Experimental Brain Research*, *236*(8), 2263–2275. doi:10.1007/s00221-018-5295-2

Naefgen, C., & Janczyk, M. (2018b). Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation. *Cognitive Processing*. doi:10.1007/s10339-018-0887-0

**Conference contributions:**

Naefgen, C., Dambacher, M., & Janczyk, M. (2016, March). *Why Free Choices Take Longer Than Forced Choices.* Presentation given at 58th Conference of Experimental Psychologists, Heidelberg, Germany.

Naefgen, C., Dambacher, M., & Janczyk, M. (2016, November). *Why Free Choices Take Longer Than Forced Choices.* Poster presented at Psychonomic Society's 57th Annual Meeting, Boston, Massachusetts.

Naefgen, C. & Janczyk, M. (2017, March). *Working Memory Support Facilitates the Generation of Free Choices.* Presentation given at 59th Conference of Experimental Psychologists, Dresden, Germany.

Naefgen, C. & Janczyk, M. (2017, September). *Working Memory Support Facilitates the Generation of Free Choices.* Poster presented at the 20th Conference of the European Society for Cognitive Psychology, Potsdam, Germany.

Naefgen, C. & Janczyk, M. (2018, March). *Working Memory Capacity Facilitates the Generation of Free Choices.* Presentation given at 60th Conference of Experimental Psychologists, Marburg, Germany.

# Author Contributions

Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: evidence from response threshold manipulations. *Psychological Research*, *82*(6), 1039–1052. doi:10.1007/s00426-017-0887-1

| Author | Author position | Scientific ideas % | Data generation % | Analysis & Interpretation | Paper writing % |
|---|---|---|---|---|---|
| Naefgen, C | 1 | 33 | 75 | 50 | 50 |
| Dambacher, M | 2 | 33 | 0 | 25 | 25 |
| Janczyk, M | 3 | 33 | 25 | 25 | 25 |
| Title of paper | | Why free choices take longer than forced choices: evidence from response threshold manipulations | | | |
| Status in publication process | | Published | | | |

Naefgen, C., & Janczyk, M. (2018a). Free choice tasks as random generation tasks: an investigation through working memory manipulations. *Experimental Brain Research*, *236*(8), 2263–2275. doi:10.1007/s00221-018-5295-2

| Author | Author position | Scientific ideas % | Data generation % | Analysis & Interpretation | Paper writing % |
|---|---|---|---|---|---|
| Naefgen, C | 1 | 75 | 75 | 75 | 75 |
| Janczyk, M | 2 | 25 | 25 | 25 | 25 |
| Title of paper | | Free choice tasks as random generation tasks: an investigation through working memory manipulations | | | |
| Status in publication process | | Published | | | |

Naefgen, C., & Janczyk, M. (2018b). Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation. *Cognitive Processing*. doi:10.1007/s10339-018-0887-0

| Author | Author position | Scientific ideas % | Data generation % | Analysis & Interpretation | Paper writing % |
|---|---|---|---|---|---|
| Naefgen, C | 1 | 40 | 75 | 75 | 75 |
| Janczyk, M | 2 | 60 | 25 | 25 | 25 |
| Title of paper | | Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation | | | |
| Status in publication process | | Published | | | |

# 1 What are free choice tasks?

*“In ancient Rome there was a poem about a dog who had two bones. He picked at one, he licked the other. He went in circles till he dropped dead.”* –Devo – Freedom of Choice

When Berlyne used the terms free and forced choice task for two specific kinds of experimental task in 1957, he intended to use the tasks to investigate the role of conflict in the formation of responses. Specifically, he wanted to investigate four variables that he posited constituted conflict: The amount of choices, the absolute strength of the choices, the relative strength of the choices (and how equal they are in strength) and how mutually exclusive the choices are. To this end he ran three response time (RT) experiments. In all of them, participants were to react to lights of different colors (and sometimes positions) in different ways: If only one color of light appears, there was a specific response they were to give (called forced choice task). Multiple colors of light appearing at the same time, on the other hand, indicated that either of the possible responses was to be given (called free choice task). He varied the responses’ mutual exclusivity by manipulating whether responses were given by two switches with one position or one switch with two positions. He further varied the absolute strength of the stimuli by, in one condition, doubling the amount of lights that served as stimuli in one experiment, as well as by varying the luminosity of the lights in another. Lastly, he varied the amount of response options by either having two or four relevant response options in the third experiment. With the exception of the last one, all manipulations affected the free and forced choice task RTs differently: Having only a single switch with two possible positions elicited slower forced choice responses but did not affect free choice RTs. Doubling the amount of lights resulted in faster forced choice responses but slower free choice responses while for the luminosity of the lights brighter lights always resulted in faster responses, but this was less pronounced for the free choice tasks. Lastly, more response options resulted in slower RTs, but not differently so for the two task types. Overall, in all experiments, free choice tasks had slower responses than forced choice tasks. Together, Berlyne interpreted these results as confirmation of the theoretical assumption that higher levels of conflict (here in the sense of competition of response tendencies) lead to longer RTs.

While the theoretical focus of Berlyne’s work was on the effects of conflict on RTs, the types of tasks and the terminology he used have been used in subsequent psychological inquiries of several types. The present work’s aim is to provide insight into the differences between free and forced choice tasks, both in a more technical sense of attributing the descriptive RT mean difference between them to parameters within a sequential sampling framework and in a sense of searching for potential cognitive mechanisms underlying these differences in parameters. To that purpose, the rest of Chapter 1 is divided into two threads: Section 1.1 will provide a definition of free and forced choice tasks as the terms are used here, which will also include a rough overview over some of the different variants of the tasks. Section 1.2 will provide a short overview over tasks and terms that are similar but which are not the subject of this work. The following two chapters will be about situating free and forced choice tasks within the literature, including an overview over the purposes for which they have been used, criticisms of these uses (Chapter 2) and an overview over fundamental similarities and dissimilarities of the two task types (Chapter 3), respectively. Chapter 4 will summarize the research questions central to the studies summarized in Chapters 5 through 7. The answers to these questions, the wider theoretical implications of these answers as well as an outlook for future research are discussed in Chapter 8.

## 1.1 What do Free Choice Tasks look like?

The defining features of free choice tasks as they will be talked about here are to be understood in contrast to the more commonly used standard forced choice task. In these, there is an unambiguous mapping of each stimulus to one response (i.e. each stimulus contains complete information about which response is correct and which response(s) is (are) incorrect). In contrast, free choice stimuli are ambiguous. This means that two or more responses to the stimulus are considered equally correct. As there are many different ways that this can be achieved on an operationalizational level, here are just a few examples, which are illustrated in Figure 1:

A. Two forced choice stimuli are presented simultaneously (e.g. Berlyne, 1957, using red and green lights, lit up separately for forced choice trials and together for free choice trials)

B. A distinct third type of stimulus is presented after an instruction to choose freely in response to it (e.g. Janczyk, Dambacher, Bieleke, & Gollwitzer, 2015, using stimuli in two colors as forced choice task and a third color of stimulus as a free choice task)

C. Similar to the previous type, multiple stimuli (as opposed to only one) are to be responded to freely (e.g. Elsner & Hommel, 2001, with auditory stimuli that previously were action consequences and in the testing phase are the action triggers, i.e. stimuli demanding a response)

D. Similar to type A, here the stimulus carries spatial information about the expected response and, for the free choice stimulus, multiple stimuli are presented. Participants are instructed to respond with the key(s) spatially (left, middle, right) corresponding to the highlighted square(s) (e.g. Janczyk, Nolden, & Jolicoeur, 2015, Experiments 2+3)

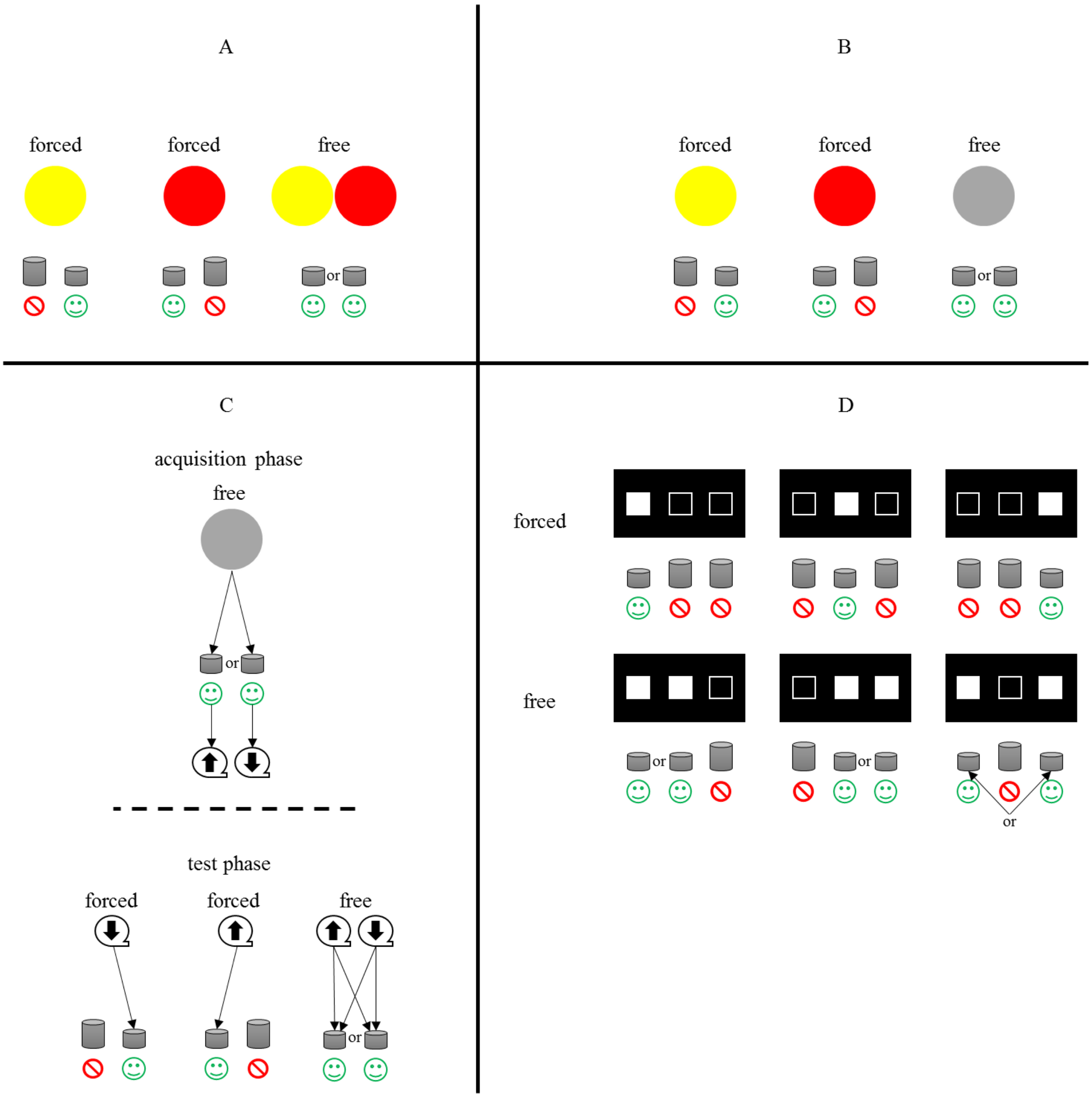

***Figure 1.*** Illustration of types of free and forced choice task stimuli. In panel A, forced choice stimuli differ in one stimulus feature (here: color) and the free choice stimulus is a composite of two forced choice stimuli. In panel B, forced choice stimuli again differ in one stimulus feature from each other and the free choice stimulus presents a third stimulus feature variant. In panel C, in the acquisition phase there is only a free choice stimulus. Different responses to this free choice stimulus lead to different action effects (here: high- or low-pitched sounds). In the test phase, high or low pitched sounds can either serve as forced choice stimuli, indicating the responses of which they previously were the effects, or they can serve as free choice stimuli. Note that these two uses in the test phase cannot happen in the same experiment. In Panel D, the stimuli carry spatial information (left, middle, right) about which response(s) are correct. If multiple boxes are highlighted, all corresponding responses are considered correct.

As part of the instructions of a given experiment, participants may or may not be instructed to try to respond in similar frequencies with all response options to the free choice stimuli and that they should try to avoid responding with obvious patterns. Sometimes, participants may even explicitly be instructed to choose response options randomly. Other times, choosing spontaneously is emphasized. An overview over the different kinds of emphases that are given in instructions is provided in Table 1. It is somewhat ambiguous on a conceptual level if tasks in which randomness is explicitly instructed should be counted among the variations of free choice tasks for the purposes of the present work. Arguably, intentional action (see Chapter 2) should be differentiated from mere random 'action'. Therefore, at this point, tasks whose instructions emphasize random generation are excluded from the working definition of free choice tasks used here (but see also Chapter 7 for an investigation into whether free choice tasks involve a component of random generation).

***Table 1.*** Illustrative examples of different instructions for free choice tasks. (adapted from Naefgen & Janczyk, 2018a)

| 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, Rushworth, Passingham, Jahanshahi, & Rothwell, 2001, see also Waszak, Wascher, Keller, Koch, Aschersleben, Rosenbaum, & Prinz, 2005; Elsner & Hommel, 2001) |
| Similar to random response instructions | Overlap between instructions and definitions of randomness<br>**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 & Kunde, 2013; see also Linser & Goschke, 2007) |
| Emphasizing spontaneity/freedom of choice | No mention of randomness<br>No particular overlap in instructions with definitions of randomness<br>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, Kiesel, & Melcher, 2010; see also Herwig, Prinz, & Waszak, 2007) |
| None reported | No explication of instructions present in the text | "When lights of both colours appeared, either response, but not both, was to be performed." (Berlyne, 1957) |

## 1.2 Which related tasks are not meant here?

There are several tasks used to investigate concepts related (either theoretically or operationalizationally) to free choice tasks as defined above. In order to clearly delineate the subject of the present work, this section will provide an overview over such tasks which are *not* the subject of this work, split into conceptually related tasks and tasks with the same name but no conceptual relation.

To identify conceptually related tasks, an exemplary framework for action will be briefly described and free choice tasks as used in the present work will be identified within that framework, by extension excluding tasks falling under different categories within it. Examples of the tasks will be presented.

The framework is the "What, When, Whether" model of intentional action by Brass and Haggard (2008). As the name suggests, this model identifies three different aspects of intentional action, which can be subject to choice: (1) The What component is concerned with the type of action that is to be executed. Brass and Haggard mention free selection tasks (which in the present text are called free choice tasks) here themselves already, but warn that they might be conflict resolution tasks instead of free choice tasks. This would be in line with Berlyne's (1957) original conceptualization of free choice tasks (but see Chapter 2 for a discussion of the different uses of free choice tasks). (2) The When component is concerned with the point in time at which the action is to be executed. The perhaps most well-known example of a task in which this component is chosen (relatively) freely can be found in the Libet, Gleason, Wright, and Pearl (1983) study, in which participants could choose freely when to give a certain response while time was displayed in the form of a clock face. The actual timing of the response and when participants were aware of their choice showed discrepancies, spurring some debates on the nature of consciousness in action. (3) Lastly, the Whether component is concerned with the question if the action will be actually executed, after the What and When are decided. One example can be found in the study by Brass and Haggard (2007) in which they instructed participants to sometimes (without prompting) decide to withhold their response to a stimulus (as opposed to providing no-go stimuli, which by themselves forbid reacting). Within this framework, free choice tasks as used here are distinguished by their freedom of choice in the What component, as the specifics of When (as soon as possible after the stimulus appeared) and Whether (always), but not of What (e.g., left or right response) are defined by the stimuli. One further specification to this is that not all tasks in which the What component is not completely specified by the stimulus necessarily fall within the category of free choice tasks as used here. One category of tasks that shows this are cued random generation tasks in which a random number is to be generated in response to a stimulus. Heinemann, Pfister, and Janczyk (2013) for example used a random generation task in which, in response to a stimulus, a number between 1 and 9 was to be said out loud. While this task is clearly not a free choice task as defined here (as the emphasis was on random generation as opposed to freely choosing a response), even if, hypothetically, the instructions did not emphasize random generation and emphasized free choice instead, it would still be somewhat problematic to fit into the particular framework used here: There is a spectrum between providing a specific set of responses (which would be a free choice task) and leaving the responses completely open (which would not be a free choice task as defined here) with higher numbers of response options shifting the task from the former more towards the latter. Of course, nine response options are still far away from an unlimited set of response options, like, for example, naming any positive whole number would be.

Other tasks with a similar name but no close conceptual relationship with the free and forced choice tasks discussed in the rest of this work include:

- The two-alternative forced choice (2AFC) task (e.g., McKenzie, Wixted, Noelle, & Gyurjyvan, 2001) needs to be distinguished from forced choice tasks as used here. 2AFC tasks require that there are always two stimuli presented for each trial of which the correct one is to be selected while for a forced choice task only one stimulus needs to be present at any given time and where different stimuli correspond to different correct responses. In some ways, the 2AFC task could be argued to be a special case of a forced choice task, as the two stimuli presented in a 2AFC trial could be conceived of as one composite stimulus to which only one of two responses is considered correct. And indeed, often in the literature the two terms are used somewhat interchangeably, despite the different specific definitions. Nevertheless, the terminological distinction as it relates to this work should be noted.

- In various fields the term “free choice” is used, but often in a more informal manner. Two examples are ethology and education. In ethology, tasks in which more than one choice is presented might be called free choice tasks. One example can be found in Arvola and Forsander (1961), who compared alcohol consumption to water consumption ratios when both are offered between six different species, calling the offering of both “Free-choice experiments” (p.819). Another can be found in Stevens, Rosati, Ross, and Hauser (2005), who presented cotton-top tamarins and common marmosets with a single food reward option in a “forced-choice session” (p.1859) and let them choose one of two food reward options in “free-choice session[s]” (p.1859). In education, for example instances in which learning material is sought out outside of formal educational contexts can be called free choice learning, as can those in which materials within such contexts are freely chosen. One instance of this can be found in Kola-Olusanya (2005), which is about how and when children learn about environmental issues in settings in which learning opportunities are presented for the children to choose from, for example, museums, zoos etc. Another can be found in Wood (2014), which is about the choices children make while playing in early education and how they relate to power structures and national educational frameworks. While the differences to the type of free choice the rest of the present work will be about might seem obvious or even trivial, it is interesting to briefly consider where the contrasts lie exactly. In the ethological examples, the parallels to the present text’s free choice tasks are potentially fairly large, with the critical exception that the choices presented are qualitatively different. As such, they are used to investigate preferences. This potentially operationalizes some ideas about self-generated action (which will be elaborated more in Chapter 2) better than the free choice tasks in which the choices are, essentially, interchangeable, as the qualitatively different response options would allow an expression of self and personal preferences. The educational example, especially those in which educational opportunities are sought out deliberately and on the subjects’ own initiative, also potentially ‘operationalize’ (the term applies, but this is not an experimental context) similar aspects of self-generated action better than the free choice tasks talked about in this work.

# 2 For what are free choice tasks used?

The purpose of this chapter is to provide an overview over the different uses free choice tasks have seen, mostly in conjunction with forced choice tasks in some way, including brief introductions into the relevant theoretical concepts as well as arguments for and against these uses. Three broader categories of purpose will be presented alongside examples for each and, if applicable, criticisms of these uses. The first use is probably the most common one: Operationalizing self-generated and externally triggered actions. The other two uses that will be presented here are free choice tasks as conflict tasks and free choice tasks as tasks without inherently clear stimulus-response links.

## 2.1 Self-generated and externally triggered actions.

In the research literature on action there is a contrast that can often be found in one shape or the other. The specific words used to describe the two sides of the contrast are often different. On the one side there are actions that will be referred to in this text as self-generated (e.g. Passingham, Bengtsson, & Lau, 2010). They are also called endogenous (e.g., Pfister, Heinemann, Kiesel, Thomaschke, & Janczyk, 2012), intentional (e.g., Brass & Haggard, 2008), internally generated (e.g., Obhi & Haggard, 2004), intention-based (e.g., Herwig et al., 2007 or Keller et al., 2006) or voluntary. On the other side of this contrast, there are those actions that here will be called externally triggered actions (e.g., Jenkins, Jahanshahi, Jueptner, Passingham, & Brooks, 2000, or Obhi & Haggard, 2004), also referred to as stimulus-based (e.g., Herwig et al., 2007, or Keller et al., 2006) or exogenous actions (e.g., Pfister et al., 2012).

These two types of actions are sometimes assumed to have two distinct types of motor control. This idea links up with arguments found in Passingham (1983). Here, Passingham argues that the arcuate premotor area is responsible for actions based on cues from outside the organism (the model organisms here being Macaca fascicularis and Macaca mulatta, two species of Old World monkeys) and that the supplementary motor area is responsible for actions based on proprioceptive cues, that is, cues from inside the organism. Similarly, some scholars theorized that there are two types of action control (or “two action control modes”), which are also claimed to be controlled by two (neuronally) distinct action control systems (Astor-Jack & Haggard, 2005; Goodale, Westwood, & Milner, 2004; Krieghoff, Waszak, Prinz, & Brass, 2011; Obhi & Haggard, 2004).

The core functional difference between the two systems is, by definition, that one is involved in enacting intentional actions with the goal of producing an effect in the environment and that the other is involved in actions that are prompted by stimuli that are encountered in the environment. One functional difference between these two systems, according to Herwig et al. (2007), is then that the stimulus-based action system enables stimulus-response (or sensorimotor) learning (cf. Hommel’s, 2000, concept of a prepared reflex, in which a perceptual antecedent is automatically and non-intentionally translated into a motor response). Meanwhile, the intention-based action system enables action-effect (or ideomotor) learning (cf. Elsner & Hommel’s, 2001, model of action control, in which, over time and automatically, bidirectional links between actions and their perceivable consequences are created).

The arguments for this distinction come from both neuropsychological and behavioral observations. Astor-Jack and Haggard (2005) used what they called the ‘truncation paradigm’. They assumed that for intentional tasks, there is a preparatory gradual accumulation of activation of the motor system while there is no such process for reactive tasks. In the truncated condition, this preparatory build-up of activation is interrupted by the presentation of a stimulus that the participants are instructed to react to with the same response as the one they were already preparing. Note that for the intentional action tasks here, participants were asked to give a response with their right finger within 2-10 s of a trial start signal. Truncation was implemented by presenting a tone during such an intentional action trial, which indicated that the response was to be given immediately. They then compared the RTs in the truncation condition with those in the reactive action condition, observing that RTs in the truncation condition were longer than in the reactive action condition. Additional analysis of pupil dilation data confirmed that participants were indeed preparing an intentional action up until the truncation happened. These longer RTs, Astor-Jack and Haggard suggest, are due to a necessary switch between the two motor systems, which they claim are not normally active at the same time.

Similarly, Obhi and Haggard (2004), also used a truncation paradigm. Here, participants were to flex a finger in response to stimuli, again either at a time of their choosing or immediately. For truncation trials, the former type of trial was interrupted by a tap to the neck of the participant by the unseen experimenter. They observed characteristic differences in the strength with which responses were given (in the form of EMG data) between the intentional and reactive actions. In truncation trials, the EMG signature of the strength of the response was the same as that observed for reactive trials, which Obhi and Haggard interpreted as evidence that it was indeed a response of the stimulus-driven type that was given. As Astor-Jack and Haggard (2004), Obhi and Haggard also reported longer RTs in the truncation condition than in the stimulus-driven condition, which they interpreted to be evidence for two separate motor systems for the two types of actions.

With regards to evidence from EEG and lesion studies (in addition to behavioral evidence and imaging studies), Krieghoff et al. (2011) reviewed the literature and, while there is still ambiguity in the literature, identified a medial-lateral dimension on which intentional action control varies. There, the fronto-medial cortex is identified with intentional action control while the fronto-lateral cortex is involved in stimulus-oriented action control. Within the former, they further specified that the more anterior regions of the fronto-medial cortex are responsible for more abstract aspects of behavior (which here map to the What and Whether aspects of Brass & Haggard's, 2008, "What When Whether" model of intentional action) while the more posterior regions are responsible for specifying other crucial aspects of the intended behavior (which for example would include the When component of Brass & Haggard's model).

Note that it is somewhat unclear in the literature whether the amount to which actions are self-generated or externally triggered is meant in a dichotomous way or in a continuous way spanning the two extremes. The former is at least suggested by both the language of entirely different substrates for each and for example Astor-Jack and Haggard's claim that normally the two systems are not active at the same time. The latter on the other hand is expressed explicitly in, for example, Krieghoff et al. (2011): "Human actions are rarely totally externally determined, nor are they ever completely internally guided. Rather, they almost always comprise external and internal components. Therefore, it might be more reasonable to assume that human actions exist along a continuum between the two extremes" (p.768). Similarly, Passingham et al. (2010) emphasize the interconnectedness of the neural systems (here: the supplementary motor area and the lateral premotor cortex) responsible for the two types of actions and write that this "would be expected given that the distinction between self-generated and externally guided actions is rarely absolute, but often one of degree" (p. 20). Janczyk (2016) largely follows these views, describing the differences between self-generated and externally triggered actions as not qualitative in nature. An exception mentioned there is that of unconditioned reflexes, which appear to belong to a category entirely distinct from actions altogether (Janczyk, Pfister, Wallmeier, & Kunde, 2014). The specific commonality between self-generated and externally triggered actions that is absent in unconditioned reflexes is whether action effects play a role in the behaviors. Action effects will be further discussed in Chapter 3.

As mentioned in the previous chapter, this use of free choice tasks to operationalize self-generated actions is not without its criticisms. For example, Krieghoff et al. (2011) themselves say that „In contrast to most situations in everyday life in which an intentional decision is associated with a certain (personal) value, there is no such value in an experimental setting" (p.774). Schüür and Haggard (2011) express a similar criticism. They surveyed the literature and categorize operationalizations of self-generated actions into (1) operant actions, (2) underdetermined actions and (3) motor consequences of integration of different types of inputs. Within those categories, free choice tasks as understood here fall, according to Schüür and Haggard, into the second category, which is defined as "actions in the absence of cues" (p.1699). Schüür and Haggard view this conceptualization of self-generated actions as problematic because, according to them, it recourses onto an agentic/reflective self, for which there, they say, is no empirical evidence. They view this conceptualization as, ultimately, sourced in introspective experience. While this is a more conceptual criticism, there also are some empirical challenges to some of the conclusions mentioned above. For example, Astor-Jack and Haggard (2005) claimed that the two motor control systems generally are not simultaneously active and therefore switching between them should incur switch costs. This was not observed when Janczyk, Nolden, and Jolicoeur (2015) reported on three experiment in which free and forced choice tasks were presented in homogenous single-task blocks, mixed single-task blocks and dual-task blocks, in which stimuli for both tasks were presented at the same time. The RTs increased from homogenous to mixed single-task to dual-task blocks and free choice tasks took longer compared to forced choice tasks. Most importantly, they did not observe any difference in dual-tasking costs between free and forced choice tasks, which Astor-Jack and Haggard's claim would have predicted here.

In a set of experiments using the psychological refractory period paradigm (Pashler, 1994) and the additive factor logic (Sternberg, 1969), Janczyk, Dambacher et al. (2015) located the source of the mean RT difference where free choice tasks take longer than forced choice tasks within a phase of pre-central processing, specifically in perceptual processes. In the additive factor experiment, the experimental manipulation targeted stimulus brightness, effectively replicating the second experiment from Berlyne (1957). However, the results were different: Stimulus brightness affected forced choice tasks, but not free choice tasks. With only two datasets directly investigating the different influence of stimulus brightness on free and forced choice tasks, however, it is hard to draw any firm conclusions. One potential explanation for the different results in these studies could be different differences in luminance between bright and dark stimuli. These cannot be directly compared, as the Janczyk, Dambacher et al. study does not report the luminances of the stimuli used. If the results of the Janczyk, Dambacher et al. study were to be replicated, this would speak for free choice tasks being less dependent on external information that forced choice stimuli.

To summarize, while there is evidence in support of the general distinction between self-generated and externally triggered actions, these categories may not be as strictly dichotomized as they sometimes are portrayed, as there is overlap between the two tasks on both a neurological and a behavioral level. There will be a further examination on the similarities and differences between the two task types in Chapter 3.

## 2.2 Conflict tasks and tasks with ambiguous stimulus-response links.

The second and third use of free choice tasks, which will be showcased in this section, are (1) free choice tasks as conflict tasks and (2) free choice tasks as tasks without clear stimulus-response links.

As already described in the previous chapter, Berlyne (1957) conceptualized free choice tasks as conflict tasks and used them to investigate the effects of different aspects of conflict ((1) the relative strength of response tendencies, (2) number of choice options, (3) absolute strength of response tendencies, and (4) the degree to which responses are incompatible to each other) on the speed of responses. His free choice stimuli were two different forced choice stimuli (lights in different colors), simultaneously presented, which also operationalized the (1) relative strength of response tendencies, with a more equal strength, that is, the free choice stimuli, resulting consistently in slower responses than with unequal response tendencies, that is, the forced choice stimuli. He operationalized the other aspects of conflict by (2) varying the amount of choice options, two or four options (more response options slowed down responses for both task types), by (3) varying the amount of forced choice stimuli presented simultaneously (doubling the presented stimuli sped up forced choice responses but slowed down free choice responses) or (3) their luminosity (the brighter the stimuli, the faster the responses, with a less pronounced effect on free choice responses), and by (4) whether responses were given by manipulating two separate switches, each only going in one direction or by operating one single switch capable of being toggled in two directions (more mutual exclusivity only slowed down forced choice responses but did not affect free choice responses). Given these results, Berlyne proposed excluding physiological incompatibility as a potential source of conflict. Further, Berlyne discussed the relationship between conflict and RTs by examining both potential explanations for the longer RTs in free choice tasks and the prerequisites of making any choice at all in them. The four explanations he discussed are that (1) "some relatively improbable combination of events must come about" (p. 114) before a response can be given, (2) approach-approach conflict is resolved by random behavioral variation that tilts the participant towards one of the choices, similar to (1), (3) information for both responses is constantly collected but for a choice to be made, some information has to be discarded and some created, which takes time, and (4) cortical organization is disrupted by conflict.

This was not the only instance in which free choice tasks were used to investigate conflict. Botvinick, Braver, Barch, Carter, and Cohen (2001) reviewed several studies which reported heightened anterior cingulate cortex activation when participants were performing free choice tasks, compared to when they were performing forced choice tasks (Deiber, Passingham, Colebatch, Friston, Nixon, & Frackowiak, 1991; Frith, Friston, Liddle, & Frackowiak, 1991; Jueptner, Frith, Brooks, Frackowiak, & Passingham, 1997; Playford, Jenkins, Passingham, Nutt, Frackowiak, & Brooks, 1992; for examples of more recent studies showing involvement of the cingulate cortex in free choice or free choice-like tasks, see Lau, Rogers, Ramnani, & Passingham, 2004; Soon, Brass, Heinze, & Haynes, 2008; Zapparoli et al., 2018). Note that in these studies participants were not instructed to respond freely or according to their own will but to respond randomly. Botvinick et al. argued that these "underdetermined responding tasks" (p. 628) result in the simultaneous activation of incompatible response tendencies (e.g., of only pressing a left button and only pressing a right button) which constitutes conflict. In this, they argue similarly to Berlyne (1957), specifically the aspect (4) of the incompatibility of response options. Another argument they used why this is evidence that conflicting responses being activated simultaneously are responsible for this increase in activity in the anterior cingulate cortex are the results reported in Raichle et al. (1994), who used a task in which participants were to respond to a noun with an appropriate verb. They reported an increased activation of the anterior cingulate cortex (compared to a task in which the nouns were to be simply repeated back) in the beginning, but this increase vanished quickly with repeated presentations of the same list of nouns. It reappeared when a new list of nouns was presented. This, Botvinick et al. argue, was, similarly to the free choice tasks, due to a newly presented noun activating multiple mutually exclusive responses at the same time, also leading to conflict.

Lastly, free choice tasks have also been used as tasks in which no clear stimulus-response links exist. Elsner and Hommel (2001), for example, used free choice tasks in their investigation of their 2-phase model of action control. In this model, in a first phase associations between motor actions and their effects are learned and in a second phase these associations are used in intentional actions to achieve these learned effects in the environment. In all four experimental designs presented in their study, free choice tasks were used in the acquisition phase in the form of auditory tones following freely chosen left or right button presses. In the test phases of one of the experiments (Experiment 1), forced choice tasks and in the other three (Experiments 2, 3, and 4), free choice tasks were used. Stimuli in the test phase were the tones which in the acquisition phase were the effects of the left and right button presses. In the experiment with forced choice tasks in the test phase, participants responded faster when the button press that a tone instructed was the same as the button press that produced the same tone in the acquisition phase than when they were different. In the three experiments in which the test phase was comprised of free choice tasks (i.e. either response was considered correct to either tone), the choices made in response to each tone were robustly biased towards the button press that was associated with the respective tone in the acquisition phase. These results were replicated by Vogel, Scherbaum, and Janczyk (2018), who used a mouse tracking paradigm to be able to examine patterns over the course of free choice trials. They identified two groups of participants who use different and distinct strategies for choosing a response option: Early choosers and late choosers. Those who chose early in (or even before) a trial which response option they were going to pick were not affected in their choice of response option by the presented previous action effect, whereas those who picked their response option during the trial were.

Overall, Elsner and Hommel’s (2001) results were taken as evidence that when a motor action is repeatedly followed by a perceived effect in the environment, an automatic association between the motor action and the effect is created, regardless of actual relevance of said association to the task at hand. This interpretation necessarily presupposes that, without the intervention of the acquisition phase, there would be no connection between the free choice stimuli and the responses given to them. As such, Elsner and Hommel used free choice tasks here as a type of task which would normally be ‘neutral’ unless a manipulation creates these links between stimuli and responses.

Another example of such a use is found in Naefgen, Caissie, and Janczyk (2017). The target of this study was to investigate the role of stimulus-response links in the genesis of the compatibility-based backward crosstalk effect (BCE) (e.g., Hommel, 1998, Hommel & Eglau, 2002, Lien & Proctor, 2002). The BCE is an effect that occurs in dual-tasking situations. Models which posit a response selection bottleneck (for overviews, see Lien & Proctor, 2002; Pashler, 1994; or Pashler & Johnston, 1998) claim that the central processing of the first and the second task in such a situation happens strictly serially. This in turn means that the second task's response selection should not influence the first task's response selection whatsoever. Nonetheless, whether or not two tasks in a dual-tasking situation are compatible on some dimension (e.g., whether the sides on which the response to the two responses are to be given are the same or different) influences not only the speed of the performance in the second task, as general assumptions of priming would predict, but also that in the first task (with Task 1 responses that are compatible to the Task 2 response as described above being faster than incompatible ones), hence the "backward" component of the BCE's name. Hommel (1998) attributed this phenomenon to an automatic stimulus-response translation that is capable of running in parallel to other processes, which would enable crosstalk between the two processes (for an elaboration, see Janczyk, Renas, & Durst, 2018). To further test the role of such stimulus-response links in the Naefgen, Caissie et al. study, three experiments were run in which Task 1 or Task 2, in a dual-tasking situation, was either a free or a forced choice task (which task was a free or a forced choice task was varied between but not within experiments) while the respective other task was always a forced choice task. The logic here was based on the assumption that free choice tasks have either no or at least weaker stimulus-response links. If these stimulus-response links are necessary in both tasks for the BCE to occur, an absence (or weaker strength) of them should lead to an absence of the (or weaker) BCE. Stimulus-response links in forced choice tasks are usually first instructed and then reinforced with repetition and feedback over the course of an experiment. This also occurs to some degree with free choice tasks, as, here, usually two responses are repeatedly given after one stimulus (albeit only one after each stimulus presentation), therefore plausibly also forming an association between the stimulus and the responses. However, these assumed associations would also only have roughly half as many chances to be reinforced as those in a forced choice task. Furthermore, there is evidence that such links can be formed rapidly (e.g., Wolfensteller & Ruge, 2011, reported evidence for response-effect learning after only 12 trials).

The experimental setups in Naefgen, Caissie et al. (2017) allowed for checking the BCE impact on Response 1 in all three experiments and for checking its impact on the choice made in Task 1 when it was a free choice task. When both tasks were forced choice tasks, there was always a BCE. When Task 1 was a free choice task, the BCE was smaller than when it was a forced choice task and it was absent when Task 2 was a free choice task. Furthermore, in all experiments there was a bias to choose the same response in the free choice task as the forced choice task instructs. This was descriptively larger in the experiments in which Task 1 was the free choice task than in the one where it was Task 2. These results overall suggest that the BCE relies at least to some extent on stimulus-response links. However, it is not possible to tell from this study whether (a) free choice tasks have no stimulus-response links but those links are not necessary (but sufficient) for the BCE, whether (b) stimulus-response links in free choice tasks are just weaker than for forced choice task and necessary for the BCE (thus limiting its strength) or (c) a combination thereof is true.

Self-generated actions, conflict tasks, and tasks with ambiguous stimulus-response links: These are the three major ways in which free choice tasks are used. The observations reported above already suggest some specific ways in which free and forced choice tasks are similar and dissimilar. The next chapter will provide a systematic overview over these specific similarities and dissimilarities.

# 3 How similar are free and forced choice tasks?

The goal of this chapter is to first review and then evaluate to what extent and in which specific ways free choice tasks and forced choice tasks are similar or different. Contrasting the tasks in that manner will by necessity be somewhat redundant to the previous chapters, as the differences and similarities become visible through the application of the tasks. One of the purposes of this chapter is to evaluate the claim that free and forced choice tasks are qualitatively different. This to-be-evaluated claim can, for example, be found in the form of an explication of Herwig et al.'s (2007) work in Pfister, Kiesel, and Melcher (2010). Herwig et al. replicated Elsner and Hommel's (2001) experiments in which the stimuli in a test phase were, in a previous acquisition phase, the effects of specific free choice task choices. While Elsner and Hommel observed in a test phase in which these previous action effects were free choice stimuli that responses congruent with the acquisition phase association were more likely and faster than incongruent responses, Herwig et al. (p. 1540 and 1549) reported that this effect "holds for endogenously driven actions only!" That is, it does not occur when the learning phase uses forced choice tasks. Pfister et al. (2010) interpreted the different action control modes posited by Herwig et al. as "fundamentally different systems" (p. 317). This interpretation (and its basis) appears especially plausible if one follows the arguments for different neural substrates governing these different kinds of actions cited by Herwig et al. (see also Chapter 2).

Lastly, in order to investigate the differences between two task types, as is the purpose of the present work, it is necessary to first establish how closely the two tasks are related. How close they are related informs which questions are reasonable when comparing them. The focus of this chapter will be on behavioral evidence, as the evidence to the differences in the neurological substrates was already discussed in Chapter 2.

## 3.1 How are free and forced choice tasks dissimilar?

First, arguments and evidence speaking for a fundamental or qualitative difference between the task types will be collated in this section.

The operational differences already described in Chapter 1 that define the two tasks in contrast to each other are perhaps the most obvious differences: In forced choice tasks, only one answer is correct in response to the stimulus while in free choice tasks, more than one answer is considered correct.

The next very consistently observed difference that is not part of their definitions, starting with Berlyne’s (1957) paper, is the RT advantage that forced choice tasks have over free choice tasks (e.g., Bodner & Mulji, 2010; Janczyk, Dambacher et al., 2015; Janczyk, Nolden et al., 2015; Naefgen, Caissie et al., 2017; Schlaghecken & Eimer, 2004). There are some exceptions to this RT advantage. It can either vanish like in the dark stimuli condition of Experiment 3 in Janczyk, Dambacher et al. (2015) or even reverse like in Naefgen, Caissie et al. (2017) in the incompatible dual-tasking trials or in Experiments 1 through 4 in Wirth, Janczyk, and Kunde (2017). Strictly speaking, the observation in Janczyk, Dambacher et al., as an absence of differences, could be seen a result that belongs in the next section, but generally speaking these exceptions are rare and thus far have not been investigated systematically. One commonality might be that this reversal the RT advantage occurs mainly under dual-tasking conditions (but not always, see Janczyk, Nolden et al., 2015).

As described in the previous chapter, Naefgen, Caissie et al. (2017) combined free and forced choice tasks in a dual-tasking situation and observed that the BCE was reduced when Task 1 was a free choice task instead of a forced choice task and absent when Task 2 was a free choice task instead of a forced choice task. This suggests that interference of this kind between the two task types is at the very least reduced. There is the alternative explanation for this reduction of the BCE that this reduction was due to immediate conflict adaptation which will investigated further in Chapter 5. However, the fact that the BCE persisted at all albeit in a reduced form when Task 1 was a free choice task also suggests similarities between the two tasks.

Another type of difference that some researchers posit exists between the two task types is that free choice tasks allow for action-effect bindings to be formed while forced choice tasks do not or at least not to the same degree.

Starting with the assumption that free and forced choice tasks are implemented by different neuronal substrates, Herwig et al. (2007) proposed that the two action control modes represented by the two task types are also associated with different modes of learning: Forced choice tasks (representing an externally-triggered action control mode) would be associated with stimulus-response/sensorimotor learning while free choice tasks (representing a self-generated action control mode) would be associated with action-effect/ideomotor learning. In three experiments similar to Elsner and Hommel's (2001) experiments (see Section 2.2 for a summary), Herwig et al. changed (among other variables) whether the action control mode in the acquisition phase was self-generated or externally-triggered by using free and forced choice tasks, respectively. Indeed, their results were compatible with their prediction that action-effect learning would only occur for the free choice task: Compatibility effects in the test phases with faster responses when the stimulus in the test phase was the same as the effect in the test phase only occurred when the acquisition phase used free choice tasks. These results were not entirely unambiguous, though. In the error rates of Experiment 2, there was a significant compatibility effect on the error rates, which was (numerically) *larger* when the acquisition phase consisted of forced choice trials. Herwig and Waszak (2009) tested an alternative explanation for these results: That the source of this difference is that for free choice task the attention is split between the response to be given and the effect that follows it (two elements) and for forced choice tasks the attention is split between these two elements as well as the stimulus (for a total of three elements). Thus, the lack of attentional resources would be the alternative explanation for action-effect associations not forming. In three experiments, they manipulated (1) the number of elements for free choice tasks to match that of forced choice tasks by introducing a stimulus discrimination task (action-effect learning happened), (2) the amount of attention drawn to the effects of forced choice tasks by adding 'wrong' action effects that had to be detected by the participants (action-effect learning did not happen) and (3) doing the same as in (2), but with free choice tasks (action-effect learning happened). Their results were consistent with their original hypothesis of only free choice tasks leading to action-effect learning. Later, Herwig and Waszak (2012) also investigated the strength and durability of action-effect associations that result from free and forced choice tasks and observed that while for forced choice tasks there are short-term action-effect associations but there is no long-term action-effect learning while for free choice tasks both short-term associations and long-term learning happen between actions and their effects (see also Janczyk, Heinemann, & Pfister, 2012). This view of the hypothesized two action control modes is, however, not undisputed. As any evidence to the contrary will necessarily lead to a view of the task types as more similar. This evidence will be presented in the next section.

## 3.2 How are free and forced choice tasks similar?

In this section arguments and evidence for similarities between free and forced choice tasks will be reviewed.

Again the most obvious similarities can be found in the descriptions of the task. In most regards, the two tasks can outwardly appear the same: The stimuli are interchangeable and (mostly) just need to be instructed differently to be one or the other, the motor execution can be virtually indistinguishable (pressing a button, flipping a switch; but see Obhi & Haggard, 2004, for an example of distinguished motor responses) and so on. But of course there are other, less superficial similarities as well.

Both task types can be influenced by masked primes (Bermeitinger & Hackländer, 2018; Bodner & Mulji, 2010; Kiesel et al., 2006; Le Bars, Hsu, & Waszak, 2012; Schlaghecken et al., 2004), with the primes both biasing the choice of the free choice tasks and influencing the RTs: free choice responses congruent to the prime are faster than responses that are incongruent to the prime. This influence by primes seems to be somewhat qualitatively different between the task types in its specifics: Mattler and Palmer (2012) reported that free choice stimuli integrate information from both internal and external sources (here: primes) while the effect primes have on forced choice stimuli is dependent on the interaction between automatic processing of the prime and the target stimulus.

Janczyk, Nolden et al. (2015) performed experiments in which either a free and a forced choice task or two forced choice tasks were presented simultaneously. If there are two distinct action control systems for the two tasks, this would imply that changing from using the one system to the other would require additional time. They reported that this was not the case and that both tasks are similarly affected by dual-task costs in such a dual tasking situation. This shows that any switch in action control systems does not show up in the form of different dual-task costs between the central processing of the two stimuli. Within the same central bottleneck framework, Janczyk, Dambacher et al. (2015) further reported that the RT difference between the two tasks arises in the pre-central stage, specifically that there is enhanced perceptual processing for forced choice tasks in comparison to free choice tasks. While this shows that the perceptual processing of these tasks might be different, it also shows that the central processing of the two tasks may be qualitatively the same.

Similarly, the fact that the BCE occurs at all (albeit reduced) with a free choice Task 1 (Naefgen, Caissie et al., 2017) is evidence for some kind of qualitative overlap in how the two tasks are processed. Bermeitinger and Hackländer (2018) applied motion primes to free and forced choice tasks and observed that, while the length of the stimulus onset asynchrony (SOA) influenced whether compatibility effects were positive (short SOA) or negative (long SOA), there was no difference between the tasks, again suggesting that interference works similarly with the two task types.

In a similar vein to Herwig and Waszak's observation (2012, see Section 3.1 for a summary) that there is no difference in short-term action-effect associations between the two task types, Janczyk, Heinemann et al. (2012), in two experiments, demonstrated that short-term action-effect bindings occur for both free and forced choice tasks. They investigated this by adapting a paradigm established by Dutzi and Hommel (2009) in which after a response is given to a first task, one of two tones was randomly selected and presented. Following this tone, a second tone was presented, which was either the same or the other tone. This second tone served as stimulus for the second task. Task 1 was either a free or a forced choice task (whereas Dutzi and Hommel only ever had Task 1 be free choice tasks) and Task 2 was always a free choice task. They observed that when the two presented tones were the same, there were also more repetitions of the Task 1 response in the Task 2 response than if the two presented tones were different. Critically, as this happened in both free and forced choice tasks (and numerically even stronger for forced choice tasks than free choice tasks), they interpreted this as evidence in favor of short-term action-effect bindings forming rapidly for both free and forced choice tasks. This contradicts the most extreme version of the claim that action effect learning only happens in an intentional action control mode (e.g., Herwig et al., 2007; see also Herwig & Waszak., 2009), but is in line with, for example, the observation of Herwig and Waszak (2012) that while short-term action-effect bindings are formed for both task types, only free choice tasks will form more persistent, longer-term action-effect associations. Nevertheless, there is ample evidence that long-term action-effect associations are also formed in free choice tasks. Pfister and Kunde (2013) for example investigated in two experiments the relative roles of anatomical features (here: which hand is used to respond) and spatial features (here: where the button that is to be pressed is located) of responses in response-effect compatibility phenomena. To this end they had participants give responses with their hands positioned either 'normally' or crossed over (the left hand pressing the right button and vice versa), which produced effects that were either on the left or the right. They presented both forced and free choice stimuli. In Experiment 1, the action effect compatibility effects were stronger in the forced choice task condition while in Experiment 2 they were the same across task types. While the results from Experiment 1 are, strictly speaking, evidence that speaks for a difference between the two task types, it is also evidence that contradicts a strong position claiming a difference going in the opposite direction, hence its inclusion here (for some other examples that run counter to the idea of only free choice tasks leading to longer-term action-effect learning see Gozli, Huffman, & Pratt, 2016, only forced choice tasks, no comparison with free choice tasks; Huffman, Gozli, Hommel, & Pratt, 2018, only forced choice; Janczyk, Durchst, & Ulrich, 2017; Janczyk, Pfister, Crognale, & Kunde, 2012; Janczyk, Pfister, Hommel, & Kunde, 2014; Kunde, 2001, Experiments 1 and 2 forced choice tasks, Experiment 3 free choice tasks, but no direct comparison; Kunde, Pfister, & Janczyk, 2012; Wolfensteller & Ruge, 2011). One potential limitation or moderating factor for some of these findings put forward by Pfister, Kiesel, and Melcher (2010) is that the mere presence of free choice tasks leads to a self-generated action control mode for both free and forced choice tasks. This would mean that when both tasks are presented together as opposed to in separate blocks, action-effect learning can happen for both.

Lastly, there are also recent conceptual criticisms of the distinction between self-generated and externally triggered action control expressed by Hommel and Wiers (2017), who argued that it would be better to view action control as a unitary system. In this view, virtually all human actions that have been investigated in psychological research are goal-directed, as this unitary action control system would be responsible for all actions which fulfill any criteria the agent in question has. Sometimes those criteria would lead to relatively faster and well-rehearsed behavior (here: forced choice tasks) or relatively slower, more novel behavior (here: free choice tasks).

### 3.3 Evaluation of similarities.

Overall, there are differences between the tasks, but also marked similarities. While cognitive interference seems to show some differences between the two task types, priming works similarly on both and the central processing of both tasks is at least similar enough to not incur switch costs. In sum, the evidence suggests that the two tasks are similar enough to be investigated through common frameworks in order to identify their differences further.

# 4 Research questions

Considering the similarities and differences described in Chapter 3 and the limits of what is currently known about those, the central question this dissertation concerns itself with is this:

**How are free and forced choice tasks similar and how are they different?**

Accordingly, this work is about specifying these similarities and differences between free and forced choice tasks further than has already been done in the literature and staking out the theoretical territory within which the two task types can be put into a common framework. This latter aspect also includes investigations into where exactly any remaining differences lie.

Because an exhaustive answer to this question is outside the scope of this work, in the following three more specific research questions that follow from the larger question are named and attempts to answer them will be described in Chapters 5 through 7.

Question 1: **Is the reduced BCE when a free choice task is involved due to conflict adaptation?** (Chapter 5) This question concerns the argument that free and forced choice tasks need to be somewhat similar because interference from one can affect performance in the other and concerns this alternative explanation of the reduction in the size of the BCE. If the reduction of the BCE is due to conflict adaptation, this would strengthen the argument for similarity from mutual interference, as this would mean that the conflict adaptation suppresses a usually even larger BCE. However, at the same time, it would speak for the specific difference that free choice tasks induce more conflict adaptation than forced choice tasks.

Question 2: **Are free choice tasks merely underdetermined tasks?** (Chapter 6) This question is about the already observed mean RT differences between free and forced choice tasks and is about identifying this difference within the established theoretical framework of sequential sampling models. As such, this research question is similar in purpose and general approach to works like Janczyk, Dambacher et al. (2015) and Janczyk, Nolden et al. (2015).

Question 3: **Are free choice tasks random generation tasks?** (Chapter 7) Question 3, in contrast to Question 2, is not concerned with placing the mean RT difference within an abstracted framework but aims to test a specific mechanical explanation of the mean RT difference.

The answers to these questions as well as their implications will be discussed in Chapter 8.

# 5 Smaller backward crosstalk effect for free choice tasks are not the result of immediate conflict adaptation

The (compatibility-based) BCE is an interference effect that occurs in dual-task situations. Specifically, it refers to the observation that when the two tasks require responses that are on some dimension incompatible (e.g., a right response in Task 1 and a left response in Task 2) the performance in Task 1 is negatively affected (mostly in the form of longer RTs), compared to when the responses are compatible on the same dimension. This happens despite the fact that, at the time, Task 2 should not have been processed yet and thus not be able to influence Task 1 performance according to central bottleneck models (e.g., Pashler, 1984, 1994).

Naefgen, Caissie et al. (2017) observed that the BCE is reduced when Task 1 is a free choice task. These results have a plausible explanation alternative to the reduced stimulus-response links assumed in this paper. It is possible that this reduction was due to conflict adaptation in reaction to the presence of the free choice task (see Section 2.2), which would result in reduced interference between tasks. As the BCE is an interference-based phenomenon, it is affected by such conflict adaptation. The general principle that interference is reduced under or after conditions of conflict has been known for a while. One example is the Gratton effect, which describes the sequential modulation of congruency effects, such as the Eriksen flanker task (Eriksen & Eriksen 1974; sequential modulation thereof: Gratton, Coles, & Donchin 1992) or the Simon effect (Simon & Rudell, 1967; sequential modulation thereof: Akçay, & Hazeltine 2007; Dignath, Janczyk, & Eder 2017). This effect can come about when there is an overlap between the task and irrelevant additional information, for example when there are task-irrelevant flanking stimuli which are either similar or dissimilar to the task-relevant stimulus in the flanker task or whether the stimulus appears in a similar place to where the response is supposed to be given in the Simon task. Like in the BCE, performance is better when the stimulus and the irrelevant information are compatible than when they are incompatible. These effects are reduced in trials following incompatible trials compared to compatible trials. This means that, following incompatible trials, the performance in incompatible trials is closer to the performance in compatible trials. Some cognitive mechanisms that are assumed to be involved in this are the level of focus on task-relevant features (Botvinick et al. 2001) and the suppression of irrelevant information

(Janczyk & Leuthold, 2018; Stürmer & Leuthold, 2003; Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002).

Janczyk (2016; see also Renas, Durst, & Janczyk, 2018, and Scherbaum, Gottschalk, Dshemuchadse, & Fischer, 2015) showed that sequential modulation generalizes to the BCE: In three experiments, the BCE was smaller in trials following incompatible trials compared to trials following a compatible trial. This shows that the BCE is, in principle, vulnerable to suppression by way of conflict adaptation. Critically, it has been shown that conflict adaptation can not only happen between two trials but also *within* one trial. Goschke and Dreisbach (2008) had participants perform a spatial compatibility task (arrows appearing in one of four locations, pointing in one of four directions) and (rarely) presented variations of these stimuli that indicated that an additional response is required of them. In those trials in which the location of the arrow was incompatible with the direction the arrow was pointing and, thus, conflict was present, more of these additional responses were not given than in non-conflict trials, which the authors interpreted as evidence for additional recruited cognitive control within these trials. Scherbaum, Fischer, Dshemuchadse, and Goschke (2011) gathered frequency tagged EEG data while participants performed a flanker task to trace the amount of attention allocated to differentiating between flanker and central stimulus. The results here also indicated that conflict adaptation happened over the course of a single trial. Together, this suggests that the adaptation to the hypothetical conflict in free choice tasks could feasibly affect the tasks of which the free choice task is one. If free choice tasks are conflict tasks, this makes it plausible that the results from the Naefgen, Caissie et al. (2017) paper could be explained by conflict adaptation.

However, these types of conflict adaptation do not generalize universally between types of interference. Sometimes conflict adaptation does not generalize between conflict tasks like in the case of Akçay and Hazeltine (2011), who observed conflict adaptation within the respective task type for both Simon and flanker tasks but did not observe it between them. Other times, there appears to be a global type of conflict adaptation or at least one that generalized across two types of task. One example for this is reported by Freitas, Bahar, Yang, and Banai (2007, Experiments 2 and 3), who interspersed flanker tasks and Stroop tasks and observed conflict adaptation between the two task types. For a review of conditions under which conflict adaptation was observed to generalize and under which it did not, see Braem, Abrahamse, Duthoo, and Notebaert (2014). As it is unclear whether conflict adaptation generalizes between free choice tasks and the BCE, it is necessary to empirically test whether or not the BCE is affected by conflict adaptation brought on by free choice tasks. This study approached this goal with two dual-task experiments in which the amount of free choice task-induced conflict was systematically varied and the effect this had on the BCE was observed.

In Experiment 1, the dual-task trials were, unbeknownst to the participants, presented in pairs: A prime trial always preceded a test trial. While all test trials were a combination of two forced choice trials (in half of which Task 1 was compatible and in half of which Task 1 was incompatible to Task 2), in half of all prime trials, Task 1 was a free choice task and in the other half a forced choice task (again, half of which were compatible and half incompatible). Task 2 was always forced choice. The trial course for both experiments is illustrated in Figure 2.

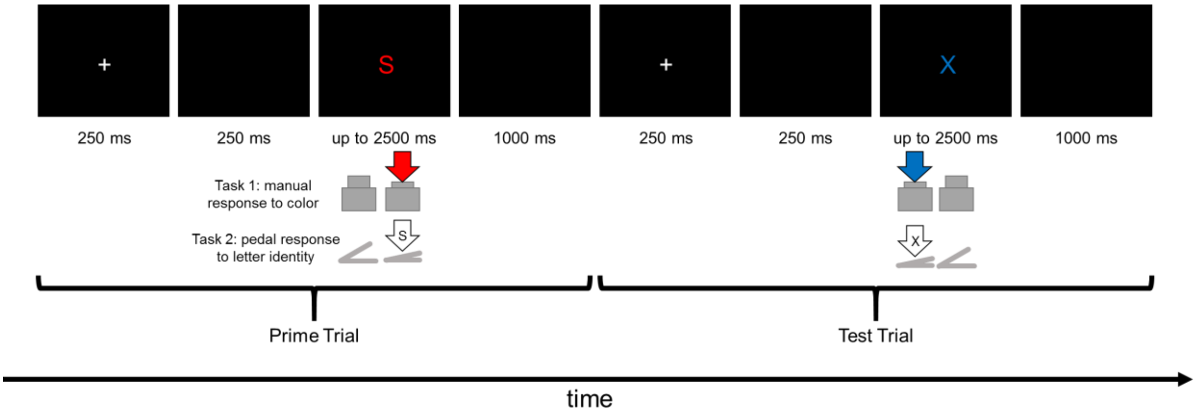

***Figure 2.*** Illustration of a pair of trials in Experiment 1. The red S is the prime trial stimulus and the blue X is the test trial stimulus. In this exemplary stimulus mapping, a red stimulus instructs the participant to give a right manual response (prime trial Task 1) and an S stimulus instructs the participants to give a right pedal response (prime trial Task 2). A blue stimulus on the other hand instructs a left response (test trial Task 1) and an X stimulus instructs a left pedal response (test trial Task 2). Therefore, this example illustrates a compatible forced choice test trial (as both responses are on the left side) following a compatible forced choice prime trial (as both responses are on the right side). Note that in Experiment 2 the general procedure was the same but there was no distinction between prime trials and test trials. (from Naefgen & Janczyk, 2018b)

We[1] predicted that the BCE should be smaller in test trials following incompatible forced choice prime trials than those following compatible forced choice prime trials, in a replication of the results from Janczyk (2016). Critically, we further predicted that, if free choice tasks lead to conflict adaptation, the BCE in test trials following compatible free choice prime trials should be smaller than in test trials following compatible forced choice trials. The restriction to compatible free choice trials (instead of including incompatible free choice trials) in the prediction was because the BCE following incompatible prime trials is

---

[1] Owing to the collaborative nature of this research, in the summaries of the studies in Chapters 5, 6 and 7, the first person pronoun was pluralized.

expected to be smaller regardless of the task type involved. Thus, if the BCE were to be smaller following such a trial, it would not be clearly attributable to either type of cognitive conflict. The results from Janczyk (2016) were replicated, there was a significant reversed BCE following incompatible forced choice prime trials. The critical results are visualized in Figure 3 and will be discussed together with the results of Experiment 2.

While in Experiment 1 a type of conflict adaptation was induced that is analogous to the sequential modulation of interference, in Experiment 2 a list-wide proportion congruency-based type was induced. Here, the amount of trials with a free choice Task 1 in a block was manipulated to be 25%, 50%, or 75%. There are multiple potential ways this could lead to heightened conflict in the critical trials with two forced choice tasks: There could be an accumulation of experienced conflict within the block or there could be a heightened expectation of conflict in any trial. If adaptation to free choice task-induced conflict were the cause of the reduced BCE in the Naefgen, Caissie et al. (2017) paper, a higher proportion of free choice tasks in Experiment 2 should also cause a reduced BCE. The results of this Experiment are visualized in Figure 2.

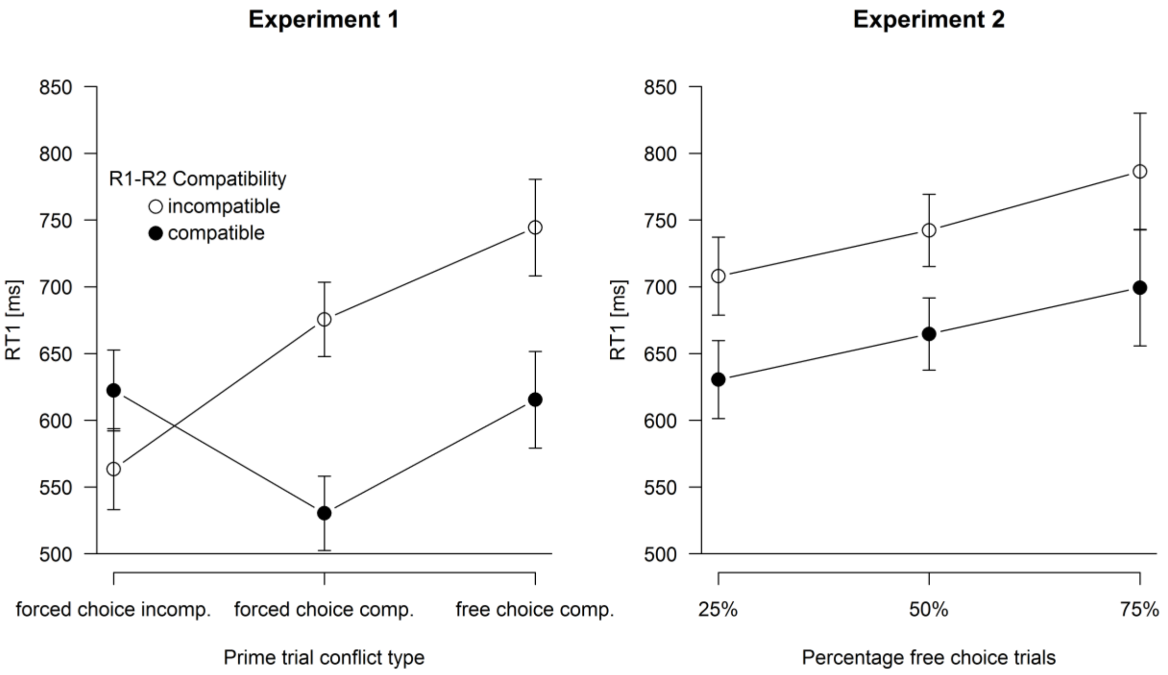

***Figure 3.*** Mean correct response times from Task 1 (RT1) of Experiments 1 and 2 as a function of prime trial conflict type (Experiment 1) or percentage of free choice trials in a block (Experiment 2) and Response 1-Response 2 (R1-R2) compatibility. Error bars are 95% within-subject confidence intervals calculated separately for each prime trial conflict type in Experiment 1 and separately for each percentage level of free choice trials in a block in Experiment 2 (see Pfister & Janczyk 2013). (from Naefgen & Janczyk, 2018b)

In summary, in both experiments there was no evidence for a reduced BCE caused by free choice tasks, as neither a preceding free choice trial nor a rising proportion of free choice trials in a block reduced the size of the BCE. This suggests that the results in Naefgen, Caissie et al. (2017) were not due to conflict adaptation in reaction to free choice tasks. Additionally, the observation that in a dual-task situation a free choice Task 1 is biased towards a spatially compatible response (i.e. a left response following a left response and a right response following a right response) was replicated in both experiments, which might be an additional dimension on which Task 2 influences Task 1, that is, in which the BCE is expressed.

# 6 Free choices compared to forced choices: Just underdetermined or is there an additional process? (Why free choices take longer than forced choices: evidence from response threshold manipulations)

Given that there is a consistently observed mean RT difference between free and forced choice tasks in which free choice tasks take longer (see Section 3.1), it is of theoretical interest to investigate the source of this difference. In order to investigate any specific differences between these tasks to which the mean RT difference could be attributed, it is necessary to first choose a common framework for the two tasks to make the differences tractable. In this study, we chose to investigate free and forced choice tasks through a sequential sampling lens, of which the drift-diffusion model (Ratcliff, 1978) is probably the most prominent example.

All of these approaches have in common that information is noisily accumulated over time until a threshold is reached and a response is initiated. While the resulting models vary in complexity, the model we assumed here is a minimal version reminiscent of Grice's (1968) variable criterion model in which there are three parameters: a) the decision thresholds, b) the non-accumulation time and c) the drift rate. The decision thresholds indicate how much evidence has to be accumulated for one or the other response in order to initiate its emission. Higher thresholds lead to longer RTs but also lower error rates, while lower thresholds lead to the opposite. This is because errors in this framework occur when the random noise shifts the accumulated information to the wrong response. As the noise is random, when there is more time for the constant accumulation in the correct direction, the amount of wrong answers will become smaller. The non-accumulation time represents how much time in the process is spent on processes that are not devoted to the accumulation of information, that is, a constant length of time that needs to be added to the whole duration of the process. The drift rate represents the speed with which information is accumulated and thus how fast a given threshold is reached. The total RT is a composite of the length of the non-accumulation time on the one hand and a combination of the drift rate and the decision thresholds on the other hand.

Within this model, there are at least two plausible explanations of how the mean RT differences between free and forced choice tasks come about. Those are that the two tasks differ in their

a) drift rates, with free choice tasks having lower drift rates.

b) non-accumulation times, with free choice tasks having longer non-accumulation times.

Explanation a) maps to the idea expressed by for example Schüür and Haggard (2011) that the defining feature of free choice tasks, as they are understood in the present work, is that they are underdetermined, that is, that there is simply a dearth of information for the participants. Meanwhile, explanation b) is less specific in to which explanation it maps. It would make theoretically necessary at least one additional process, but would not give any information on what that process is, with the exception of an upper limit of how long it can take. Alternatively, one of the already existing non-accumulation processes could be slowed down/elongated.

As both explanations predict longer RTs, it is necessary to identify what would make them distinguishable. Because the decision thresholds only interact with the drift rate but not with the non-accumulation time to determine the total RTs, this is theoretically fairly simple: If one of the drift rates is lower than the other, then increasing the decision thresholds should increase the difference between the two task types. Conversely, lowering the decision threshold should then reduce this difference. Should changing the decision thresholds not affect the size of the difference between free and forced choice tasks, this would suggest that the mean RT difference can be attributed to a difference in non-accumulation times. These two explanations as well as how the proposed manipulation would make them distinguishable are illustrated in Figure 4. If there is a difference in drift rates, there should be a significant statistical interaction between a manipulation of the decision thresholds and whether the task is a free or a forced choice task.

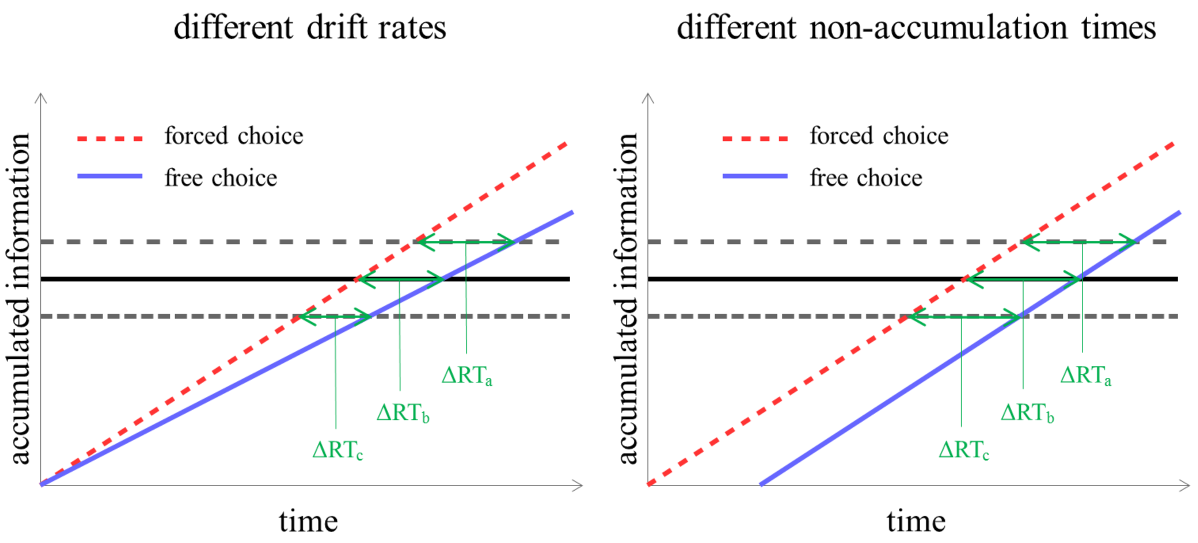

***Figure 4***. Two accounts of the mean RT difference between free and forced choice tasks. The continuous black line represents a medium decision threshold. The dashed line stands for a task in which there is an increased decision threshold and the dotted line for one in which it is lowered. Under the "different drift rates" account, the RT difference between forced and free choice tasks becomes smaller with lower thresholds ($\Delta RT_a > \Delta RT_b > \Delta RT_c$). In other words, task and threshold manipulation interact with each other. In contrast, with differences in non-accumulation times, the RT difference remains the same irrespective of the threshold ($\Delta RT_a = \Delta RT_b = \Delta RT_c$) and therefore reflects an additive relation between task and threshold manipulation. (from Naefgen, Dambacher et al., 2018)

For the manipulation of the decision thresholds, we used two approaches in three experiments. In all experiments, participants fulfilled relatively simple free and forced choice tasks in which they had to react to differently colored circles with, depending on the color, either a left or right button press or a button press of their choice. In Experiment 1, we used varying amounts of so-called catch trials in a given block while in Experiments 2 and 3 we used varying amounts of time pressure.

Catch trials are trials in which no stimulus appears and in which participants are instructed to not give a response. A higher proportion of these catch trials in a block increases the separation of the decision thresholds (e.g., Gordon, 1967; Näätänen, 1972). This is theorized to result from lowered expectations that the stimulus will appear, in turn leading to a heightened decision threshold (Brysbaert, 1994; Grice et al., 1982; Seibold, Bausenhart, Rolke, & Ulrich, 2011). For time pressure, it was demonstrated that higher time pressure (i.e. less time to react) leads to lower decision thresholds (e.g., Diederich, 1997; Dror, Basola, & Busemeyer, 1999; Forstmann, Dutilh, Brown, Neumann, Cramon, Ridderinkhof, & Wagenmakers, 2008).

For all three experiments, responses were manual button presses and participants were instructed to try to respond to the white free choice stimuli with both response options about equally often and to avoid patterns in their responses. All three types of trials (free choice, forced choice left, forced choice right) occurred equally often.

In Experiment 1, the amounts of catch trials per block were 0%, 25%, 50%, and 75%. Experiments 2 and 3 both had measurement blocks in the beginning, where the mean and the standard deviation of the participant were measured, so that in subsequent blocks individualized time limits within which a response had to be given could be applied. Both experiments had a condition of low, medium and high time pressure, which was applied blockwise. The baseline for each condition was the mean RT, which was then modified by adding or subtracting a multiple of the SD of the RT. For Experiment 2, those modifiers were +1 SD for the low, +0 SD for the medium and -1 SD for the high time pressure condition and for Experiment 3 they were +0 SD, +0.5 SD and +1.5 SD, respectively.

The results of the three experiments are illustrated in Figure 5. The slowing RTs in Experiment 1 with higher percentages of catch trials suggest that the decision thresholds were indeed influenced as intended. The critical results of Experiment 1 yield some ambiguity: While overall, there is a significant interaction between the amount of catch trials in a block and the task type, this is entirely driven by the condition in which there are no catch trials whatsoever. Here, the RT difference between free and forced choice task was reduced. Furthermore, while a higher decision threshold should result in slower responses with fewer errors, in Experiment 1 the conditions with the ostensibly higher decision thresholds had higher instead of the expected lower error rates. This could be indicative of the manipulation targeting the drift rates instead of the decision thresholds. This issue was not apparent, at least in terms of the RTs and PEs changing in the expected directions, in Experiments 2 and 3 (but for a discussion of the non-specificity of time pressure manipulations, including effects on drift rates, see e.g., Arnold, Bröder, & Bayen, 2015; Dambacher & Hübner, 2015; Rae, Heathcote, Donkin, Averell, & Brown, 2014; Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004).

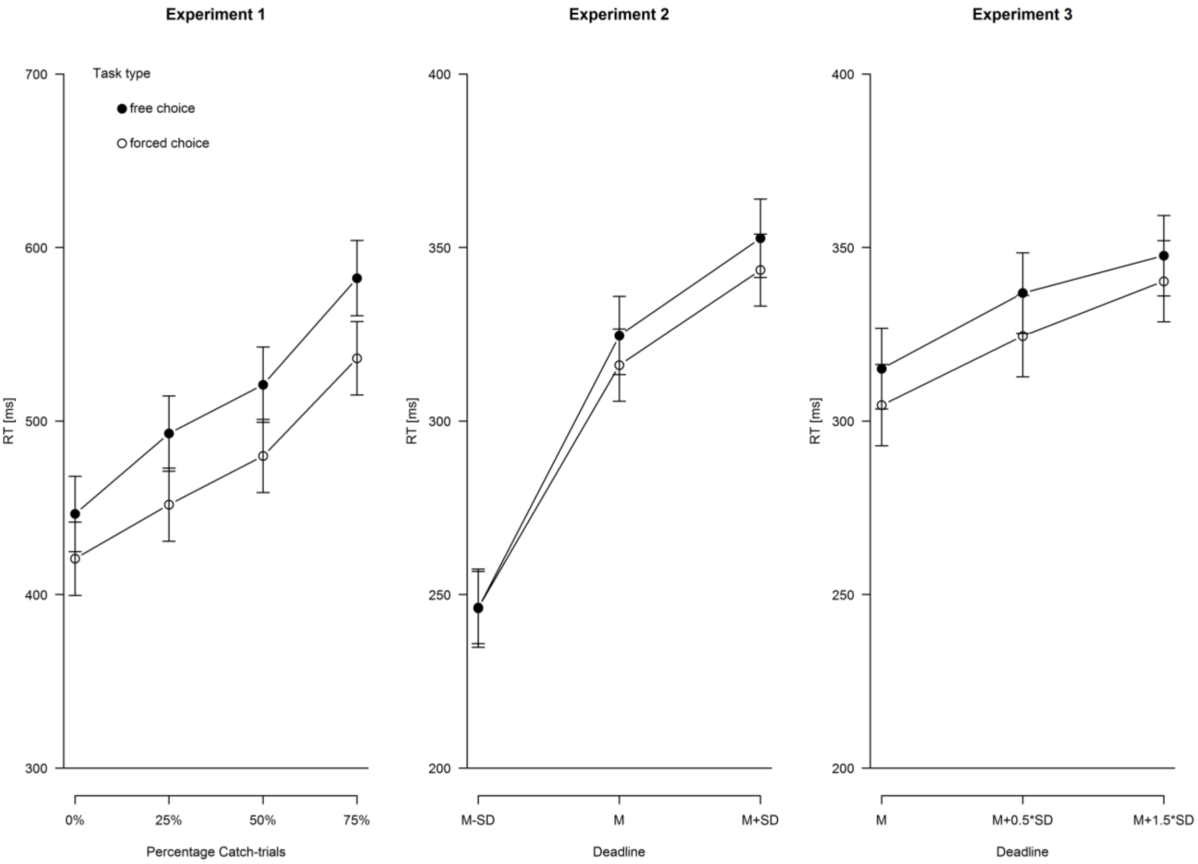

***Figure 5.*** Mean correct RTs from all three experiments as a function of task type and block type. Error bars are 95% within-subject confidence intervals calculated for the difference between free and forced choice tasks collapsed across block types (see Pfister & Janczyk, 2013). (adapted from Naefgen, Dambacher et al., 2018)

In Experiment 2, the results were again somewhat ambiguous: There was a significant interaction between the threshold manipulation and the task type on the RTs, but this result was driven by a collapse of the RT difference between free and forced choice tasks in the high time pressure condition. To check whether this was the result of a large proportion of fast guesses (as there were high error rates in this condition that, in part, reached almost chance level), we performed a median split categorization of the participants based on the error rates in the high time pressure condition. We then analyzed the two data sets separately, with a significant interaction between the threshold manipulation and task type only in the high-error rate subset of the sample. To gain more clearly interpretable data, a third experiment with less strict time limits was run.

In Experiment 3, the results were unambiguous: The threshold manipulation affected RTs and PEs as expected and there was no significant interaction between the threshold manipulation and the task type on the RTs.

In sum, as there was no stable interaction between the decision threshold manipulations and the size of the mean RT difference between free and forced choice tasks in Experiments 1-3, it is reasonable to attribute it to different non-accumulation times. As previously mentioned, while this rules out mere underdeterminedness of free choice tasks, it does not, by itself, illuminate *what* the additional processes involved in free choice tasks are. An attempt to answer that question is described in the next chapter.

One limitation of this study is that the assumed underlying parameters were not directly assessed for both types of tasks. Furthermore, when EZ (Wagenmakers, Van Der Maas, & Grasman, 2007) was used to extract the parameters of the forced choice tasks, in all three experiments the non-accumulation times were affected by the thresholds manipulation. In Experiments 1 and 2, the drift rates were affected by the threshold manipulation. In Experiment 2, this effect was largest between the high time pressure condition and the other two conditions, adding to the problems of interpretability here. Lastly, in Experiment 1, the response thresholds were not significantly affected at all, indicating potential problems with the interpretability of it.

# 7 Is the additional process a random generation task? (Free Choice Tasks as Random Generation Tasks: An Investigation through Working Memory Manipulations)

Accepting the former chapter’s conclusion that free choice tasks are not merely underdetermined but involve an additional process all but spells out the next question when trying to understand free choice tasks: What is this additional process?

This study aimed to answer this question. One specific potential mechanism was considered: whether the additional process is one of random generation. Frith (2013) noted that for free choice tasks as understood in the present work “in essence, the experimenter is asking her subjects to try to be unpredictable and random” (p. 291). The argument here is rooted in evidence that random choices are perceived as more free (Ebert & Wegner, 2011) and that, in neuroimagining studies, random generation tasks and free choice tasks activate similar brain regions (Jahanshahi, Dirnberger, Fuller, & Frith, 2000; Jenkins et al., 2000).

This becomes especially apparent when inspecting the instructions usually accompanying free choice tasks. Even when they don’t explicitly mention randomness as a goal (e.g., Hadland et al., 2001, p. 1105; see also Waszak et al., 2005; Elsner & Hommel 2001), they often still resemble instructions for randomness in that, for example, strategies or patterns in the responses ought to be (actively) avoided (e.g., Pfister & Kunde, 2013, p. 650; see also Linser et al., 2007). While this is often the case, there are exceptions to this. In those, often an emphasis is put on spontaneously choosing responses or acting freely (e.g., Pfister et al., 2010, p. 319; see also Herwig et al., 2007). For examples of these instruction types, see Table 1 in Section 1.1.

To make the assertion that free choice tasks are random generation tasks testable, we wanted to replicate an observation from a random generation context with free choice task. The logic here was that if free choice tasks are random generation tasks, they should react similarly to manipulations with known consequences in the latter task type’s case. There are several types of findings to choose from, such as time constraints (Baddeley, 1962, as cited in Baddeley, 1966) or concurrently performed tasks (Baddeley, 1966). We chose a specific version of the latter: the relationship between random generation and working memory. This relationship has been pinpointed by Miyake, Friedman, Emerson, Witzki, Howerter, and

Wager (2000) using principal component analysis as being specifically located within the executive functions of inhibition and updating. In addition to that, it has been directly demonstrated that concurrent working memory-intensive tasks both increase RTs and reduce the randomness of responses on various measures of such (Cooper, Karolina, & Davelaar, 2012). Overall, the literature suggests that the working memory plays a critical role in the generation of random responses and that manipulations of working memory load lead to according increases and decreases of randomness in responses.

While there are many measures of randomness in behavioral experiments (see e.g., Towse & Neil, 1998, alone, for 14 different measures), most of them are more applicable in contexts where there are more than two response options to choose from. In this study we chose the so called local unevenness (LU), which has been used before in behavioral studies (e.g., Heuer, Janczyk, & Kunde, 2010; Heuer, Kohlisch, & Klein, 2005).

The LU of a given sequence is calculated by first dividing the sequence into sub-sequences of a chosen size (called the ‘window size’, as the sub-divided sequences can be thought of as being looked at through a window running along the total sequence). Then it is calculated how far the proportion of each response option deviates from the expected proportion of the respective response option under an assumption of randomness (here always 0.5). These deviations are then squared, added together and divided by how many response options there are. The square root of the result of that number is the LU for the window. For an example of calculating the LUs and the mean LU for a sequence of choices in different window sizes see Figure 6. The formula for the LU can be found in Formula 1.

$$(1)\ LU_w = \sqrt{\frac{(p_{left} - 0.5)^2 + (p_{right} - 0.5)^2}{2}}$$

| Window Size 2 | Window Size 4 | Window Size 6 | Window Size 8 | All Sequences (Window Size 4) |
|---|---|---|---|---|
| LRLLRRLRRRLL | LRLLRRLRRRLL | LRLLRRLRRRLL | LRLLRRLRRRLL | LLLL .5 |
| LR 0 | LRLL .25 | LRLLRR 0 | LRLLRRLR 0 | LLLR .25 |
| RL 0 | RLLR 0 | RLLRRL 0 | RLLRRLRR .125 | LLRL .25 |
| LL .5 | LLRR 0 | LLRRLR 0 | LLRRLRRR .125 | LLRR 0 |
| LR 0 | LRRL 0 | LRRLRR .166 | LRRLRRRL .125 | LRLL .25 |
| RR .5 | RRLR .25 | RRLRRR .333 | RRLRRRLL .125 | LRLR 0 |
| RL 0 | RLRR .25 | RLRRRL .166 | | LRRL 0 |
| LR 0 | LRRR .25 | LRRRLL 0 | | LRRR .25 |
| RR .5 | RRRL .25 | | | RLLL .25 |
| RR .5 | RRLL 0 | | | RLLR 0 |
| RL 0 | | | | RLRL 0 |
| LL .5 | | | | RLRR .25 |
| | | | | RRLL 0 |
| | | | | RRLR .25 |
| | | | | RRRL .25 |
| | | | | RRRR .5 |
| Mean .227 | .156 | .156 | .1 | .1875 |

***Figure 6.*** The leftmost four panels show examples of (average) values of local unevennesses (LUs) in an example sequence for window sizes 2, 4, 6, and 8. The rightmost panel shows all sequences that can occur for window size 4 and the respective LUs. The resulting ideal LU value for this window size is then .1875. (adapted from Figures 1 and 2 in Naefgen & Janczyk, 2018a)

To arrive at a measure of randomness for a given window size, all possible combinations of responses and their respective LUs are calculated. Under assumptions of randomness, all of them are equally likely to appear. The mean LU of all possible combinations of choices, then, is the ideal LU. This is illustrated in Figure 6, rightmost panel. The distance between this ideal LU and the observed LU (LUD) is a measure for lack of randomness. This somewhat awkward phrasing is due to the fact that while a LUD farther away from 0 is indicative of certain kinds of patterns (i.e. more or fewer unbalanced sequences than expected), a LUD close to 0 is not proof positive of the presence of randomness.

It is important to note that, while this measure's goal is assessing randomness, we are using it not for this purpose but for the purpose of assessing whether responses in free choice tasks behave similarly to responses in random generation tasks. Whether humans are capable of creating truly random sequences is a question that does not directly touch upon the question we are trying to answer here.

With this measure of randomness and the theoretical framework described above, we designed two experiments. In one, we supported the functioning of working memory and in the other we increased the load on it. If free choice tasks are random generation tasks, we expected them to behave like random generation tasks: Working memory support should increase the randomness of responses, while increased working memory load should reduce the randomness of the responses.

In Experiment 1, participants were instructed to give responses freely (and could do so at their own pace), to give both response options about equally often and to avoid patterns in their responses. The working memory support manipulation was implemented by displaying either zero, three or seven of the previous choices given above a constantly visible fixation circle (see Figure 7 for an illustration). This manipulation was chosen under the assumption that one of the ways in which working memory load decreases how random responses in a random generation task are is by hindering the formation of a memory trace of previous responses, thereby making it harder to monitor the chain of responses for patterns that would violate randomness. Conversely, then, making it easier to keep track of potential patterns should make it easier to avoid them.

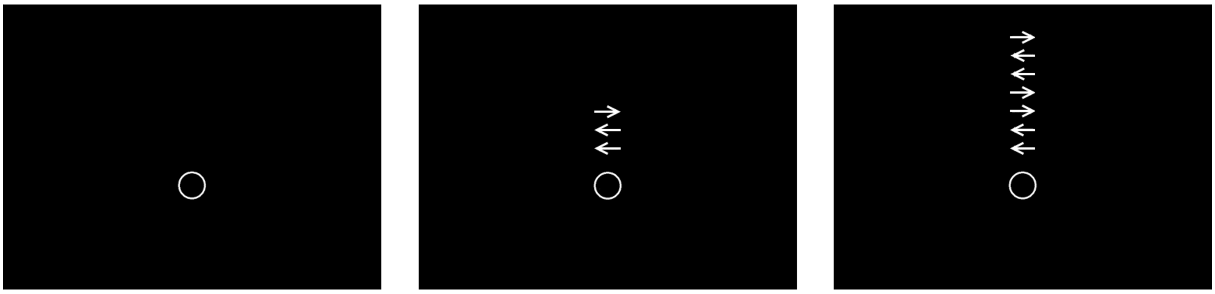

***Figure 7.*** Examples of the different working memory support conditions. In the left panel, no working memory support is given, in the middle panel three previous choices are displayed, and in the right panel seven previous choices are displayed. Not visible here are the arrows that appear within the circle for 50 ms after a response. (from Naefgen & Janczyk, 2018a)

The increase of working memory load in Experiment 2 was implemented by having participants alternatingly respond to a free choice tasks and then an *n*-back task (Kirchner, 1958, for examples of this use of the *n*-back task see Jonides, Schumacher, Smith, & Lauber, 1997; Mitchell, Macrae, & Gilchrist, 2002; Phillips, Tunstall, & Channon, 2007; Watter, Geffen, & Geffen, 2003). In the free choice task, participants were again instructed to choose freely between the two response options, choose both about equally often and to avoid patterns. In the *n*-back task, depending on the block, participants had to press a button when either the position or the color of the stimulus matched a pre-defined condition. Which condition this was also depended on the block: The stimulus could have to match a pre-defined color/position shown at the beginning of the block (0-back) or it could have to match the color/position of the stimulus going back 1, 2, or 3 trials (1-, 2- or 3-back, respectively).

As this requires constantly monitoring the last *n* *n*-back stimuli, a higher *n*-back difficulty should induce higher working memory load and, thus, reduce the randomness of the responses. For an illustration, see Figure 8.

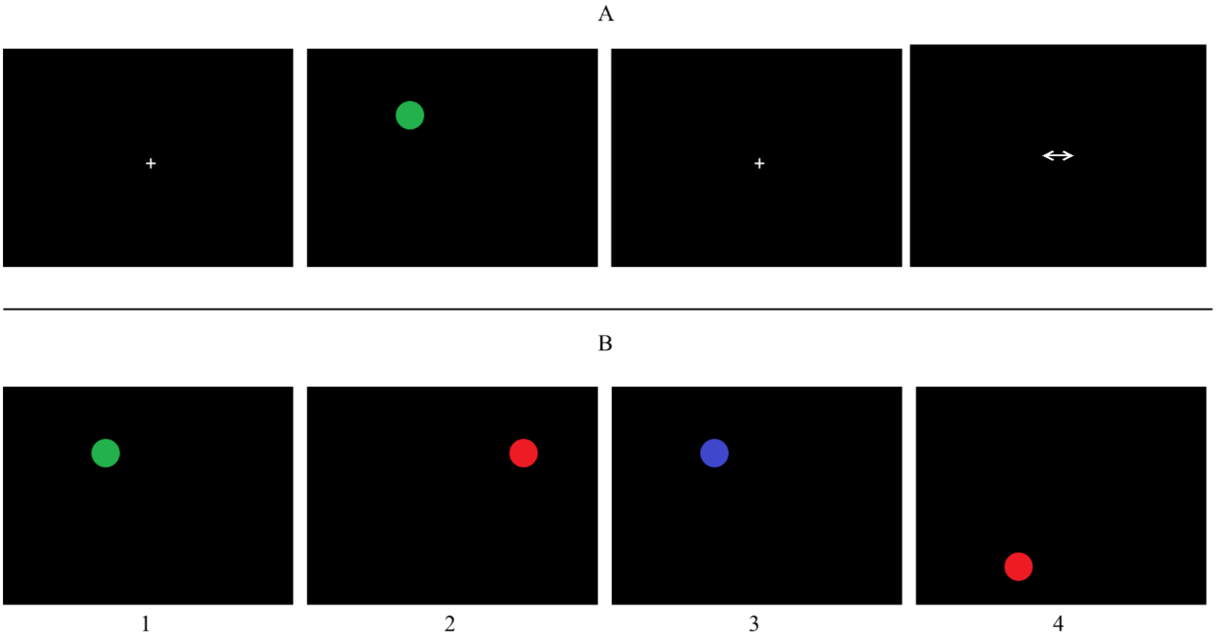

***Figure 8.*** (A) Example of a sequence of displayed stimuli and fixation crosses on the screen. (B) Example of a sequence of *n*-back stimuli (free choice stimuli not displayed). In the color-based 2-back condition, only panel (4) would require a response, while in the location-based 2-back condition, panel (3) would require a response. (from Naefgen & Janczyk, 2018a)

The LUD results of both experiments are illustrated in Figure 9. Overall, the LUDs all change in a manner consistent with predictions based on the assumption that free choice tasks are random generation tasks. That is, working memory support decreases the LUDs while working memory load increased LUDs. There were some contrasts in the results that did not reach significance, the most critical one being that more working memory support did not lead to a larger decrease of LUDs. Furthermore, the mostly positive value of the LUDs suggest a higher-than-expected proportion of unbalanced sequences (e.g., in window size 4 more sequences such as L-L-L-R, R-R-R-R, or R-L-R-R). This in turn could, for example, be indicative of more long sequences of repetition than would be expected.

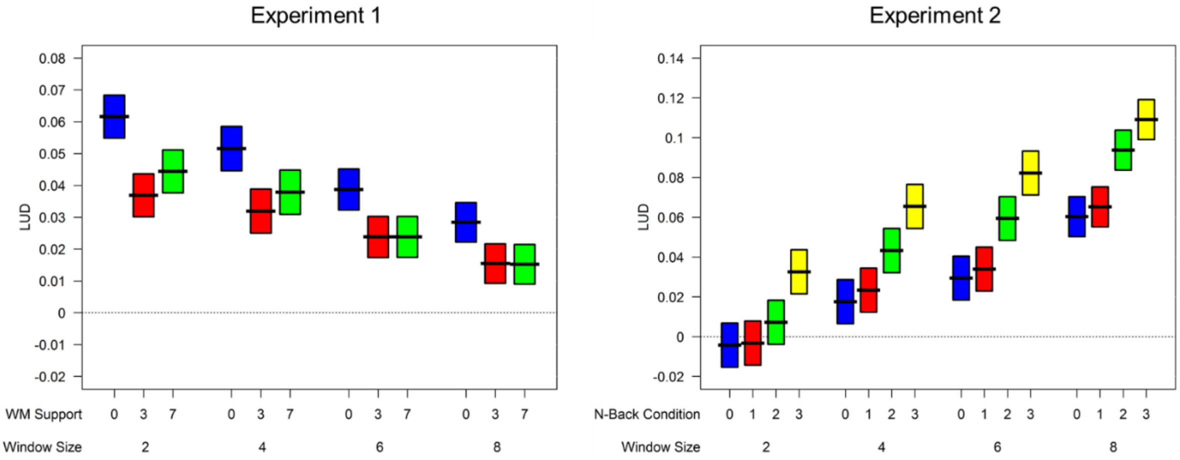

***Figure 9.*** Mean distances to ideal local unevenness (LUDs) for the window sizes 2, 4, 6, and 8 in the free choice tasks in Experiments 1 and 2 for each level of working memory (WM) support and WM load, respectively. The dotted line indicates ideal local unevenness. Error bars are 95% within confidence intervals (separate for all window sizes) (Loftus & Masson 1994).

These results are compatible with the assumption that free choice tasks are similar to random generation tasks. A careful interpretation of the results would be that free choice tasks are random-generation-like, while the strongest interpretation would be that they, in fact, *are* random generation tasks. What further reaching consequences, if any, this has for the literature on voluntary action largely depends on the assumptions one has about the relationship between freedom of choice and randomness and will be discussed in Chapter 8.

# 8 Discussion

The overall goal of the present work is to contribute to the total body of work that aims to describe and explain human action. It does so by reporting and examining evidence about free and forced choice tasks, two types of task that are used to operationalize self-generated and externally triggered actions, respectively. One distinguishing feature of this work is that the focus is on the nature of free choice tasks itself instead of trying to investigate another question by using free choice tasks. In this chapter, the central results and conclusions will be summarized. Their implications for the literature on action will be discussed and potential directions for future research on free choice tasks will be laid out.

## 8.1 Central results and conclusions

These are the answers to the research questions from Chapter 4 that the major conclusions of the previous three chapters suggest:

Question 1: **Is the reduced BCE when a free choice task is involved due to conflict adaptation?**

Study 1 suggests that the answer to this is “no”. The reduced or vanished BCE in Naefgen, Caissie et al. (2017) when Task 1 or Task 2, respectively, is a free choice task cannot be explained by conflict adaptation induced by free choice tasks, providing indirect support for the conclusions put forward by Naefgen, Caissie et al. that this is due to weaker stimulus-response links in free choice tasks. The fact that despite the reduction there was still a BCE present in Naefgen, Caissie et al. suggests some degree of similarity between the task types, as otherwise mutual interference would not be possible.

Question 2: **Are free choice tasks merely underdetermined tasks?**

Study 2 suggests that the answer to this is “no”. The mean RT difference between free and forced choice tasks can, within a sequential sampling framework, not be attributed to a difference in drift rates but rather to a difference in non-accumulation times.

Question 3: **Are free choice tasks random generation tasks?**

Study 3 suggests that the answer to this is “yes”. When working memory load is manipulated, free choice tasks are affected like random generation tasks.

The purpose of Study 1 was to investigate a potential alternative explanation for the results of Naefgen, Caissie et al. (2017). The results of that study were that when in a dual-task situation one of the two tasks was a free choice task, the BCE was reduced compared to when it was a forced choice task (when it was Task 1) or that it vanished altogether (when it was Task 2). These results were interpreted in light of the hypothesis that stimulus-response links are what drives the BCE, given the assumption that stimulus-response links are weaker or absent in free choice tasks. Study 1 was an investigation of an alternative explanation of these results: That free choice tasks, given that they are assumed to be conflict tasks (e.g., Berlyne, 1957, or Botvinick et al., 2001), could lead to immediate conflict adaptation, thereby reducing the BCE. This hypothetical conflict was manipulated by changing (1) whether or not the prime (dual-tasking) trial before the target trial had a free or a forced choice Task 1 and (2) the amount of trials in a block that had one or the other task type as Task 1. As neither manipulation had an effect on the size of the BCE, there was no evidence for this alternative explanation. Another notable observation from Study 1 is that the choices in a free choice Task 1, as in Naefgen, Caissie et al., were biased towards the response required by the Task 2 stimulus, which is potentially another way backward crosstalk can manifest.

The purposes of Studies 2 and 3, in contrast, were more direct investigations of free choice tasks themselves. Study 2 viewed free and forced choice tasks through a sequential sampling lens and sought to attribute the on average longer RTs of free choice tasks compared to forced choice tasks to parameters within such a framework. The minimal framework (cf. Grice, 1968) used here relied on just three parameters: the non-accumulation time, the drift rate, and the decision thresholds. The manipulation of the decision thresholds, which was used to locate the RT difference within the other two parameters, was operationalized blockwise by changing the amount of catch trials (Gordon, 1967, Näätänen, 1972) and the time limits of trials (Diederich, 1997, Dror et al., 1999). As there was no reliable change in the mean RT differences, the main conclusion was that the mean RT difference between the task types can be attributed to a difference in the non-accumulation times.

Study 3 sought to investigate the question whether free choice tasks have one specific mechanical quality previously (Frith, 2013) attributed to them: Are they (like) random generation tasks? As forced choice tasks obviously are not random generation tasks, this study did not take a comparative approach between the two task types. The way the question was investigated was by manipulating working memory load while free choice tasks were performed. In random generation tasks, it has been shown that this influences the randomness of the responses (e.g., Baddeley, 1966, Cooper et al., 2012). Working memory load was experimentally decreased (by showing previous responses to a free choice task on screen) and increased (by having participants perform a simultaneous *n*-back task). The randomness of the responses was measured by their distance from ideal local unevenness (Heuer et al., 2005, 2010), with a larger distance corresponding to less randomness. The results were consistent with the interpretation of free choice tasks as random generation tasks, that is, working memory support increased and higher working memory load decreased the randomness of the responses.

## 8.2 Implications and outlook

The purpose of this section is to place the results of the studies included in this work in a larger theoretical context and discuss their wider-ranging implications.

Free choice tasks as conflict tasks. While the purpose of Study 1 was to investigate an alternative explanation for Naefgen, Caissie et al. (2017), it also has implications for the interpretation of free choice tasks as conflict tasks[2]. There are two basic interpretations with regards to the results as they presented themselves: Either free choice tasks produce no cognitive conflict or the type of conflict they produce does somehow not affect the BCE. The former appears unlikely, given that free choice tasks by definition involve conflict on an operational level, as at least two responses are connected to each stimulus. But this does not necessarily mean that conflict arises also on a level of cognitive mechanisms. To illustrate this point: While at all times we have a radical freedom to do anything we are physically capable of doing, we usually do not feel conflicted about all of these options. In order to perceive this conflict as conflict, we have to cognitively process our options in some specific ways, usually by restricting the space of possible actions in some way (e.g., by asking ourselves what we should eat for dinner[3]). However, the space of options in free choice tasks is both highly limited and highly salient, so it appears unlikely that the conflict is not perceived by participants. Furthermore, neuroimaging evidence (summarized in Botvinick et al., 2001)

---

[2] Conflict is here to be understood differently than in Berlyne (1957). Berlyne assumed that free choice tasks are tasks involving conflict between response options a priori and then used them to investigate what effectively influences the strength of this conflict. As such, in Berlyne's work it was never in question whether or not free choice tasks are conflict tasks, whereas here the question is more whether free choice tasks lead to *cognitive* conflict in the sense of, for example, Botvinick et al. (2001), or if there is another cognitive mechanism at work that does not involve conflict.

[3] Asking someone else what we should eat for dinner can also lead to conflict, albeit of a different kind than discussed here.

shows that similar brain regions are active when performing free choice tasks and other conflict-inducing tasks. It seems therefore more likely that there is cognitive conflict but that this conflict does not translate into increased task shielding from the BCE. One possibility can be found in the reported sensitivity of conflict adaptation to the types of conflict involved. According to a prediction by Braem et al. (2014), the strength of the transfer of congruency sequence effects has a U-shaped relationship with the (dis)similarity of the task contexts. If this is true, it could be that here the tasks contexts were neither sufficiently similar nor dissimilar to lead to conflict adaptation.

There are two closely related venues for further research here: First, investigating to which, if any, conflict tasks the conflict created by free choice tasks generalizes for the purposes of conflict adaptation. Second, investigating the reverse, that is, finding out which conflict tasks can induce conflict adaptation that affects free choice tasks. The latter would potentially reap additional theoretical benefits, as it would simultaneously be an investigation into whether or not free choice tasks actually are conflict tasks. If a conflict task were to be identified that induces conflict adaptation that in turn increases response speed in free choice tasks, that would imply that a part of the mean RT difference between free and forced choice tasks is due to conflicting response activation. A straightforward design for investigating this generalizability similar to the design used in Study 1, Experiment 1, would be to have alternating prime and target trials. The target trials are either free or forced choice tasks while the prime trials are trials of conflict tasks that are either compatible or incompatible in any given prime-target pair of trials. Following an incompatible prime trial, free choice tasks should then be closer in their RTs to forced choice tasks, compared to when the two tasks are following a compatible prime trial. This leaves open the question of which conflict tasks are suitable candidates. Following Braem et al.'s (2014) guidelines, the task context should either be very similar to the assumed conflict in free choice tasks or very dissimilar. As the conflict in free choice tasks is assumed to arise from conflicting activated response options, a task that potentially is similar could be the Eriksen flanker task (Eriksen & Eriksen, 1974). In this task type, a central stimulus that indicates one response is flanked by either compatible or incompatible stimuli that either indicate the same response or the other response, respectively. Something to keep in mind when following this approach is that it is on some level exploratory, which here implies that negative results are not conclusive, as it could just be the wrong type of conflict task that was used. Note that according to Braem et al.'s overview, conflict adaptation is not only dependent on the type of conflict task but, possibly to a much larger extent, on the context and implementation of these tasks, like for example the types of stimuli used or the respective response modes. This means that great care needs to be taken when selecting the specifics of both prime and target trial tasks.

Backwards biasing of free choice task response choices — A new type of BCE? Study 1 also replicated the observation from Naefgen, Caissie et al. (2017) that the choice made in a free choice Task 1 is significantly more likely to be the same as the choice instructed by a forced choice Task 2 than a different choice. In Naefgen, Caissie et al. the same congruency bias was reported when the free choice task was the second task. An open question here is if the process which brings rise to this bias has the same transmission mechanism as the BCE that affects the RTs.

A possible general approach to answering this question could be to apply a manipulation that reduces the RT BCE and see if it also reduces the choice bias “BCE”. A potential candidate for this could be the sequential modulation reported in Janczyk (2016). Here, only forced choice tasks were used and when a trial followed an incompatible trial, there was no BCE while in trials following a compatible trial, there was a BCE. If, then, a trial in which Task 1 is a free choice trial follows a (forced choice-forced choice) trial that is incompatible, the RT BCE should vanish (assuming that the mechanism behind the free choice BCE is the same as the one in the forced choice BCE) and whether or not the bias vanishes would be evidence for or against, respectively, the same mechanism involved in the choice “BCE”.

This was possible with the dataset of Experiment 1 in Naefgen and Janczyk (2018b). An analysis of the choices in the free choice prime trials indicated that when the previous forced choice-forced choice trial was a compatible trial, participants were more likely than not (71.0%) to choose the same response in the free choice Task 1 as instructed by the forced choice Task 2, one sample *t*-test against 50%: $t(35) = 11.59$, $p < .001$, $d = 2.73$. This effect was not present when the previous forced choice-forced choice trial was incompatible and even nominally in the opposite direction (46.0%), one sample *t*-test against 50%: $t(35) = -1.70$, $p = .098$, $d = -0.40$. The difference in the percentages of same free choice Task 1 answers between these two conditions was significant, $t(35) = 12.65$, $p < .001$, $d = 2.98$. This suggests that the same mechanisms are involved in this choice BCE and the BCE that affects RTs. While further research into this similarity between the kinds of BCE is necessary, this result supports claims of similarity between free and forced choice tasks, as there needs to be at least some similarity for crosstalk to be possible.

Locating the RT difference between free and forced choice tasks. If the RT difference results from pre-central perceptual processing advantages (Janczyk, Dambacher et al., 2015) and the rate of information accumulation is identified as the parameter where perceptual processing is located[4], attributing the RT difference to a difference in non-accumulation time like Study 2 suggests leads to a contradiction. Also, the results of Study 3 suggest that free choice tasks involve the generation/selection of random responses. While it is conceivable that the additional non-accumulation time consists of additional encoding as well as the generation/selection of a random response, it would be necessary to investigate whether Janczyk, Dambacher et al.'s postulated perceptual advantage can be attributed to non-accumulation times instead of drift rates. Resolving this issue certainly is a task for future research. Another open question here is why Berlyne (1957) reported an advantage for brighter free choice stimuli while Janczyk, Dambacher et al., in their Experiment 3, did not.

Formal models of free and forced choice tasks. A potential way to learn more about free choice tasks would be to create a unified formal model of them. Such a model would have to incorporate the limitations set by the literature and would bring with it the potential for further predictions and insights. Mattler and Palmer (2012) for example applied an accumulator model to priming effects in free and forced choice tasks. Their accumulator model was based on the accumulator model described by Vorberg, Mattler, Heinecke, Schmidt, and Schwarzbach (2003), which in turn was an attempt to model priming in forced choice tasks. In this model, there are two accumulators of sensory evidence (one for each of the response options), inhibiting each other. As long as the stimulus from which these accumulators receive activation is present, even if it is just in a sensory buffer (like what a masked prime would cause), it will create activation in the corresponding accumulator. Once the difference between the activation of the accumulators is high enough, the response corresponding to the accumulator with the higher accumulation is emitted. Together with some other slight deviations from this model, Mattler and Palmer added two assumptions to adapt this model to free choice tasks. (1) That a free choice stimulus leads to "a randomly chosen activation of one of the two accumulators from some internal source with an activation rate that is independent from stimulus characteristics" (p. 356) and (2) that for free choice trials the evidence required to emit a response is reduced over time. In other words, the decision thresholds (see Chapter 6) are lowered once the free choice stimulus appears. While

---

[4] For example Voss, Rothermund, and Voss (2004) who interpreted "the drift [rate] as a measure of perceptual sensitivity (in a between-person comparison) or as a measure of task difficulty (in a between-condition comparison)" (p. 1208).

Mattler and Palmer’s data from three experiments fit this model quite well, there are some theoretical questions about the assumptions of the model that still need to be answered. While acceptable as a provisional assumption, it is entirely unclear *how* the activation of one of the accumulators is “chosen randomly” and needs to be specified further in the future. Furthermore, while the mathematical modeling-side of the falling decision thresholds is clear, it is somewhat unclear what they mean from a viewpoint of mechanistic explanations. Presumably, there would need to be some sort of encoding process that changes the behavior of the decision thresholds from static (in forced choice tasks) to dynamic (in free choice tasks). Lastly, there are phenomena that this model cannot predict yet, like for example the tendency for repeated responses in free choice tasks (i.e., giving the same response to multiple free choice stimuli in a row in a frequency above chance level, reported in Naefgen & Janczyk, 2018a).

Another notable example of a mathematical model is the one presented by Devaine, Waszak, and Mamassian (2013). They reported on a two-stage model of action control where the two stages represent the dissociation in the first and the combination of two assumed types of action control in the second stage. In this model there are two variables, one for internal actions and one for external actions. In the first stage, they accumulate independently and constantly. The internal variable accumulates for a randomly chosen response option. Once both pass a first threshold, the second phase begins in which, depending on whether a trial is 'congruent' (i.e. the two variables code for the same responses) or not, the internal variable is either facilitated (congruent) or inhibited (incongruent) by the external variable. Once one of the variables crosses a second threshold, the response it codes for is emitted. This model has the downside that it is somewhat specific in its application to the experimental setup on which it was tested, so some work would have to be done to generalize it to other contexts. It also is somewhat questionable to what extent this model applies to free choice tasks per se, as the internal generation in the experimental setup here is framed as an anticipation of a future (forced choice) response and not as a self-directed choice. However, despite the specificity, this model also provides a theoretical element that builders of a future models may want to consider using: Viewing the two types of action control not as incompatible systems but as systems that work in parallel.

A potential further source for insights into what constraints are sensible for formal models of free choice tasks are dynamic measurements such as the mouse tracking used by

Vogel et al. (2018), as they provide insight into what happens between stimulus onset and emission of a response.

The relationship between free choice tasks, voluntary action, and randomness. But what are the broadest implications of the results presented in this work? What would be the implications of free choice tasks only differing from forced choice tasks by an additional necessary step in which a random response is chosen, making them essentially random generation tasks? Does this completely invalidate any work on self-generated, voluntary actions in which free choice tasks are used as an operationalization?

These are, of course, questions that cannot easily be decided. The answers largely depend on the assumptions one has about voluntary actions and freedom of choice: If one assumes that, at least, some randomness is necessary for choices to be truly free, these results and their strongest interpretation would not be problematic whatsoever, even expected and encouraging for this use of free choice tasks. If, on the other hand, one assumes that randomness is in contradiction with intentional actions, as it lacks any expression of directed will, it would render free choice tasks an untenable operationalization of this type of action. Either way it is necessary to be open about these assumptions, for reasons of both clarity of the theoretical constructs one is working with and of clarity of communication, so that readers of research can categorize the definitions for themselves.

Another important point to keep in mind here is that while the present work, most specifically in Study 3, mainly investigated whether free choice tasks are random generation tasks, but not necessarily whether or not responses to these tasks are, in fact, random (or, in other words, unpredictable). However, the issue of randomness and predictability of free choices is one empirically touched upon both here and in other works. One example here are Lages and Jaworska (2012), who used multivariate pattern analysis, a type of machine learning, to predict free choice responses with better than chance accuracy (note that the timing for the responses was unspecific and instructed as “when participants felt the urge to do so” (p. 2), similarly to Libet et al. (1983)). If supposedly free choices can be predicted, they cannot be entirely random. But again, the deeper implications of this for the use of free choice tasks are entirely dependent on the assumptions one has about them.

But maybe there is a criterion, at least for empirical psychologists, to decide between those premises: The endeavor of describing and explaining phenomena in a scientific manner, at its practical core, runs counter to the idea of fundamentally unpredictable phenomena (for a similar discussion, see Prinz, 2004, which has a stronger focus on free will as opposed to free choice). The attempt may not succeed, but without the assumption that it is, at least in principle, even if not practically, possible to explain a phenomenon, it would be moot to even try. Therefore, people who try to do this, may be inclined towards deterministic ideas of voluntary action. Whether this is true, is, of course, an empirical question.

Freedom of choice: Multiple concepts within one term? I will end this discussion with an appeal. I claim that the term *freedom* of choice is not a useful (or even clearly defined) psychological concept and should be abandoned and replaced with multiple other terms. I argue that this term actually refers to multiple, different concepts that fall under the umbrellas of (1) freedom as the presence of viable options and (2) freedom as expressions of one's own will. While these two categories of concepts are not necessarily mutually exclusive, their focuses are distinct and they lead to different questions. For the former class, the amount of different response options (including the response but also potentially including the time of the response, such as in Libet et al., 1983) that fulfill a given goal are critical, be they the different buttons in a free choice task experiment, several brands of identical-seeming paper towels in a store or something else. It lends itself well to being operationalized in a laboratory, as its focus is on the environment in which an action is performed. The latter class, on the other hand, is not just about the context of an action but about a subjective experience, an attribution, of the goal of an action (see e.g., Chapter 2 of Prinz, 2004, for a discussion of related conceptualizations of the self and free will). Here it matters more, for example, how much a person identifies with the action goals that drive an action and how much the person attributes the action to themselves.

It is important to be mindful and explicit about which type of concept one wants to investigate, as the two represent very different phenomena. Free choice tasks as understood here, for example, would clearly operationalize category (1) but it is dubious at best to what extent being allowed to press either a left or a right button represents the own will of the participants in the sense of category (2) better than being allowed only one of these options at a time. Similarly, it is difficult to gauge to what extent participants' "urge" to give a response in Libet et al.'s (1983) study was their own or an artifact of the instructions given. Changing the focus away from freedom and towards these more specific definitions also entirely bypasses the discussion of randomness and how it relates to freedom of choice (or at least puts it in a framework founded on psychological theories instead of ontological assumptions, the former of which are much easier agreed upon).

## 8.3 Summary

The aim of the present work was to investigate the nature of the free choice task and contrast it with that of the forced choice task. In three studies, it was shown that free choice tasks do not lead to a type of conflict adaptation that reduces the BCE, that within a sequential sampling framework the slower responses in free choice tasks than in forced choice tasks can be attributed to a longer non-accumulation time instead of a dearth of evidence for one response or the other and that free choice tasks are affected like random generation tasks by working memory load manipulations, suggesting that random response generation is a component of the free choice task.

Future research should focus on integrating seemingly contradictory results in the literature, potentially by creating a unified formal model of free and forced choice tasks.

# References

Akçay, Ç., & Hazeltine, E. (2007). Conflict monitoring and feature overlap: Two sources of sequential modulations. *Psychonomic Bulletin & Review*, *14*(4), 742–748. doi:10.3758/BF03196831

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, *79*(5), 882–898. doi:10.1007/s00426-014-0608-y

Arvola, A., & Forsander, O. (1961). Comparison between water and alcohol consumption in six animal species in free-choice experiments. *Nature*, *191*(4790), 819–820. doi:10.1038/191819a0

Astor-Jack, T., & Haggard, P. (2005). Intention and reactivity. In G. Humphreys & J. Riddoch (Eds.), *Attention in Action* (pp. 109–130). London, UK: Taylor & Francis. Retrieved from doi:10.4324/9780203449226

Baddeley, A. D. (1962). Some factors influencing the generation of random letter sequences. *Medical Research Council Applied Psychological Research Unit Report*, *422/62*.

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, *18*(2), 119–129. doi:10.1080/14640746608400019

Berlyne, D. E. (1957). Conflict and choice time. *British Journal of Psychology*, *48*(2), 106–118. doi:10.1111/j.2044-8295.1957.tb00606.x

Bermeitinger, C., & Hackländer, R. P. (2018). Response priming with motion primes: negative compatibility or congruency effects, even in free-choice trials. *Cognitive Processing*, 1–11. doi:10.1007/s10339-018-0858-5

Bodner, G. E., & Mulji, R. (2009). Prime proportion affects masked priming of fixed and free-choice responses. *Experimental Psychology*, *57*(5), 360–366. doi:10.1027/1618-3169/a000043

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Braem, S., Abrahamse, E. L., Duthoo, W., & Notebaert, W. (2014). What determines the specificity of conflict adaptation? A review, critical analysis, and proposed synthesis. *Frontiers in Psychology*, *5*, 1134. doi:10.3389/fpsyg.2014.01134

Brass, M., & Haggard, P. (2007). To do or not to do: The neural signature of self-control. *Journal of Neuroscience*, *27*(34), 9141–9145. doi:10.1523/JNEUROSCI.0924-07.2007

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist*, *14*(4), 319–325. doi:10.1177/1073858408317417

Brysbaert, M. (1994). Behavioral estimates of interhemispheric transmission time and the signal detection method: A reappraisal. *Perception & Psychophysics*, *56*(4), 479–490. doi:10.3758/BF03206739

Cooper, R. P., Karolina, W., & Davelaar, E. J. (2012). Differential contributions of set-shifting and monitoring to dual-task interference. *Quarterly Journal of Experimental Psychology*. Retrieved from http://journals.sagepub.com/doi/10.1080/17470218.2011.629053

Dambacher, M., & Hübner, R. (2015). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological Research*, *79*(1), 83–94. doi:10.1007/s00426-014-0542-z

Deiber, M.-P., Passingham, R. E., Colebatch, J. G., Friston, K. J., Nixon, P. D., & Frackowiak, R. S. J. (1991). Cortical areas and the selection of movement: a study with positron emission tomography. *Experimental Brain Research*, *84*(2), 393–402. doi:10.1007/BF00231461

Devaine, M., Waszak, F., & Mamassian, P. (2013). Dual process for intentional and reactive decisions. *PLOS Computational Biology*, *9*(4). doi:10.1371/journal.pcbi.1003013

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*(3), 260–274. doi:10.1006/jmps.1997.1167

Dignath, D., Janczyk, M., & Eder, A. B. (2017). Phasic valence and arousal do not influence post-conflict adjustments in the Simon task. *Acta Psychologica*, *174*, 31–39. doi:10.1016/j.actpsy.2017.01.004

Dror, I. E., Basola, B., & Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & Cognition*, *27*(4), 713–725. doi:10.3758/BF03211564

Dutzi, I. B., & Hommel, B. (2009). The microgenesis of action-effect binding. *Psychological Research PRPF*, *73*(3), 425–435. doi:10.1007/s00426-008-0161-7

Ebert, J. P., & Wegner, D. M. (2011). Mistaking randomness for free will. *Consciousness and Cognition*, *20*(3), 965–971. doi:10.1016/j.concog.2010.12.012

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240. doi:10.1037/0096-1523.27.1.229

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. doi:10.3758/BF03203267

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Cramon, D. Y. von, Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538–17542. doi:10.1073/pnas.0805903105

Freitas, A. L., Bahar, M., Shan Yang, & Banai, R. (2007). Contextual adjustments in cognitive control across tasks. *Psychological Science (0956-7976)*, *18*(12), 1040–1043. doi:10.1111/j.1467-9280.2007.02022.x

Frith, C. (2013). The psychology of volition. *Experimental Brain Research*, *229*(3), 289–299. doi:10.1007/s00221-013-3407-6

Frith, C. D., Friston, K., Liddle, P. F., & Frackowiak, R. S. (1991). Willed action and the prefrontal cortex in man: a study with PET. *Proceedings. Biological Sciences*, *244*(1311), 241–246. doi:10.1098/rspb.1991.0077

Goodale, M. A., Westwood, D. A., & Milner, A. D. (2004). Two distinct modes of control for object-directed action. *Progress in Brain Research*, *144*, 131–144.

Gordon, I. E. (1967). Stimulus probability and simple reaction time. *Nature*, *215*(5103), 895–896. doi:10.1038/215895a0

Goschke, T., & Dreisbach, G. (2008). Conflict-triggered goal shielding: Response conflicts attenuate background monitoring for prospective memory cues. *Psychological Science*, *19*(1), 25–32. doi:10.1111/j.1467-9280.2008.02042.x

Gozli, D. G., Huffman, G., & Pratt, J. (2016). Acting and anticipating: Impact of outcome-compatible distractor depends on response selection efficiency. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1601–1614. doi:10.1037/xhp0000238

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506. doi:10.1037/0096-3445.121.4.480

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, *75*(5), 359–373.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982). Human reaction time: Toward a general theory. *Journal of Experimental Psychology: General*, *111*(1), 135–153. doi:10.1037/0096-3445.111.1.135

Hadland, K. A., Rushworth, M. F. S., Passingham, R. E., Jahanshahi, M., & Rothwell, J. C. (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. *Journal of Cognitive Neuroscience*, *13*(8), 1097–1108. doi:10.1162/089892901753294392

Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1540–1554. doi:10.1080/17470210601119134

Herwig, A., & Waszak, F. (2009). Intention and attention in ideomotor learning. *The Quarterly Journal of Experimental Psychology*, *62*(2), 219–227. doi:10.1080/17470210802373290

Herwig, A., & Waszak, F. (2012). Action-effect bindings and ideomotor learning in intention- and stimulus-based actions. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00444

Heuer, H., Janczyk, M., & Kunde, W. (2010). Random noun generation in younger and older adults. *The Quarterly Journal of Experimental Psychology*, *63*(3), 465–478. doi:10.1080/17470210902974138

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. *Quarterly Journal of Experimental Psychology: Section A*, *58*(2), 275–307. doi:10.1080/02724980343000855

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1368–1384. doi:10.1037/0096-1523.24.5.1368

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and Performance, Vol. XVIII* (pp. 247–273). Cambridge, MA: MIT Press.

Hommel, B., & Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends in Cognitive Sciences*, *21*(12), 940–949. doi:10.1016/j.tics.2017.09.009

Huffman, G., Gozli, D. G., Hommel, B., & Pratt, J. (2018). Response preparation, response selection difficulty, and response-outcome learning. *Psychological Research*, 1–11. doi:10.1007/s00426-018-0989-4

Jahanshahi, M., Dirnberger, G., Fuller, R., & Frith, C. D. (2000). The role of the dorsolateral prefrontal cortex in random number generation: a study with positron emission tomography. *NeuroImage*, *12*(6), 713–725. doi:10.1006/nimg.2000.0647

Janczyk, M. (2016). Sequential modulation of backward crosstalk and task-shielding in dual-tasking. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(5), 631–647. doi:10.1037/xhp0000170

Janczyk, M., Dambacher, M., Bieleke, M., & Gollwitzer, P. M. (2015). The benefit of no choice: goal-directed plans enhance perceptual processing. *Psychological Research*, *79*(2), 206–220. doi:10.1007/s00426-014-0549-5

Janczyk, M., Durst, M., & Ulrich, R. (2017). Action selection by temporally distal goal states. *Psychonomic Bulletin & Review*, *24*(2), 467–473. doi:10.3758/s13423-016-1096-4

Janczyk, M., Heinemann, A., & Pfister, R. (2012). Instant attraction: Immediate action-effect bindings occur for both, stimulus- and goal-driven actions. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00446

Janczyk, M., & Leuthold, H. (2018). Effector system-specific sequential modulations of congruency effects. *Psychonomic Bulletin & Review*, *25*(3), 1066–1072. doi:10.3758/s13423-017-1311-y

Janczyk, M., Nolden, S., & Jolicoeur, P. (2015). No differences in dual-task costs between forced- and free-choice tasks. *Psychological Research*, *79*(3), 463–477. doi:10.1007/s00426-014-0580-6

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General*, *141*(3), 489–501. doi: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*(1), 30–43. doi:10.1016/j.cognition.2014.03.001

Janczyk, M., Pfister, R., Wallmeier, G., & Kunde, W. (2014). Exceptions to the PRP effect? A comparison of prepared and unconditioned reflexes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 776–786. doi:10.1037/a0035548

Janczyk, M., Renas, S., & Durst, M. (2018). Identifying the locus of compatibility-based backward crosstalk: Evidence from an extended PRP paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 261–276. doi:10.1037/xhp0000445

Jenkins, I. H., Jahanshahi, M., Jueptner, M., Passingham, R. E., & Brooks, D. J. (2000). Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. *Brain: A Journal of Neurology*, *123 ( Pt 6)*, 1216–1228.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, *9*(4), 462–475. doi:10.1162/jocn.1997.9.4.462

Keller, P. E., Wascher, E., Prinz, W., Waszak, F., Koch, I., & Rosenbaum, D. A. (2006). Differences between intention-based and stimulus-based actions. *Journal of Psychophysiology*, *20*(1), 9–20. doi:10.1027/0269-8803.20.1.9

Kiesel, A., Wagener, A., Kunde, W., Hoffmann, J., Fallgatter, A. J., & Stöcker, C. (2006). Unconscious manipulation of free choice in humans. *Consciousness and Cognition*, *15*(2), 397–408. doi:10.1016/j.concog.2005.10.002

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358. doi:10.1037/h0043688

Kola-Olusanya, A. (2005). Free-choice environmental education: understanding where children learn outside of school. *Environmental Education Research*, *11*(3), 297–307. doi:10.1080/13504620500081152

Krieghoff, V., Waszak, F., Prinz, W., & Brass, M. (2011). Neural and behavioral correlates of intentional actions. *Neuropsychologia*, *49*(5), 767–776. doi:10.1016/j.neuropsychologia.2011.01.025

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 387–394. doi:10.1037/0096-1523.27.2.387

Kunde, W., Pfister, R., & Janczyk, M. (2012). The locus of tool-transformation costs. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 703–714. doi:10.1037/a0026315

Lages, M., & Jaworska, K. (2012). How predictable are “spontaneous decisions” and “hidden intentions”? Comparing classification results based on previous responses with multivariate pattern analysis of fMRI BOLD signals. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00056

Lau, H. C., Rogers, R. D., Ramnani, N., & Passingham, R. E. (2004). Willed action and attention to the selection of action. *NeuroImage*, *21*(4), 1407–1415. doi:10.1016/j.neuroimage.2003.10.034

Le Bars, S., Hsu, Y.-F., & Waszak, F. (2016). The impact of subliminal effect images in voluntary vs. stimulus-driven actions. *Cognition*, *156*, 6–15. doi:10.1016/j.cognition.2016.07.005

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*.

Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. *Psychonomic Bulletin & Review*, *9*(2), 212–238. doi:10.3758/BF03196277

Linser, K., & Goschke, T. (2007). Unconscious modulation of the conscious experience of voluntary control. *Cognition*, *104*(3), 459–475. doi:10.1016/j.cognition.2006.07.009

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi:10.3758/BF03210951

Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition*, *123*(3), 347–360. doi:10.1016/j.cognition.2012.03.002

McKenzie, C. R. M., Wixted, J. T., Noelle, D. C., & Gyurjyan, G. (2001). Relation between confidence in yes–no and forced-choice tasks. *Journal of Experimental Psychology: General*, *130*(1), 140–155. doi:10.1037/0096-3445.130.1.140

Mitchell, J. P., Macrae, C. N., & Gilchrist, I. D. (2002). Working memory and the suppression of reflexive saccades. *Journal of Cognitive Neuroscience*, *14*(1), 95–103. doi:10.1162/089892902317205357

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Näätänen, R. (1972). Time uncertainty and occurence uncertainty of the stimulus in a simple reaction time task. *Acta Psychologica*, *36*(6), 492–503. doi:10.1016/0001-6918(72)90029-7

Naefgen, C., Caissie, A. F., & Janczyk, M. (2017). Stimulus-response links and the backward crosstalk effect — A comparison of forced- and free-choice tasks. *Acta Psychologica*, *177*, 23–29. doi:10.1016/j.actpsy.2017.03.010

Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: evidence from response threshold manipulations. *Psychological Research*, *82*(6), 1039–1052. doi:10.1007/s00426-017-0887-1

Naefgen, C., & Janczyk, M. (2018a). Free choice tasks as random generation tasks: an investigation through working memory manipulations. *Experimental Brain Research*, *236*(8), 2263–2275. doi:10.1007/s00221-018-5295-2

Naefgen, C., & Janczyk, M. (2018b). Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation. *Cognitive Processing*. doi:10.1007/s10339-018-0887-0

Obhi, S. S., & Haggard, P. (2004). Internally generated and externally triggered actions are physically distinct and independently controlled. *Experimental Brain Research*, *156*(4), 518–523. doi:10.1007/s00221-004-1911-4

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(3), 358–377.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*(2), 220–244.

Passingham, R. E. (2007). Two cortical systems for directing movement. In *Ciba Foundation Symposium 132 - Motor Areas of the Cerebral Cortex* (pp. 151–170). Wiley-Blackwell. doi:10.1002/9780470513545.ch10

Passingham, Richard E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14*(1), 16–21. doi:10.1016/j.tics.2009.11.001

Pfister, R., Heinemann, A., Kiesel, A., Thomaschke, R., & Janczyk, M. (2012). Do endogenous and exogenous action control compete for perception? *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 279–284. doi:10.1037/a0026658

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74–80. doi:10.2478/v10053-008-0133-x

Pfister, R., Kiesel, A., & Melcher, T. (2010). Adaptive control of ideomotor effect anticipations. *Acta Psychologica*, *135*(3), 316–322. doi:10.1016/j.actpsy.2010.08.006

Pfister, R., & Kunde, W. (2013). Dissecting the response in response–effect compatibility. *Experimental Brain Research*, *224*(4), 647–655. doi:10.1007/s00221-012-3343-x

Phillips, L. H., Tunstall, M., & Channon, S. (2007). Exploring the role of working memory in dynamic social cue decoding using dual task methodology. *Journal of Nonverbal Behavior*, *31*(2), 137–152. doi:10.1007/s10919-007-0026-6

Playford, E. D., Jenkins, I. H., Passingham, R. E., Nutt, J., Frackowiak, R. S. J., & Brooks, D. J. (1992). Impaired mesial frontal and putamen activation in Parkinson's disease: A positron emission tomography study. *Annals of Neurology*, *32*(2), 151–161. doi:10.1002/ana.410320206

Prinz, W. (2004). Kritik des freien Willens: *Psychologische Rundschau*, *55*(4), 198–206. doi:10.1026/0033-3042.55.4.198

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226–1243. doi:10.1037/a0036801

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi:10.1037/0033-295X.85.2.59

Renas, S., Durst, M., & Janczyk, M. (2018). Action effect features, but not anatomical features, determine the backward crosstalk effect: Evidence from crossed-hands experiments. *Psychological Research*, *82*(5), 970–980. doi:10.1007/s00426-017-0873-7

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*(2), 261–282. doi:10.1037/0096-3445.133.2.261

Scherbaum, S., Fischer, R., Dshemuchadse, M., & Goschke, T. (2011). The dynamics of cognitive control: Evidence for within-trial conflict adaptation from frequency-tagged EEG. *Psychophysiology*, *48*(5), 591–600. doi:10.1111/j.1469-8986.2010.01137.x

Scherbaum, S., Gottschalk, C., Dshemuchadse, M., & Fischer, R. (2015). Action dynamics in multitasking: the impact of additional task factors on the execution of the prioritized motor movement. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00934

Schlaghecken, F., & Eimer, M. (2004). Masked prime stimuli can bias “free” choices between response alternatives. *Psychonomic Bulletin & Review*, *11*(3), 463–468. doi:10.3758/BF03196596

Schüür, F., & Haggard, P. (2011). What are self-generated actions? *Consciousness and Cognition*, *20*(4), 1697–1704. doi:10.1016/j.concog.2011.09.006

Seibold, V. C., Bausenhart, K. M., Rolke, B., & Ulrich, R. (2011). Does temporal preparation increase the rate of sensory information accumulation? *Acta Psychologica*, *137*(1), 56–64. doi:10.1016/j.actpsy.2011.02.006

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. doi:10.1037/h0020586

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545. doi:10.1038/nn.2112

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. doi:10.1016/0001-6918(69)90055-9

Stevens, J. R., Rosati, A. G., Ross, K. R., & Hauser, M. D. (2005). Will travel for food: Spatial discounting in two New World monkeys. *Current Biology*, *15*(20), 1855–1860. doi:10.1016/j.cub.2005.09.016

Stürmer, B., & Leuthold, H. (2003). Control over response priming in visuomotor processing: a lateralized event-related potential study. *Experimental Brain Research*, *153*(1), 35–44. doi:10.1007/s00221-003-1579-1

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1345–1363. doi:10.1037/0096-1523.28.6.1345

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*(4), 583–591. doi:10.3758/BF03209475

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences*, *100*(10), 6275–6280. doi:10.1073/pnas.0931489100

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. doi:10.3758/BF03196893

Wagenmakers, E.-J., Maas, H. L. J. V. D., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. doi:10.3758/BF03194023

Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, *162*(3), 346–356. doi:10.1007/s00221-004-2183-8

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003. doi:10.1111/1469-8986.3860998

Wirth, R., Janczyk, M., & Kunde, W. (2018). Effect monitoring in dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 553–571. doi:10.1037/xlm0000474

Wolfensteller, U., & Ruge, H. (2011). On the timescale of stimulus-based action–effect learning. *The Quarterly Journal of Experimental Psychology*, *64*(7), 1273–1289. doi:10.1080/17470218.2010.546417

Wood, E. A. (2014). Free choice and free play in early childhood education: troubling the discourse. *International Journal of Early Years Education*, *22*(1), 4–18. doi:10.1080/09669760.2013.830562

Zapparoli, L., Seghezzi, S., Scifo, P., Zerbi, A., Banfi, G., Tettamanti, M., & Paulesu, E. (2018). Dissecting the neurofunctional bases of intentional action. *Proceedings of the National Academy of Sciences*, *115*(28), 7440–7445. doi:10.1073/pnas.1718891115

# Appendix A – Study 1

The following represents the article described in Chapter 5. Springer Nature granted the license to reproduce the final author's accepted manuscript here.

Naefgen, C., & Janczyk, M. (2018b). Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation. *Cognitive Processing*. doi:10.1007/s10339-018-0887-0

The published version can be found under: https://link.springer.com/article/10.1007/s10339-018-0887-0

# Smaller backward crosstalk effects for free choice tasks are not the result of immediate conflict adaptation

Christoph Naefgen & Markus Janczyk

Department of Psychology

Eberhard Karls University of Tübingen, Tübingen, Germany

**Address correspondence to:**
Christoph Naefgen
Eberhard Karls University of Tübingen
Department of Psychology
Schleichstraße 4
72076 Tübingen
Germany
Phone: +49 (0)7071 29 76769
Email: christoph.naefgen@uni-tuebingen.de

## Abstract

In dual-task situations, mutual interference phenomena are often observed. One particularly interesting example of such phenomena is that even Task 1 performance is improved if Task 2 requires a compatible (e.g., both responses are given on the left side) instead of an incompatible response (e.g., one response is given on the right side, and the other on the left side). This is called the compatibility-based backward crosstalk effect (BCE). In a previous paper, we observed support for a critical role of stimulus-response (S-R) links in causing this effect: the BCE was smaller when one of the two tasks was a free choice task. However, an alternative explanation for this observation is that free choice tasks lead to immediate conflict adaptation, thereby reducing the interference from the other task. In the present two experiments, we tested this explanation by varying the amount of conflict assumed to be induced by a free choice task either sequentially (Exp. 1) or block-wise (Exp. 2). While we replicated a sequential modulation of the BCE with two forced choice tasks, we observed (1) no reduction of the BCE induced by (compatible) free choice trials nor (2) an effect of block-wise manipulations of the frequency of free choice trials on the size of the BCE. Thus, while the BCE is sensitive to sequential modulations induced by the (in)compatibility of two forced choice responses, which might point to conflict adaptation, the reduced BCE in dual-task situations involving a free choice task is likely due to its weaker S-R links.

**Keywords:** conflict adaptation ; backward crosstalk effect ; free choice tasks ; dual-task

## 1. Introduction

***1.1 Dual-Tasking and the Backward Crosstalk Effect.*** When humans work on two tasks simultaneously, performance in one or both tasks usually becomes worse. These dual-task costs can be influenced in various ways, depending on the tasks' specific characteristics. In the case that characteristics of Task 2 performance influence even performance in Task 1, this is called a *backward crosstalk effect* (BCE). The example we investigate here is based on spatial compatibility between the responses required in both tasks: If both tasks require spatially compatible responses (e.g., a manual left button press in Task 1 followed by a left pedal button press or a "left" vocal response in Task 2), response times (RTs) in Task 1 are shorter in comparison to trials with spatially incompatible responses (e.g., a manual left button press followed by a right pedal button or a "right" vocal response). This is the compatibility-based BCE, which was first demonstrated by Hommel (1998; see also Ellenbogen & Meiran 2008, 2011; Giammarco, Thomson, & Watter 2014; Hommel & Eglau 2002; Janczyk, Pfister, Hommel, & Kunde 2014; Janczyk, Renas, & Durst 2018; Lien & Proctor 2002; Renas, Durst, & Janczyk 2017; Watter & Logan 2006).

Such observations are difficult to reconcile with the broadly accepted central bottleneck theory of dual-tasking. This theory assumes that task processing comprises three stages: A pre-central perceptual stage, a central response selection stage, and a post-central motor stage. The central response selection stage is conceived as the only stage incapable of parallel processing and interaction with other stages of its kind, hence the term bottleneck (Pashler 1984, 1994). In other words, response selection in Task 2 can only start when response selection in Task 1 has finished and the bottleneck becomes again available. However, the existence of BCEs has challenged this idea. It was argued that some response selection related processes in Task 2 must already be ongoing even during Task 1 response selection. Thus, some authors argued to split the response selection stage into two stages: (1) A first stage of *response activation*, capable of being processed in parallel with other stages and interacting with them (and thus being the stage where the BCE results from) and (2) a bottleneck stage of *(final) response selection* (Hommel 1998; Hommel & Eglau 2002; Lien & Proctor 2002). Recently, other authors have argued, however, that automatic Task 2 response activation directly affects Task 1 response selection (Janczyk et al. 2018; Thomson, Danis, & Watter 2015).

In most studies on the BCE, both component tasks were *forced choice tasks*, which means that for every presented stimulus exactly one response is considered correct. A different type of task is the *free choice task*, in which for one stimulus, two (or more) responses are considered equally valid (Berlyne 1957). Typically, these free choice tasks are accompanied by the instruction to try to respond with both responses about equally often and to avoid obvious patterns in the responses. A typical observation are longer RTs in free than in forced choice tasks. There are multiple explanations for this observation: Some have attributed it to different modes of sensorimotor integration (i.e., intention-based vs. stimulus-based actions; see Herwig, Prinz, & Waszak 2007). Others have ascribed this RT difference to implementation intentions (Gollwitzer 1999) that do not exist for free choice tasks, but only for forced choice tasks (Janczyk, Dambacher, Bieleke, & Gollwitzer 2015a). Implementation intention here means that participants form an “if-then” plan on how to achieve the goal in question. In the case of forced choice tasks, this may, for example, be “If I see a red stimulus, I press the left button”. Such plans are assumed to facilitate early perceptual processing for forced choice stimuli, resulting in the observed RT difference. Naefgen, Dambacher, and Janczyk (2017b) looked at the RT difference from a sequential sampling perspective. In such a framework, information is noisily accumulated at some speed over time until it reaches a threshold, which initiates giving a response. Within this framework, they manipulated the decision thresholds and provided evidence for longer phases in which no information is accumulated in free choice tasks when compared to forced choice tasks, which may be devoted to random generation in the free choice task (Naefgen & Janczyk 2018). Moreover, in line with these latter studies that attribute the RT difference to a process outside response selection, both free and forced choice tasks are similarly affected by dual-task interference (Janczyk, Nolden, & Jolicoeur 2015b).

In a recent study, we compared the size of the BCE between conditions in which one of the two tasks was either a free choice task or a forced choice task (Naefgen, Caissie, & Janczyk 2017a). We assumed that free choice tasks entail weaker stimulus-response (S-R) links than forced choice tasks do: Even if in free choice tasks S-R links are formed, they would be less consistent and therefore weaker than in forced choice tasks. S-R links (or more precisely: automatic S-R translations occurring in Task 2) have been proposed as the mechanism leading to the BCE by various authors (Hommel 1998; Hommel & Eglau 2002; Janczyk et al. 2018; Lien & Proctor 2002). The general observation in the study by Naefgen et al. was a smaller BCE when one of the tasks was a free choice task – a result that would be consistent with the assumption of weaker S-R links in the free choice task.

***1.2 Cognitive Conflict and Control.*** Berlyne (1957) already conceptualized free choice tasks as response-response (R-R) conflict-laden tasks. Essentially, whenever a free choice stimulus is presented, the (two) response options compete with each other. In order to produce a response, some sort of conflict resolution needs to take place. This view suggests an alternative explanation for our earlier observation of a smaller BCE with free choice tasks as Task 1 (see also the General Discussion in Naefgen et al. 2017a): In particular, the smaller BCE may in fact also result from conflict adaptation. In other words, encountering a free choice task may result in (cognitive) conflict which then leads to immediate processes of conflict adaptation which reduce the impact of Task 2 on Task 1 performance.

Botvinick, Braver, Barch, Carter, and Cohen's (2001) conflict-monitoring theory posits that cognitive control is determined by conflict monitoring and arises whenever conflict is detected. In particular, it suggests that conflict arises and leads to increases in cognitive control mechanisms in conflict tasks (e.g., Stroop tasks), but also in underdetermined tasks (e.g., such as the free choice task investigated here; cf. Exp. 2 from Frith, Friston, Liddle, & Frackowiak

1991, who used a similar task and observed that it activates the anterior cingulate cortex, which Botvinick et al. identified as involved in cognitive control),. These mechanism can, for example, be an increased focus on task-relevant features (Botvinick et al. 2001) or a suppression of task-irrelevant information (Janczyk & Leuthold 2018; Stürmer & Leuthold 2003; Stürmer, Leuthold, Soetens, Schröter, & Sommer 2002).

One particularly important effect in support of this theory is the sequential modulation of the congruency effect observed in conflict tasks. For example, in the Eriksen flanker task (Eriksen & Eriksen 1974), a central stimulus is flanked by task-irrelevant stimuli that are either congruent (i.e., they suggest the same response option as the central stimulus) or incongruent (i.e., they suggest the other response option). Responses to congruently flanked stimuli are generally faster than responses to incongruently flanked stimuli (the congruency effect). Importantly, the size of this congruency effect depends on the congruency status of the preceding Trial *n*-1 with larger congruency effects following congruent than following incongruent Trials *n*-1; a sequential modulation sometimes referred to as the Gratton effect (Gratton, Coles, & Donchin 1992). Similar results are also obtained for other conflict tasks (Simon task: Akçay, & Hazeltine 2007; Dignath, Janczyk, & Eder 2017; Stroop: Mayr, & Awh 2009; Notebaert, Gevers, Verbruggen, & Liefooghe 2006), and also occur for the BCE which is only observed following compatible Trials *n*-1 (Janczyk 2016; Renas et al. 2017; Scherbaum, Gottschalk, Dshemuchadse, & Fischer 2015; see also Schuch, Dignath, Steinhauser, & Janczyk 2018). Importantly for the present purposes, it has been shown that adaptation to cognitive conflict can happen even within one trial (Goschke & Dreisbach 2008; Scherbaum, Fischer, Dshemuchadse, & Goschke 2011). Thus, it is in fact possible that R-R conflict occurring upon encountering a free choice task (Berlyne 1957) could have affected the size of the BCE by way of immediate conflict adaptation in our previous experiments (Naefgen et al. 2017a). For an illustration of how different kinds of conflict (could) affect the size of the BCE, see Figure 1.

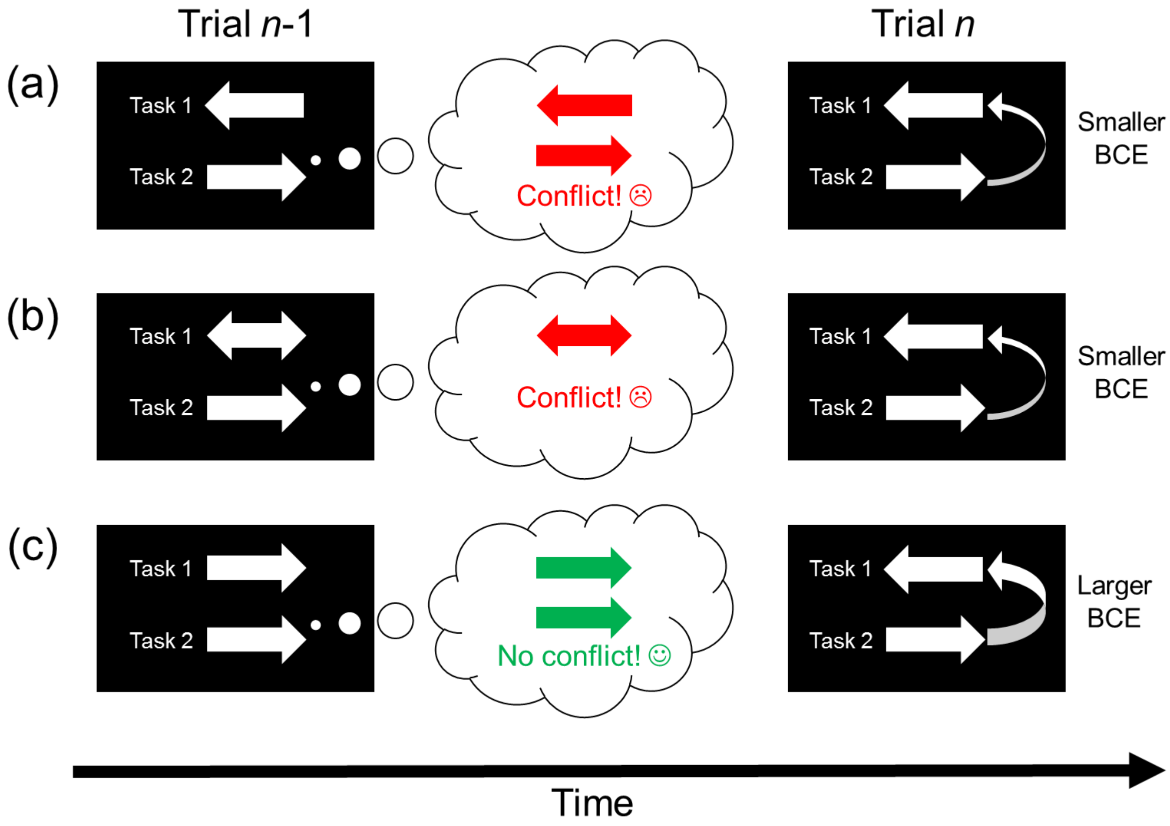

***Figure 1.*** Illustration of possible different kinds of conflicts in the backward crosstalk paradigm. In the first row (a), conflict arises from incompatible Task 1 and Task 2 responses. In the second row (b), conflict arises from the free choice Task 1 (indicated by the double-headed arrow). In the third row (c), no conflict is present. In cases where conflict occurred in Trial *n*-1 a smaller BCE is expected in Trial *n* than when there was no conflict present in Trial *n*-1.

With the present study we aim to address (and rule out) this alternative explanation. To this end, we manipulated the level of conflict in two BCE experiments. In Experiment 1, we manipulated the degree of conflict in a previous dual-task trial; that is, in Trial *n*-1, and focused on the size of the following BCE. In Experiment 2, we manipulated the conflict level block-wise, by varying the proportion of trials in which the first task was a free choice task.

## 2. Experiment 1

Experiment 1 employed a standard BCE paradigm with the simultaneous onset of two stimuli (the color and the identity of a letter, see Figure 2 for an illustration). Task 1 responses were manual left/right key presses, and Task 2 responses were left/right foot pedal presses. Unbeknown to the participants, trials were presented in pairs of a prime and a test trial. We systematically manipulated the type of Task 1 in the prime trials (free vs. forced choice) and, in case of forced choice Task 1s (50 % of the prime trials, 100 % of the test trials), the compatibility relation between both responses. Half of the forced choice prime trials preceding each compatible and incompatible test trials were compatible; the other half was incompatible. This experimental setup produced data that are similar in nature to Experiment 1 from Naefgen et al. (2017a). However, presenting the trials in pairs allowed us to achieve roughly equal numbers of trials in the relevant design cells. (Note that for free choice tasks some variance between participants regarding the proportions of compatible and incompatible trials is to be expected.) The critical analyses focused on the size of the BCE in test trials as a function of the nature of the prime trial. The first prediction concerns trials where Task 1 in the prime trial was a forced choice task. Here, we expect to replicate the observation of Janczyk (2016) that the BCE is smaller or absent following incompatible trials and large following compatible trials. The critical comparison is the one between these latter trials and trials where Task 1 in the prime trial was a free choice task and participants responded in a compatible way. If the free choice task in fact induces cognitive conflict that leads to initiation of adaptation processes, we expect a smaller BCE after compatible free choice prime trials than after compatible forced choice prime trials. If, however, differences in the strength of S-R links are important, the size of the BCE in the test trial is expected similar in test trials following compatible forced choice and compatible free choice prime trials. As there were no predictions concerning trials where the prime trials was a free choice and participants responded in an incompatible way, these trials were not included in the main analyses reported here. However, analyses of the full 2 × 2 × 2 design are provided in the Appendix for completeness.

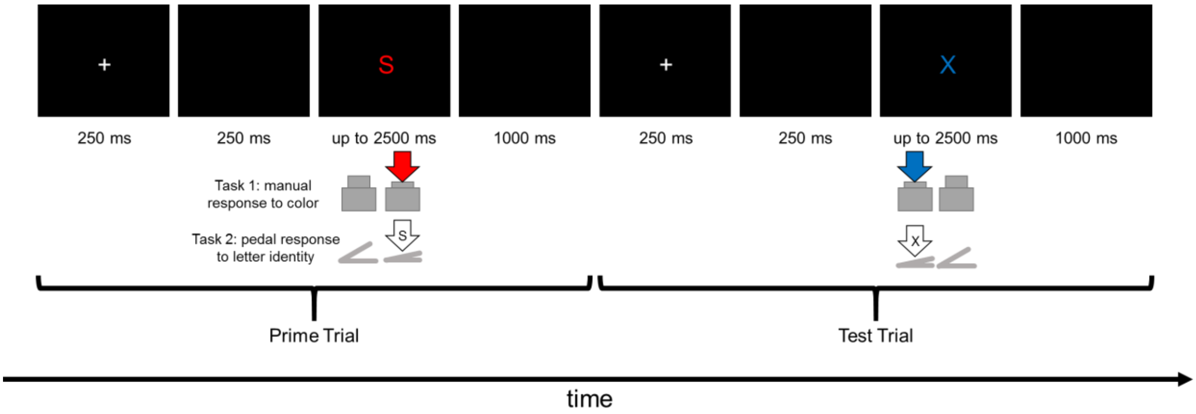

***Figure 2.*** Illustration of a pair of trials in Experiment 1. The red S is the prime trial stimulus and the blue X is the test trial stimulus. In this exemplary stimulus mapping, a red stimulus instructs the participant to give a right manual response (prime trial Task 1) and an S stimulus instructs the participants to give a right pedal response (prime trial Task 2). A blue stimulus on the other hand instructs a left response (test trial Task 1) and an X stimulus instructs a left pedal response (test trial Task 2). Therefore, this example illustrates a compatible forced choice test trial (as both responses are on the left side) following a compatible forced choice prime trial (as both responses are on the right side). Note that in Experiment 2 the general procedure was the same but there was no distinction between prime trials and test trials.

## 2.1 Methods

***2.1.1 Participants.*** Thirty-six people from the Tübingen area participated (Mean age = 22 years, 31 female) for course credit or monetary compensation. All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection.

Data from participants who favored a left manual response in ≤ 15 % or ≥ 85 % of the free choice prime trials or whose Task 1 free choice prime trial response were ≤ 15 % compatible or ≥ 85 % compatible with the Task 2 response were discarded and replaced with data from new participants (*n* = 11).

***2.1.2 Apparatus and stimuli.*** Stimulus presentation and response collection were controlled by a PC connected to a 17-inch CRT monitor. Stimuli were colored letters, that is, the letters ‘X’ and ‘S’ presented in red, green, or blue color. In particular, Task 1 stimuli (S1) were the respective colors, and Task 2 stimuli (S2) were the letter identities. Stimuli were presented against a black background. Manual responses in Task 1 (R1) were collected with two custom-built response keys placed on a table to the left and right of the participants. Foot pedal responses in Task 2 (R2) were given on response keys placed under the left and right foot of the participants in a position that allowed them to sit in a comfortable position.

***2.1.3 Tasks and procedure.*** Task 1 was either to give a predefined R1 in response to two of the possible colors (forced choice task) or to freely choose one of the possible responses in response to the third color (free choice task). Task 2 was to give R2 in response to the letter identity (thus Task 2 was always forced choice). A trial began with the presentation of a small fixation cross (250 ms), followed by a blank screen (250 ms) and the letter stimulus onset. The stimulus remained on screen until both responses were made. A trial was cancelled if no response was given within 2500 ms after stimulus onset. General errors (no response, response too early, wrong response order) and erroneous responses in one or both tasks were fed back (1000 ms), and the next trial started after an inter-trial interval (ITI) of 1000 ms. Trials were (without the participants' knowledge) presented in pairs, with the first trial in each pair being the prime and the second being the test trial. In half of the prime trials, Task 1 was a free choice task, in the other half a forced choice task. In test trials, Task 1 was always forced choice.

Following ten randomly drawn practice trial pairs (not analyzed), nine blocks of 32 trial pairs each were administered; the first two of these blocks were excluded from the analyses as practice. The 32 trial pairs per block represent the combination of the different stimuli that can occur in prime and the test trial, with the free choice stimulus appearing twice as the prime trial's S1: 4 (Prime S1: free choice, free choice, forced choice left, and forced choice right) × 2 (Prime S2: forced choice left and right) × 2 (Test S1: Forced choice left and right) × 2 (Test S2: Forced choice left and right). These trial pairs were presented in a random order.

Participants were tested individually in one single session of about 45 minutes. Written instructions emphasized speed and accuracy and, for the free choice trials, an even distribution of left and right responses as well as avoiding patterns to maintain this distribution. Participants were also instructed to always give first R1 and then R2. The mappings of stimuli to tasks/responses were counterbalanced across participants.

***2.1.4 Design and analyses.*** Only test trials following entirely correct prime trials were considered for analyses. A trial was considered compatible when both R1 and R2 were given on the same side; otherwise, a trial was incompatible.

Test trials with general errors were excluded first (wrong response, no response, response too early, wrong response order). Further, to control for possible response grouping (e.g., Miller & Ulrich 2008; Ulrich & Miller 2008), only trials were analyzed where both responses were separated by an inter-response interval (IRI) of at least 50 ms (excluding 1.2 % of trials; using IRIs of 100 ms and 150 ms changed none of the significance patterns). For RT analyses, we considered only test trials in which both R1 and R2 were correct, and trials were further excluded as outliers if RTs deviated more than 2.5 *SD*s from the respective cell mean (calculated separately for each participant).

The two independent variables of interest were: (1) R1-R2 compatibility in the test trial (compatible vs. incompatible) and (2) the conflict level in the prime trial (forced choice incompatible vs. forced choice compatible vs. free choice compatible). RT and error data were analyzed with two orthogonal Helmert contrasts on the variable conflict level and its interaction with the variable compatibility in the test trial. For the latter we expected a main effect. Contrast 1 coded incompatible forced choice primes against the other two levels and we expected an interaction of this contrast with test trial compatibility (revealing the sequential modulation observed, e.g., in Janczyk 2016). Contrast 2 then coded compatible forced choice primes against compatible free choice primes. If the free choice prime induced some sort of conflict adaption, this should yield a decreased BCE in the test trial, and thus an interaction of this contrast with the test compatibility. Both RTs and percentages of errors (PEs) in Task 1 were analyzed with this approach. Task 2 results are provided in the Appendix.

Lastly, analyzing the proportion of compatible (Task 1) response choices in prime trials involving a free choice task, gave the opportunity to replicate the observations in Naefgen et al. (2017a, Experiments 1 and 2) that the choice in a free choice task is influenced by the response required in a subsequent forced choice task. In particular, participants' choices were biased towards choosing a compatible response.

### 2.3. Results and discussion

In the free choice tasks, participants chose the left key on average 43.8 % of the time (Range 18.0-80.7 %), which is significantly different from 50 %, *t*(35) = -2.25, *p* = .031, *d* = -0.53.

Mean correct RTs in Task 1 (2.14 % excluded as outliers) are visualized in Figure 3 and are summarized in Table 1. Responses were faster in compatible trials than in incompatible trials, *t*(35) = 6.47, *p* < .001, showing an overall BCE. Both contrasts were significant, Contrast 1: *t*(35) = 5.50, *p* < .001; Contrast 2: *t*(35) = 7.60, *p* < .001. Most importantly, Contrast 1 interacted with compatibility in the test trial, *t*(35) = 10.66, *p* < .001, whereas Contrast 2 did not, *t*(35) = 0.99, *p* = .328.

Paired *t*-tests indicated significant BCEs for trials preceded by compatible free choice trials (129 ms), *t*(35) = 7.24, *p* < .001, *d* = 1.71, as well as preceded by compatible forced choice trials (146 ms), *t*(35) = 10.59, *p* < .001, *d* = 2.50. When preceded by incompatible forced choice trials, the BCE was reversed (-59 ms), *t*(35) = -3.96, *p* < .001, *d* = -0.93.

Mean PEs are summarized in Table 1. The compatibility in the test trial, *t*(35) = 5.28, *p* < .001, had a significant influence on the PEs with – overall – fewer errors in compatible compared with incompatible trials. As in the RT analyses, Contrast 1, *t*(35) = 4.66, *p* < .001, Contrast 2, *t*(35) = 4.28, *p* < .001, and the interaction of Contrast 1 with compatibility in the test trial, *t*(35) = 5.47, *p* < .001, were significant. The interaction of Contrast 2 and compatibility was not significant, *t*(35) = 0.03, *p* = .974. Paired *t*-tests indicated significant differences in PEs between compatible and incompatible test trials when preceded by compatible free choice primes, *t*(35) = 5.57, *p* < .001, *d* = 1.31, and compatible forced choices primes, *t*(35) = 6.13, *p* < .001, *d* = 1.45, but not for trials preceded by incompatible forced choice primes, *t*(35) = -0.97, *p* = .339, *d* = -0.23.

The last analysis focused on prime trials involving a free choice Task 1 (2.86 % outliers). In these trials, participants chose the same response location as required in Task 2, thus a compatible choice, in 58.9 % of trials. This value is significantly different from 50 %, *t*(35) = 4.72, *p* < .001, *d* = 1.11.

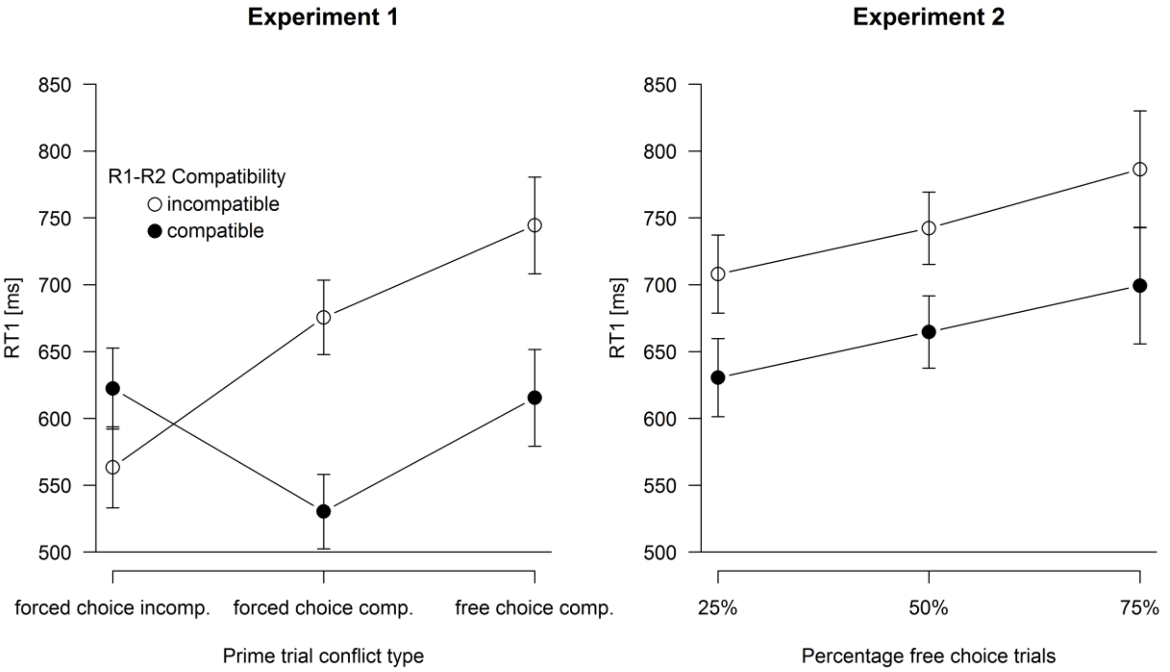

***Figure 3.*** Mean correct response times from Task 1 (RT1) of Experiments 1 and 2 as a function of prime trial conflict type (Experiment 1) and percentage of free choice trials in a block (Experiment 2) and R1-R2 compatibility. Error bars are 95 % within-subject confidence intervals calculated separately for each prime trial conflict type in Experiment 1 and separately for each percentage level of free choice trials in a block in Experiment 2 (see Pfister & Janczyk 2013).

***Table 1.*** Mean correct response times (RT1) in milliseconds and percentages of errors (PE1) from Task 1 of Experiment 1 as a function of prime trial conflict type and R1-R2 compatibility in the test trial. The BCE (backward crosstalk effect) row reports the difference between the mean of the compatible and incompatible trials.

| | Prime trial conflict Type | | | | | |
|---|---|---|---|---|---|---|
| R1-R2 compatibility | Forced choice incompatible | | Forced choice compatible | | Free choice compatible | |
| | RT1 | PE1 | RT1 | PE1 | RT1 | PE1 |
| incompatible | 563 | 3.7 | 676 | 11.3 | 744 | 15.2 |
| compatible | 622 | 5.4 | 530 | 1.2 | 615 | 5.2 |
| BCE | -59 | -1.7 | 146 | 10.1 | 129 | 10.0 |

In sum, this experiment yields two main results. First, we replicated the smaller (or even inversed) BCE following incompatible trials (Janczyk 2016), thus a sequential modulation revealing conflict adaptations. Second, however, the BCE in the test trial was not smaller following a compatible free choice than following a compatible forced choice prime trial. Such a reduction would have been predicted if the smaller BCEs with free choice tasks (Naefgen et al. 2017a) were due to (immediate) conflict adaptations triggered by the free choice task. Thus, from Experiment 1 we tentatively conclude that the diminished BCE observed with free choice tasks in Naefgen et al. did not arise from immediate conflict adaptation processes triggered upon encountering a free choice task. Experiment 2 further investigates this with a different approach.

The choice results for the free choices in prime trials replicate the observations reported in Naefgen et al. (2017a) that the choice in a free choice task is biased by a subsequent Task 2 forced choice response toward a compatible response. This lends additional credibility to the idea discussed there that free choice task choices are biased both by preceding primes (see also Kiesel, Wagener, Kunde, Hoffmann, Fallgatter, & Stöcker 2006, and Mattler & Palmer 2012) but also by subsequent 'primes' such as the forced choice Task 2 in the present study.

## 3. Experiment 2

In addition to the congruency status of the immediately preceding trial, the proportion of congruent trials modulates the size of congruency effects which become larger with an increasing proportion of congruent trials in a block. This observation is called the list-wide proportion congruency (LWPC) effect (Gratton el al. 1992; for a review, see Bugg & Crump 2012). A variant of this effect is the context-specific proportion congruency (CSPC) effect (Crump, Gong, & Milliken 2006), where the proportion of congruent trials is manipulated as a function of context (e.g., location), while the overall proportion of congruent and incongruent trials is 50 % each. Fischer, Gottschalk, and Dreisbach (2014) reported that the BCE indeed is sensitive to CSPC manipulations (see also Fischer & Dreisbach 2015, who used Task 1 stimuli that conveyed information about the stimulus onset asynchrony and reported a reduced BCE when Task 1 predicted a short SOA). Assuming that the reduced impact of incompatible information under conditions with high proportions of incongruent information is due to an adaptation of how much ‘irrelevant’ information (as this is in part determined by how irrelevant the information actually is) is used (see Botvinick et al. 2001; Schmidt 2013), a similar adaptation to varying proportions of free choice trials should be observed if free choice tasks also induce (R-R) conflict.

Experiment 2 therefore employed the same BCE paradigm as Experiment 1 with the following differences: Trials were no longer presented in pairs. The critical variables manipulated were the proportion of free choice trials in a block (75 % vs. 50 % vs. or 25 %) and the compatibility of R1 and R2. Over the course of a block, participants should adapt to the proportion of free choice trials they are confronted with. In particular, a higher percentage of free choice trials should lead to higher perceived conflict, which in turn should strengthen conflict adaptation. In other words, if the reduced BCE in Naefgen et al. (2017a) was indeed due to R-R conflict-induced conflict adaptation, higher percentages of free choice trials in a block should lead to adaptation to this conflict, and thus to smaller BCEs in the trials where both Task 1 and 2 were forced choice of the same block.

## 3.1 Methods

***3.1.1 Participants.*** Thirty-six people from the Tübingen area participated (Mean age = 23 years, 27 female) for course credit or monetary compensation. All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection.

Data from participants who favored a left manual response in ≤ 15 % or ≥ 85 % of the free choice trials were discarded and replaced with data from new participants (*n* = 5).

***3.1.2 Apparatus and stimuli.*** The apparatus and stimuli were the same as in Experiment 1.

***2.1.3 Tasks and procedure.*** The general procedure was the same as in Experiment 1 except that trials were not presented in pairs. Again, only Task 1 could be free choice. After a practice block of ten trials (excluded from the analysis), three sets of six blocks of 32 trials were presented. Each set of six blocks comprised one of the levels of the free choice frequency manipulation in Task 1 (25 % [or 8 per block] vs. 50 % [or 16 per block] vs. 75 % [or 24 per block]). The order of the three sets of blocks was counterbalanced across participants and trials within each block appeared in a randomized order. The 32 trials in each block result from the combination of even numbers of the types of forced choice trials and the respective percentage of free choice trials.

Participants were tested individually in one single session of about 45 minutes. Written instructions emphasized speed and accuracy and, for the free choice trials, an even distribution of left and right responses as well as avoiding patterns to maintain this distribution. Participants were also instructed to always give first R1 and then R2. The mappings of stimuli to tasks/responses were counterbalanced across participants.

***2.1.4 Design and analyses.*** Only trials where both Task 1 and Task 2 were forced choice tasks were considered for the main analyses.

Two independent variables were varied within participants: (1) R1-R2 compatibility (compatible vs. incompatible; for forced choice Task 1 trials this could be manipulated by the experimenters) and (2) the amount of free choice Task 1 trials in the block (25 % vs. 50 % vs. 75 %). Accordingly, RTs and PEs from Task 1 were mainly analyzed in terms of a 2 × 3 ANOVA. Trials with general errors were excluded first (wrong response, no response, response too early, wrong response order). Again, only trials were analyzed where both responses were separated by an IRI of at least 50 ms (excluding 1.4 % of trials; using IRIs of 100 ms and 150 ms changed none of the significance patterns). For RT analyses, only trials in which both R1 and R2 were correct were considered, and trials were excluded as outliers if RTs deviated more than 2.5 *SD*s from the respective cell mean (calculated separately for each participant). When the assumption of sphericity was violated, Greenhouse-Geisser corrections were applied and the respective $\varepsilon$ is reported. Results for Task 2 are reported in the Appendix.

In addition, the 50 % free choice blocks offered an opportunity to replicate the results from Naefgen et al. (2017a, Exp. 1 and 2), where we observed a smaller BCE when T1 was a free choice task. Accordingly, a 2 × 2 ANOVA with compatibility and task type as repeated measures was performed on RT1 for data from these blocks. As there cannot be errors in a free choice task, PEs in Task 1 were analyzed with a paired *t*-test for Task 1 forced choice trials.

As in Experiment 1, we again took the opportunity to replicate the observations reported in Naefgen et al. (2017a, Experiments 1 and 2) that the choice in a free choice task is influenced by the required response in a subsequent forced choice task.

### 2.3. Results and discussion

In the free choice tasks, participants chose the left key on average 45.6 % of the time (Range 20.2-80.7 %), which is significantly different from 50 %, $t(35) = -2.12$, $p = .041$, $d = -0.50$.

Mean correct RT1s (2.43 % excluded as outliers) are visualized in Figure 3 (right panel) and are summarized in Table 2. Responses were faster in compatible trials than in incompatible trials, $F(1,35) = 38.94$, $p < .001$, $\eta_p^2 = .53$, showing an overall BCE. There also was a significant effect of the amount of free choice tasks in a block, $F(2,70) = 7.92$, $p = .001$, $\eta_p^2 = .18$ with more free choice tasks in a block leading to slower responses. Most importantly, there was no significant interaction between compatibility and the amount of free choices, $F(2,70) = 0.17$, $p = .802$, $\eta_p^2 < .01$, $\varepsilon = .83$.

Mean PEs are summarized in Table 2. Compatibility had a significant effect on PEs with fewer errors in compatible trials, $F(1,35) = 23.03$, $p < .001$, $\eta_p^2 = .40$. The main effect of the amount of free choices, $F(2,70) = 1.14$, $p = .319$, $\eta_p^2 = .03$, $\varepsilon = .84$, as well as the interaction, $F(2,70) = 0.14$, $p = .830$, $\eta_p^2 < .01$, $\varepsilon = .84$, were not significant.

***Table 2.*** Mean correct response times (RT1) in milliseconds and percentages of errors (PE1) from Task 1 of Experiment 2 as a function of block type and R1-R2 compatibility. The BCE (backward crosstalk effect) row reports the difference between the mean of the compatible and incompatible trials.

| | Block Type | | | | | |
|---|---|---|---|---|---|---|
| R1-R2 compatibility | 25 % Free Choices | | 50 % Free Choices | | 75 % Free Choices | |
| | RT1 | PE1 | RT1 | PE1 | RT1 | PE1 |
| incompatible | 708 | 11.3 | 742 | 12.4 | 786 | 13.2 |
| compatible | 630 | 4.3 | 664 | 4.7 | 699 | 5.3 |
| BCE | 78 | 7.0 | 78 | 7.7 | 87 | 7.9 |

For the additional analysis on free choice trials (as done by Naefgen et al. 2017a), mean correct RT1s (2.50 % excluded as outliers) are summarized in Table 3. Responses were faster in compatible trials than in incompatible trials, $F(1,35) = 20.48$, $p < .001$, $\eta_p^2 = .37$, showing an overall BCE. RT1s in trials with free choices were shorter than in trials with forced choices, $F(1,35) = 38.54$, $p < .001$, $\eta_p^2 = .52$. In addition, there was a significant interaction between these two factors, $F(1,35) = 11.85$, $p = .002$, $\eta_p^2 = .25$, with a smaller BCE in trials with free choices. Paired *t*-tests indicated that there was only a significant BCE for trials with a forced choice Task 1, $t(35) = 5.83$, $p < .001$, $d = 1.37$, but not for trials with a free choice Task 1, $t(35) = 1.21$, $p = .235$, $d = 0.28$. Fewer errors were made in compatible than in incompatible trials (see Table 3), $t(35) = 4.07$, $p < .001$, $d = 0.96$.

When analyzing data from the trials involving a free choice Task 1 (2.50 % outliers), participants chose the same response location as in Task 2, thus a compatible choice, in 59.9 % of the Task 1 free choice trials. This value is significantly different from 50 %, $t(35) = 5.23$, $p < .001$, $d = 1.23$. Thus, we could again replicate the respective observations reported in Naefgen et al. (2017a, Experiments 1 and 2).

***Table 3.*** Mean correct response times (RT1) in milliseconds and percentages of errors (PE1) in % from Task 1 of Experiment 2 (50 % free choices block) as a function of task type and R1-R2 compatibility. The BCE (backward crosstalk effect) row reports the difference between the mean of the compatible and incompatible trials.

| | Trial Type | | |
|---|---|---|---|
| R1-R2 compatibility | Free Choice | Forced Choice | |
| | RT1 | RT1 | PE1 |
| incompatible | 624 | 742 | 12.4 |
| compatible | 607 | 665 | 4.7 |
| BCE | 17 | 77 | 7.7 |

In summary, the non-significant interaction in the main analysis suggests that the BCE was of the same size irrespective of the amount of free choice trials in a block. In other words, a decreasing size of the BCE with increasing proportions of free choice trials (as predicted from a conflict adaptation account) was not observed. This result further supports the conclusion from Experiment 1. Furthermore, we replicated results reported by Naefgen et al. (2017a, Exp. 1 and 2) with regards to the reduced BCE when Task 1 is a free choice trial as well as the influence of the forced choice Task 2 on the actual choice that is made in the free choice task. Interestingly, free choice RTs were also shorter than forced choice RTs. This has also been observed in our previous study as well as in other dual-task studies (e.g., Wirth, Janczyk, & Kunde 2018). Note, however, that the opposite was true in the dual-task study by Janczyk, Nolden et al. (2015).

## 4. General Discussion

The present study aimed at testing an alternative explanation for our recent observation that the compatibility-based BCE in dual-tasking is smaller when Task 1 is a free choice task (Naefgen et al. 2017a). We attributed this result to weaker S-R links in free compared with forced choice tasks, but suspected that it may alternatively result from immediate conflict adaptation when participants encountered a free choice Task 1. In two experiments, we examined the size of the BCE depending on the conflict level of the preceding trial (Experiment

1) or the amount of free choice Task 1 trials in a block (Experiment 2). If the alternative explanation were true, we expected smaller BCEs in cases with larger potential conflict.

### 4.1. Summary of results

In both experiments, we first replicated the standard R1-R2 compatibility-based BCE (e.g., Hommel 1998; Janczyk et al. 2018). Second, in Experiment 1 we also replicated the results reported in Janczyk (2016) that the BCE is large in trials following compatible trials and small/absent or even reversed following an incompatible forced choice trial. Third, we were also able to replicate the observation reported in Naefgen et al. (2017a) that the BCE was smaller in free choice Task 1 RTs, but that the actual choices were biased into a compatible direction by the forced choice Task 2 (Experiments 1 and 2). Most importantly, however, we observed no difference in the size of the BCE for trials following compatible free choice and compatible forced choice trials (Experiment 1), and the proportion of free choice trials in a block did not affect the size of the BCE.

### 4.2. Limitations and theoretical implications

Overall, our results replicated critical aspects of the Naefgen et al. (2017a) study and they offer support for our original conclusion: the smaller BCE in dual-task trials with one of the tasks being a free choice task is likely due to weaker S-R links in this kind of task, rather than by an immediate conflict adaptation upon encountering a free choice trial.

One might object that we did not test for rapid within-trial conflict adaptation in Experiment 1 as described by Scherbaum et al. (2011). Perhaps, conflict adaptation occurs very rapidly and all consequences vanish immediately after the trial. This, however, is implausible as an explanation for the lack of conflict adaptation. While rapid conflict adaptation effects were reported for BCE tasks from mouse tracking experiments, nonetheless a sequential modulation of the BCE in the subsequent trial occurred (Scherbaum et al. 2015). A potential objection to the reasoning behind Experiment 2 is that thus far LWPC effects have not been reported in the context of the BCE. However, Fischer et al. (2014) reported a CSPC modulation of the BCE, arguably even stronger evidence for its susceptibility to LWPC-like manipulations.

The fact that we did not observe any hint of conflict adaptation induced by R-R conflict inherent in free choice tasks has theoretical implications. It is of course possible that free choice tasks do not create R-R conflict as originally assumed by Berlyne (1957). In this case, of course, no conflict adaptation (e.g., in the form of sequential modulations) should occur.

Alternatively, it is possible that free choice tasks elicit R-R conflict and also conflict adaptation but that this conflict adaptation does not generalize to other tasks. This is plausible, because for standard conflict tasks a generalization of conflict adaptation from one task to another (e.g., from a flanker to a Stroop task) does not always occur (see Braem, Abrahamse, Duthoo, & Notebaert 2014, for a review). Further, one may conceive BCE trials with a free choice Task 1 as ones instantiating a different context than those with a forced choice Task 1. If this were true, a sequence with a prime trial that entailed a free choice Task 1 would mean a change of context to the test trial. Indeed, there is some evidence that sequential modulations (within dual-task settings) seem to depend on repetitions of task contexts (Fischer, Plessow, Kunde, & Kiesel 2010).

### 4.3. Conclusion

We investigated an alternative explanation to reduced S-R links for diminished BCEs in dual-task trials involving a free choice trial (Naefgen et al. 2017a). However, we observed no evidence supporting the idea that conflict adaptation induced by free choice tasks led to these smaller BCEs.

## 5. Appendix

This Appendix reports the analyses of Task 2 performance in Experiments 1 and 2 (Sections 5.1 and 5.2) and the full analyses of the 2 × 2 × 2 design employed in Experiment 1 (Section 5.3 for Task 1 performance and Section 5.4 for Task 2 performance).

***5.1 Experiment 1: Task 2 results.*** Mean correct RT2s (2.61 % excluded as outliers) are summarized in Table A1. Responses were faster in compatible trials than in incompatible trials, $t(35) = 6.05$, $p < .001$, showing an overall forward crosstalk effect (FCE). Contrast 1 was significant, $t(35) = 4.30$, $p < .001$, as was Contrast 2, $t(35) = 6.07$, $p < .001$. Contrast 1 interacted with compatibility, $t(35) = 11.61$, $p < .001$ and so did Contrast 2, $t(35) = 2.13$, $p = .040$. The latter indicating a reduced FCE following compatible free (vs. compatible forced) choice Task 1 trials.

Paired *t*-tests indicated significant FCEs for trials following compatible free choice prime trials, $t(35) = 7.14$, $p < .001$, $d = 1.68$, as well as compatible forced choice prime trials, $t(35) = 10.38$, $p < .001$, $d = 2.45$. Following incompatible forced choice prime trials, the FCE was reversed, $t(35) = -3.72$, $p = .001$, $d = -0.88$.

Mean PE2s are summarized in Table A1. The compatibility in the test trial, $t(35) = 3.33$, $p = .002$, had a significant influence on the PEs. Furthermore, Contrast 1 was significant, $t(35) = 3.84$, $p < .001$, while Contrast 2 was not, $t(35) = 0.73$, $p = .469$. Contrast 1 interacted with compatibility, $t(35) = 6.30$, $p < .001$, while Contrast 2 did not $t(35) = 0.64$, $p = .523$. For the differences in PEs between compatible and incompatible trials, paired *t*-tests indicated significant differences for trials following compatible free, $t(35) = 4.13$, $p < .001$, $d = 0.97$, and forced choice prime trials, $t(35) = 5.41$, $p < .001$, $d = 1.27$, as well as, in the other direction, those following incompatible forced choice prime trials, $t(35) = -3.35$, $p = .002$, $d = -0.79$.

***Table A1.*** Mean correct response times (RT2) in milliseconds and percentages of errors (PE2) from Task 2 of Experiment 1 as a function of prime trial conflict type and R1-R2 compatibility. The FCE (forward crosstalk effect) row reports the difference between the mean of the compatible and incompatible trials.

| | Prime trial conflict Type | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| R1-R2 compatibility | Forced choice incompatible | | Forced choice compatible | | Free choice compatible | | Free choice incompatible | |
| | RT2 | PE2 | RT2 | PE2 | RT2 | PE2 | RT2 | PE2 |
| incompatible | 919 | 2.4 | 1065 | 13.1 | 1124 | 11.7 | 1098 | 8.6 |
| compatible | 994 | 8.4 | 864 | 3.6 | 967 | 3.8 | 1041 | 5.0 |
| FCE | -75 | -6.0 | 201 | 9.5 | 157 | 7.9 | 57 | 3.6 |

***5.2 Experiment 2: Task 2 results.*** Mean correct Task 2 RTs (2.18 % excluded as outliers) are summarized in Table A2. There was a significant FCE, $F(1,35) = 38.59$, $p < .001$, $\eta_p^2 = .97$, as well as a significant effect of the block type, $F(2,70) = 7.94$, $p = .001$, $\eta_p^2 = .18$, but no significant interaction, $F(2,70) = 0.46$, $p = .630$, $\eta_p^2 = .01$. Paired $t$-tests indicated significant FCEs for all block types, 25 % free choices, $t(35) = 5.99$, $p < .001$, $d = 1.41$; 50 % free choices, $t(35) = 5.34$, $p < .001$, $d = 1.26$; and 75 % free choices, $t(35) = 4.20$, $p < .001$, $d = 0.99$.

Mean PE2s are summarized in Table A2. The compatibility in the test trial had a significant effect on PE2s with fewer errors in compatible trials, $F(1,35) = 10.53$, $p = .003$, $\eta_p^2 = .23$. Neither the block type, $F(2,70) = 0.01$, $p = .985$, $\eta_p^2 < .01$, nor its interaction with compatibility, $F(2,70) = 0.10$, $p = .905$, $\eta_p^2 < .01$, were significant.

***Table A2.*** Mean correct response times (RT2) in milliseconds and percent error (PE2) from Task 2 of Experiment 2 as a function of block type and R1-R2 compatibility. The FCE (forward crosstalk effect) row reports the difference between the mean of the compatible and incompatible trials.

| | Block Type | | | | | |
|---|---|---|---|---|---|---|
| R1-R2 compatibility | 25 % Free Choices | | 50 % Free Choices | | 75 % Free Choices | |
| | RT2 | PE2 | RT2 | PE2 | RT2 | PE2 |
| incompatible | 1114 | 8.2 | 1163 | 8.6 | 1214 | 8.6 |
| compatible | 1009 | 4.8 | 1072 | 4.7 | 1102 | 4.6 |
| BCE | 105 | 3.4 | 91 | 3.9 | 112 | 4.0 |

***5.3 Experiment 1, Full Design: Task 1 results.*** This section describes the Task 1 RTs (2.09 % excluded as outliers) and PEs of the full 2 × 2 × 2 (compatibility in the test trial ×

compatibility in the prime trial × task type in the prime trial) design of Experiment 1. Only test trials were used in this analysis. Mean values are summarized in Table A3 There was a significant main effect of compatibility in the test trial, $F(1,35) = 31.77$, $p < .001$, $\eta_p^2 = .48$, and of task type in the prime trial, $F(1,35) = 65.61$, $p < .001$, $\eta_p^2 = .65$. There was no main effect of compatibility in the prime trial, $F(1,35) = 0.03$, $p = .874$, $\eta_p^2 < .01$. There were significant interactions between compatibility in the test and in the prime trial, $F(1,35) = 92.87$, $p < .001$, $\eta_p^2 = .73$, the compatibility in the test trial and the task type in the prime trial, $F(1,35) = 4.76$, $p = .036$, $\eta_p^2 = .12$, and between all three factors, $F(1,35) = 17.75$, $p < .001$, $\eta_p^2 = .34$. There was no significant interaction between task type and compatibility in the prime trial, $F(1,35) = 2.81$, $p = .103$, $\eta_p^2 = .07$.

For the PE1s, there were main effects for the compatibility in the test trial, $F(1,35) = 12.44$, $p = .001$, $\eta_p^2 = .26$, the compatibility in the prime trial, $F(1,35) = 5.01$, $p = .032$, $\eta_p^2 = .13$, as well as task type in the prime trial, $F(1,35) = 25.69$, $p < .001$, $\eta_p^2 = .42$. There was an interaction between compatibility of the test and the prime trial, $F(1,35) = 28.71$, $p < .001$, $\eta_p^2 = .45$. There was no interaction between the compatibility of the test trial and task type in the prime trial, $F(1,35) = 0.04$, $p = .852$, $\eta_p^2 < .01$, task type in the prime trial and compatiblility in the prime trial, $F(1,35) = 0.48$, $p = .493$, $\eta_p^2 = .01$, nor between all three factors, $F(1,35) = 0.06$, $p = .815$, $\eta_p^2 < .01$.

***Table A3.*** Mean correct response times (RT1) in milliseconds and percentages of errors (PE1) from Task 1 of Experiment 1 as a function of prime trial R1-R2 compatibility, prime trial task type, and test trial R1-R2 compatibility. The BCE row reports the difference between the mean of the compatible and incompatible trials.

| | Prime trial conflict Type | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| R1-R2 compatibility | Forced choice incompatible | | Forced choice compatible | | Free choice compatible | | Free choice incompatible | |
| | RT1 | PE1 | RT1 | PE1 | RT1 | PE1 | RT1 | PE1 |
| incompatible | 563 | 3.7 | 676 | 11.3 | 744 | 15.2 | 704 | 7.3 |
| compatible | 622 | 5.4 | 530 | 1.2 | 615 | 5.2 | 680 | 8.5 |
| BCE | -59 | -1.7 | 146 | 10.1 | 129 | 10.0 | 24 | -1.2 |

***5.4 Experiment 1, Full Design: Task 2 results.*** This section describes the Task 2 RTs (2.45 % excluded as outliers) and PEs of the full 2 × 2 × 2 (compatibility in the test trial × compatibility in the prime trial × task type in the prime trial) design of Experiment 1. Only test trials were used in this analysis. Mean values are summarized in Table A1. There was a significant main effect of compatibility in the test trial, $F(1,35) = 32.48$, $p < .001$, $\eta_p^2 = .48$, and task type in the prime trial, $F(1,35) = 57.53$, $p < .001$, $\eta_p^2 = .62$. There was no main effect of compatibility in the prime trial, $F(1,35) = 0.91$, $p = .346$, $\eta_p^2 = .03$. There were significant interactions between compatibility in the test and the prime trial, $F(1,35) = 103.50$, $p < .001$, $\eta_p^2 = .75$, the compatibility in the test trial and task type in the prime trial, $F(1,35) = 5.97$, $p = .020$, $\eta_p^2 = .15$, and all three factors, $F(1,35) = 29.99$, $p < .001$, $\eta_p^2 = .46$. There was no significant interaction between compatibility and task type in the prime trial, $F(1,35) = 3.44$, $p = .072$, $\eta_p^2 = .09$.

For the PE2s, there were main effects for the compatibility in the test trial, $F(1,35) = 12.33$, $p = .001$, $\eta_p^2 = .26$, and the compatibility in the prime trial, $F(1,35) = 10.28$, $p = .003$, $\eta_p^2 = .23$. There were significant interactions between compatibility in the test and in the prime trial, $F(1,35) = 33.38$, $p < .001$, $\eta_p^2 = .49$, the compatibility in the test trial and task type in the prime trial, $F(1,35) = 10.43$, $p = .003$, $\eta_p^2 = .23$, the compatibility and task type in the prime trial, $F(1,35) = 6.08$, $p = .019$, $\eta_p^2 = .15$, as well as all three factors, $F(1,35) = 11.45$, $p = .002$, $\eta_p^2 = .25$. There was no main effect for task type in the prime trial, $F(1,35) = 0.37$, $p = .548$, $\eta_p^2 = .01$.

## Compliance with Ethical Standards

*Funding*: This research was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), grant JA 2307/1-2 awarded to Markus Janczyk. Work of MJ is further supported by the Institutional Strategy of the University of Tübingen (DFG ZUK 63).

*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 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## 6. References

Akçay Ç, Hazeltine E (2007) Conflict monitoring and feature overlap: Two sources of sequential modulations. Psychon Bull Rev 14:742–748. doi: 10.3758/BF03196831

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

Botvinick MM, Braver TS, Barch DM, et al. (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652 doi: 10.1037/0033-295X.108.3.624

Braem S, Abrahamse EL, Duthoo W, Notebaert W (2014) What determines the specificity of conflict adaptation? A review, critical analysis, and proposed synthesis. Front Psychol 5:1134. doi: 10.3389/fpsyg.2014.01134

Bugg JM, Crump MJC (2012) In Support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. Front Psychol 3:367. doi: 10.3389/fpsyg.2012.00367

Crump MJC, Gong Z, Milliken B (2006) The context-specific proportion congruent Stroop effect: Location as a contextual cue. Psychon Bull Rev 13:316–321. doi: 10.3758/BF03193850

Dignath D, Janczyk M, Eder AB (2017) Phasic valence and arousal do not influence post-conflict adjustments in the Simon task. Acta Psychol 174:31–39. doi: 10.1016/j.actpsy.2017.01.004

Ellenbogen R, Meiran N (2008) Working memory involvement in dual-task performance: Evidence from the backward compatibility effect. Mem Cogn 36:968–978. doi: 10.3758/MC.36.5.968

Ellenbogen R, Meiran N (2011) Objects and events as determinants of parallel processing in dual tasks: Evidence from the backward compatibility effect. J Exp Psychol Hum Percept Perform 37:152–167. doi: 10.1037/a0019958

Eriksen BA, Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 16:143–149. doi: 10.3758/BF03203267

Fischer R, Dreisbach G (2015) Predicting high levels of multitasking reduces between-tasks interactions. J Exp Psychol Hum Percept Perform 41:1482–1487. doi: 10.1037/xhp0000157

Fischer R, Gottschalk C, Dreisbach G (2014) Context-sensitive adjustment of cognitive control in dual-task performance. J Exp Psychol Learn Mem Cogn 40:399–416. doi: 10.1037/a0034310

Fischer R, Plessow F, Kunde W, Kiesel A (2010) Trial-to-trial modulations of the Simon effect in conditions of attentional limitations: Evidence from dual tasks. J Exp Psychol Hum Percept Perform 36:1576–1594. doi: 10.1037/a0019326

Frith CD, Friston K, Liddle PF, Frackowiak RS (1991) Willed action and the prefrontal cortex in man: a study with PET. Proc Biol Sci 244:241–246. doi: 10.1098/rspb.1991.0077

Giammarco M, Thomson SJ, Watter S (2016) Dual-task backward compatibility effects are episodically mediated. Atten Percept Psychophys 78:520–541. doi: 10.3758/s13414-015-0998-y

Gollwitzer PM (1999) Implementation intentions: Strong effects of simple plans. Am Psychol 54:493–503. doi: 10.1037/0003-066X.54.7.493

Goschke T, Dreisbach G (2008) Conflict-triggered goal shielding: Response conflicts attenuate background monitoring for prospective memory cues. Psychol Sci 19:25–32. doi: 10.1111/j.1467-9280.2008.02042.x

Gratton G, Coles MGH, Donchin E (1992) Optimizing the use of information: Strategic control of activation of responses. J Exp Psychol Gen 121:480–506. doi: 10.1037/0096-3445.121.4.480

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. doi: 10.1080/17470210601119134

Hommel B (1998) Automatic stimulus–response translation in dual-task performance. J Exp Psychol Hum Percept Perform 24:1368–1384. doi: 10.1037/0096-1523.24.5.1368

Hommel B, Eglau B (2002) Control of stimulus-response translation in dual-task performance. Psychol Res 66:260–273. doi: 10.1007/s00426-002-0100-y

Janczyk M (2016) Sequential modulation of backward crosstalk and task-shielding in dual-tasking. J Exp Psychol Hum Percept Perform 42:631–647. doi: 10.1037/xhp0000170

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

Janczyk M, Leuthold H (2018) Effector system-specific sequential modulations of congruency effects. Psychon Bull Rev 25 1066-1072. doi: 10.3758/s13423-017-1311-y

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

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

Janczyk M, Renas S, Durst M (2018) Identifying the locus of compatibility-based backward crosstalk: Evidence from an extended PRP paradigm. J Exp Psychol Hum Percept Perform 44:261–276. doi: 10.1037/xhp0000445

Kiesel A, Wagener A, Kunde W, et al. (2006) Unconscious manipulation of free choice in humans. Conscious Cogn 15:397–408. doi: 10.1016/j.concog.2005.10.002

Lien M-C, Proctor RW (2002) Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. Psychon Bull Rev 9:212–238. doi: 10.3758/BF03196277

Mattler U, Palmer S (2012) Time course of free-choice priming effects explained by a simple accumulator model. Cogn 123:347–360. doi: 10.1016/j.cognition.2012.03.002

Mayr U, Awh E (2009) The elusive link between conflict and conflict adaptation. Psychol Res 73:794–802. doi: 10.1007/s00426-008-0191-1

Miller J, Ulrich R (2008) Bimanual response grouping in dual-task paradigms. Q J Exp Psychol 61:999–1019. doi: 10.1080/17470210701434540

Naefgen C, Caissie AF, Janczyk M (2017a) Stimulus-response links and the backward crosstalk effect — A comparison of forced- and free-choice tasks. Acta Psychol 177:23–29. doi: 10.1016/j.actpsy.2017.03.010

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

Naefgen C, Janczyk M (2018) Free choice tasks as random generation tasks: An investigation through working memory manipulations. Exp Brain Res 1–13. doi: 10.1007/s00221-018-5295-2

Notebaert W, Gevers W, Verbruggen F, Liefooghe B (2006) Top-down and bottom-up sequential modulations of congruency effects. Psychon Bull Rev 13:112–117. doi: 10.3758/BF03193821

Pashler H (1984) Processing stages in overlapping tasks: Evidence for a central bottleneck. J Exp Psychol Hum Percept Perform 10:358–377 doi: 10.1037/0096-1523.10.3.358

Pashler H (1994) Dual-task interference in simple tasks: data and theory. Psychol Bull 116:220–244 doi: 10.1037/0033-2909.116.2.220

Pfister R, Janczyk M (2013) Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. Adv Cogn Psychol 9:74–80. doi: 10.2478/v10053-008-0133-x

Renas S, Durst M, Janczyk M (2017) Action effect features, but not anatomical features, determine the Backward Crosstalk Effect: Evidence from crossed-hands experiments. Psychol Res. doi: 10.1007/s00426-017-0873-7

Scherbaum S, Fischer R, Dshemuchadse M, Goschke T (2011) The dynamics of cognitive control: Evidence for within-trial conflict adaptation from frequency-tagged EEG. Psychophysiol 48:591–600. doi: 10.1111/j.1469-8986.2010.01137.x

Scherbaum S, Gottschalk C, Dshemuchadse M, Fischer R (2015) Action dynamics in multitasking: The impact of additional task factors on the execution of the prioritized motor movement. Front Psychol 6:934. doi: 10.3389/fpsyg.2015.00934

Schmidt JR (2013) Temporal learning and list-level proportion congruency: Conflict adaptation or learning when to respond? PLOS ONE 8:e82320. doi: 10.1371/journal.pone.0082320

Schuch, S., Dignath, D., Steinhauser, M., & Janczyk, M. (2018) Monitoring and control in multitasking. Psychon Bull Rev https://doi.org/10.3758/s13423-018-1512-z

Stürmer B, Leuthold H (2003) Control over response priming in visuomotor processing: A lateralized event-related potential study. Exp Brain Res 153:35–44. doi: 10.1007/s00221-003-1579-1

Stürmer B, Leuthold H, Soetens E, et al. (2002) Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. J Exp Psychol Hum Percept Perform 28:1345–1363. doi: 10.1037/0096-1523.28.6.1345

Thomson SJ, Danis LK, Watter S (2015) PRP training shows Task1 response selection is the locus of the backward response compatibility effect. Psychon Bull Rev 22:212–218. doi: 10.3758/s13423-014-0660-z

Ulrich R, Miller J (2008) Response grouping in the psychological refractory period (PRP) paradigm: Models and contamination effects. Cogn Psychol 57:75–121. doi: 10.1016/j.cogpsych.2007.06.004

Watter S, Logan GD (2006) Parallel response selection in dual-task situations. Percept Psychophys 68:254–277. doi: 10.3758/BF03193674

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

# Appendix B – Study 2

The following represents the article described in Chapter 6. Springer Nature granted the license to reproduce the final author's accepted manuscript here.

Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: evidence from response threshold manipulations. *Psychological Research*, *82*(6), 1039–1052. doi:10.1007/s00426-017-0887-1

The published version can be found under: https://link.springer.com/article/10.1007/s00426-017-0887-1

RUNNING HEAD: RTs in free and forced choice tasks

# Why free choices take longer than forced choices: Evidence from response threshold manipulations

[a] Christoph Naefgen, [b] Michael Dambacher & [a] Markus Janczyk

[a] Eberhard Karls University of Tübingen, Tübingen, Germany

[b] University of Konstanz, Konstanz, Germany

**Address correspondence to:**
Christoph Naefgen
Eberhard Karls University of Tübingen
Department of Psychology
Schleichstraße 4
72076 Tübingen
Germany
Phone: +49 (0)7071 29 76769
Email: christoph.naefgen@uni-tubingen.de

## Abstract

Response times (RTs) for free choice tasks are usually longer than those for forced choice tasks. We examined the cause for this difference in a study with intermixed free and forced choice trials, and adopted the rationale of sequential sampling frameworks to test two alternative accounts: Longer RTs in free choices are caused (1) by lower rates of information accumulation, or (2) by additional cognitive processes that delay the start of information accumulation. In three experiments, we made these accounts empirically discriminable by manipulating decision thresholds via the frequency of catch trials (Exp. 1) or via inducing time pressure (Exp. 2 and 3). Our results supported the second account, suggesting a temporal delay of information accumulation in free choice tasks, while the accumulation rate remains comparable. We propose that response choice in both tasks relies on information accumulation towards a specific goal. While in forced choice tasks, this goal is externally determined by the stimulus, in free choice tasks it needs to be generated internally, which requires additional time.

**Key words:** Free choice ; Forced choice ; Sequential-sampling ; Response threshold

# Why free choices take longer than forced choices: Evidence from response threshold manipulations

In 1980 the New Wave band Devo claimed that “freedom of choice is what [we] got” and that “freedom from choice is what [we] want”. Indeed, it appears that a *lack* of freedom is what we want in order to speed up our decisions: an increase in choice options can slow down decisions, which has been shown in situations ranging from complex decision making contexts (e.g., Hanoch, Wood, Barnes, Liu, & Rice, 2011) to minimalist laboratory experimental setups (e.g., Merkel, 1885). A specific example of these latter setups are comparisons between so-called forced choice and free choice tasks (Berlyne, 1957).

***Forced choice and free choice tasks and their use in research.*** In the simplest version of forced and free choice tasks (see, e.g., Berlyne, 1957), participants have two response options (e.g., a left and a right key) and are confronted with three different stimuli (e.g., letters or color patches). Participants are instructed to respond to two of these stimuli with prescribed responses (e.g., red → left key press; blue → right key press) – the *forced choice task*. In case of the third stimulus (e.g., white), in contrast, they can choose “freely” from the two response options – the *free choice task*[5].

Notably, and of particular importance to the present study, the vast majority of studies comparing forced and free choice tasks report shorter response times (RTs) in forced choice compared with free choice tasks (e.g., Berlyne, 1957; Janczyk, Nolden, & Jolicoeur, 2015). It is the purpose of the present study to elucidate where this RT difference results from.

One interpretation of the RT difference is that both tasks differ in terms of their underlying response/action selection systems or processes. In this vein, free and forced choice

---

[5] It should be noted that this freedom of choice is often constrained to some degree by instructions such as “choose both response options about equally often”.

tasks have often been used to operationalize qualitatively different *self-generated* (or intentional, internally generated, intention-based, voluntary) and *externally-triggered* (or stimulus-based) actions (e.g., Brass & Haggard, 2008; Herwig, Prinz, & Waszak, 2007; Passingham, Bengtsson, & Lau, 2010; Keller et al., 2006; Waszak et al., 2005). Evidence for such a distinction comes, for example, from research on learning and using associations between bodily movements and their environmental consequences (i.e., their action effects), a field that was inspired by Ideomotor Theory (e.g., Harleß, 1861; Greenwald, 1970; Shin, Proctor, & Capaldi, 2010; Stock & Stock, 2004). In particular, when specific bodily movements are consistently followed by an auditory stimulus as an action effect (e.g., left key → low-pitch tone, right key → high-pitch tone), results from some studies suggested that associations between the movements and the effects are only learned in free choice tasks, that is, in an intention-based action control mode (Herwig et al., 2007; see also Gaschler & Nattkemper, 2012; Herwig & Waszak, 2009, 2012; Pfister, Kiesel, & Melcher, 2010).

This claim is, however, controversial. For example, Pfister, Kiesel, and Hoffmann (2011) reported learning of action effects even in forced choice tasks, and many other studies observed clear evidence for a role of action effects for performance in forced choice tasks (e.g., Janczyk, Pfister, Crognale, & Kunde, 2012; Janczyk, Pfister, Hommel, & Kunde, 2014; Janczyk, Pfister, & Kunde, 2012; Janczyk, Skirde, Weigelt, & Kunde, 2009; Kühn, Elsner, Prinz, & Brass, 2009, Exp. 3; Kunde, 2001; Kunde, Pfister, & Janczyk, 2012; Wolfensteller & Ruge, 2011). Furthermore, studies using the response-effect (R-E) compatibility paradigm (Kunde, 2001) reported R-E compatibility effects of the same size in forced and free choice tasks (e.g., Janczyk, Durst, & Ulrich, 2017), and the size of dual-task interference is also comparable for both tasks (Janczyk, Nolden, et al., 2015). In addition, Janczyk, Dambacher, Bieleke, and Gollwitzer (2015) used the Psychological Refractory Period (PRP) paradigm in combination with the locus of slack logic (Schweickert, 1978; see also Janczyk, 2013, 2017, or

Miller & Reynolds, 2003, for applications) to identify the source of the RT difference within the stream of processing. Based on Gollwitzer's (1999) implementation intention account, they argued for a perceptual locus, and indeed reported evidence in support of this idea in their study. Essentially, their observations suggest that the RT difference actually results from facilitated perceptual processing of forced choice stimuli.

In light of the evidence summarized in the last paragraph and the importance of forced and free choice tasks in contemporary research, we argue that *effect* or *goal state anticipation* drives response selection in both forced and free choice tasks, but that for the latter task the effect must be self-generated, which comes with additional demands. Here, we investigate further whether both tasks and their RT difference can be described within a common theoretical framework. This will help understanding the sources of the RT difference between the tasks.

***A sequential sampling account of the RT difference.*** Sequential sampling models offer tools to delineate the source(s) of the RT difference between free and forced choice tasks. These approaches assume that evidence for one or the other response is (noisily) accumulated until a decision threshold is reached and the corresponding response is initiated (for an overview, see Ratcliff, Smith, Brown, & McKoon, 2016). The best-known model of this class is the drift-diffusion model, proposed by Ratcliff (1978). While many sophisticated models from this family feature high complexity, the present study focuses on a very simple model with three parameters reminiscent of the features in Grice's (1968) variable criterion model: (1) the decision thresholds that must be reached in order to count as a decision and to initiate emission of a response, (2) the non-accumulation time reflecting all the time before and after the accumulation time proper (i.e., early perceptual processing, motor execution, and perhaps other additional processes), and (3) the drift rate reflecting the strength of evidence for one particular response, and thus the amount of evidence for each response added at each time-step.[6] With higher drift rates, for example, a threshold is on average reached earlier resulting in shorter RTs and fewer errors (which occur when the incorrect threshold is reached, e.g., due to the noise in the accumulation process). Further, lowering the threshold (using a more liberal criterion) yields shorter RTs but more errors (because the chance of reaching the incorrect threshold increases), and augmenting the thresholds (using a more conservative criterion) yields longer RTs and fewer errors. Importantly, the exact kind of evidence that is accumulated is not further specified within this model. In a simple two-alternative forced choice task, one may think of an individual stimulus as the immediate cause of evidence accumulation into one or the other direction, but as already noted in the previous section, it is also conceivable that an anticipated effect or goal state is the source of evidence being accumulated.

Assuming that effect or goal anticipation in the case of a free choice must happen endogenously (without the stimulus entirely determining the goal as in the case of forced choice trials), at least two scenarios can explain the RT difference between forced and free choice tasks within the framework described above: (1) Accumulation starts at the same time in both tasks, but the evidence driving the accumulation process towards one of the response thresholds is weaker in free choice tasks and thus the drift rate is lower (see Figure 1, left panel). (2) Longer RTs in free choice tasks can also result when drift rates are the same in both tasks, but additional time is needed before (or after) the start of accumulation in free choice tasks (for a more thorough description of the consequences of different onsets of information accumulation, see

---

[6] Mattler and Palmer (2012) also used a sequential sampling approach to investigate how priming affects performance and choices in both types of tasks. They observed that while masked primes always influence forced choice RTs, free choices are not influenced when the stimuli (prime and target) are of arbitrary shape. They also specified an accumulator model to explain the data, with the notable assumption of rapidly shrinking threshold separations after onset of a free choice stimulus. In their paper, they conclude that forced choice priming is a result of the integration of the automatic processing of primes and evidence from the stimulus while free choice priming is based on the integration of "external stimulation by the prime and internal response tendencies".

Bausenhart, Rolke, Seibold, & Ulrich, 2010). In this case, the additional delay would be reflected in the non-accumulation time (see Figure 1, right panel).

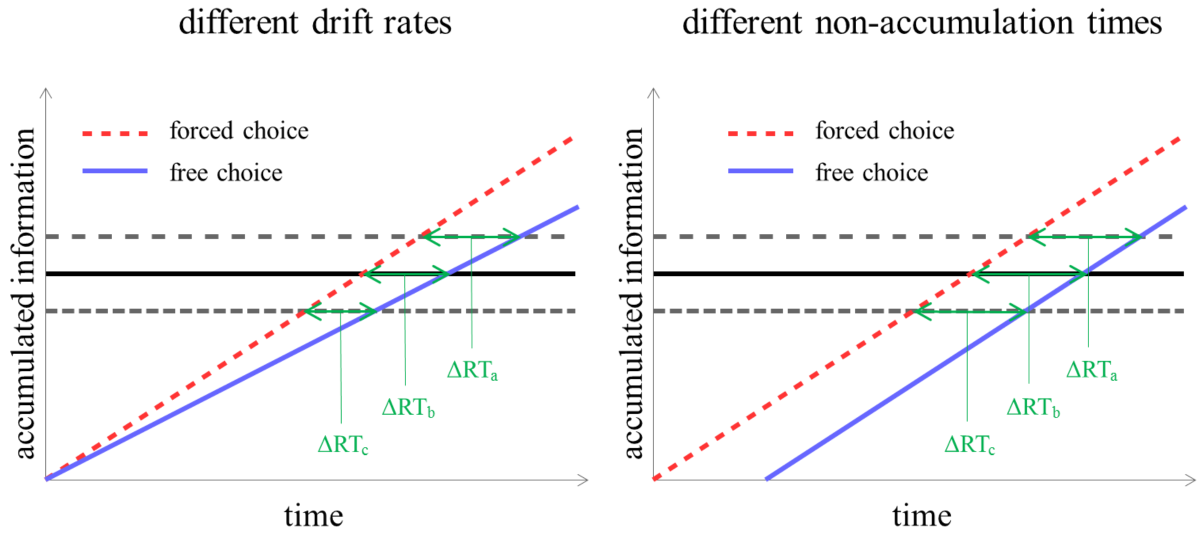

**Figure 1**. Two accounts of the mean RT difference between free and forced choice tasks. The continuous black line represents a medium decision threshold. The dashed line indicates an increased decision threshold and the dotted line indicates a lowered threshold. Under the "drift rate" account, the RT difference between forced and free choice tasks becomes smaller with lower thresholds ($\Delta RT_a > \Delta RT_b > \Delta RT_c$). In other words, task and threshold manipulation interact with each other. In contrast, with differences in non-accumulation times, the RT difference remains the same irrespective of the threshold ($\Delta RT_a = \Delta RT_b = \Delta RT_c$) and therefore reflects an additive relation between task and threshold manipulation.

Even though on the global level both accounts predict longer RTs in free than in forced choice tasks, there is a way to empirically distinguish them by manipulating the decision thresholds. Under the assumption of different non-accumulation times but equal drift rates, the RT difference between free and forced choice tasks should be independent of the actual threshold (see Figure 1, right panel) and therefore of the same size under liberal and conservative criteria. Thus, task type and the manipulation of the decision thresholds should combine additively because gathering the required additional information takes the same amount of time when both types of task have the same speed of information acquisition. To use a metaphor: If two horses in a horse race run at the same speed but one horse starts five meters closer to the goal than the other horse, the distance between the two horses when they cross the finishing line will not change, even if the goal is moved closer to or farther away from the starting point of the race. In contrast, if there is a difference in the drift rates between the two tasks, the RT difference should become smaller the lower the threshold and bigger the higher the threshold is (see Figure 1, left panel). In other words, task type and the manipulation should statistically interact. In the horse race metaphor this means that one horse is faster than the other but they start in the same position. Over the course of the race, the distance between the two horses would increase. If the race is short (liberal criterion), there is less time for the distance to increase, whereas distance can increase in a longer race (conservative criterion) resulting in larger differences.

Two previously established methods of manipulating decision thresholds are the amount of catch-trials in an experimental block and time pressure. Catch-trials are trials in which no stimulus appears at the time when a stimulus would normally appear. Participants are instructed not to react to this absence of a stimulus. Generally, the more catch-trials there are, the longer the reaction will take (e.g., Gordon, 1967; Näätänen, 1972). It has been theorized, that this is because a higher amount of catch-trials leads to a decreased stimulus expectancy, which in turn leads to a higher and thus more conservative decision threshold (e.g., Brysbaert, 1994; Grice, Nullmeyer, & Spiker, 1982; Seibold, Bausenhart, Rolke, & Ulrich, 2011). Another manipulation of the threshold is to vary the time available for responding, that is, varying the time pressure. Increasing time pressure has been repeatedly theorized and empirically shown to lower the decision criterion (e.g., Diederich, 1997; Dror, Basola, & Busemeyer, 1999; Forstmann et al., 2008; Ratcliff & McKoon, 2008).

***The present experiments.*** The aim of the present study was to investigate the RT difference between forced and free choice tasks and to distinguish between the two accounts introduced in the previous section. In Experiment 1, we varied the amount of catch-trials in order to manipulate thresholds (Näätänen, 1972; Seibold et al., 2011). In Experiments 2 and 3, we manipulated the response deadline (thus inducing time pressure) to manipulate the thresholds (Ratcliff & McKoon, 2008).

## Experiment 1

In Experiment 1, participants worked on forced and free choice tasks that were randomly intermingled. We expected longer RTs in the free than in the forced choice task (Berlyne, 1957). The critical manipulation was the proportion of catch-trials within a block (0, 25, 50, or 75%), in which no stimulus appeared and thus no response was to be given. If task type and the catch-trial manipulation affect RTs additively, this would support the idea of comparable drift rates but longer non-accumulation times in free choice tasks. In contrast, if both interact in a way that the RT difference increases with the amount of catch-trials, this would favor an account in terms of different drift rates.

### Methods.

***Participants.*** Thirty-two persons from the Tübingen area participated (mean age = 24 years; standard deviation = 3 years; 27 female; one unknown value for age) for monetary compensation or course credit. 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 were done via a standard PC connected to a 17-inch CRT monitor. Stimuli were red, green, and white circles, presented against a black background. Manual responses were collected with the two CTRL keys on a standard keyboard placed on the table in front of the participants.

***Tasks and procedure.*** The task was either to give a predefined response to two of the possible colors (forced choice task: red and green stimuli), or to freely choose one of the two possible responses to the third color (free choice task: white stimulus). On catch-trials, where no stimulus appeared, the participants were instructed not to respond at all. Prior to each block, participants were informed about the percentage of catch-trials in this block. A trial began with the presentation of a small fixation cross (250 ms; see Figure 2). Following a blank screen (250 to 350 ms), the stimulus appeared and remained on screen until the response was made. A trial was terminated if no response was given within 1500 ms after stimulus onset. General errors (i.e., no response in non-catch-trials within the time limit of 1500 ms and responses before stimulus appearance) and erroneous responses (response in a catch-trial or wrong key in forced choice trials) triggered respective feedback (1000 ms). The next trial started after an inter-trial interval (ITI) of 1000 ms. Eight blocks of 120 trials (all three stimuli appeared equally often in the normal non-catch-trials) were administered. The amount of catch-trials was varied across four block types (0%, 25%, 50%, 75%). The first four blocks (one of each type) were ordered by a Latin Square, and the order of the next four blocks was the reverse of the first four blocks.

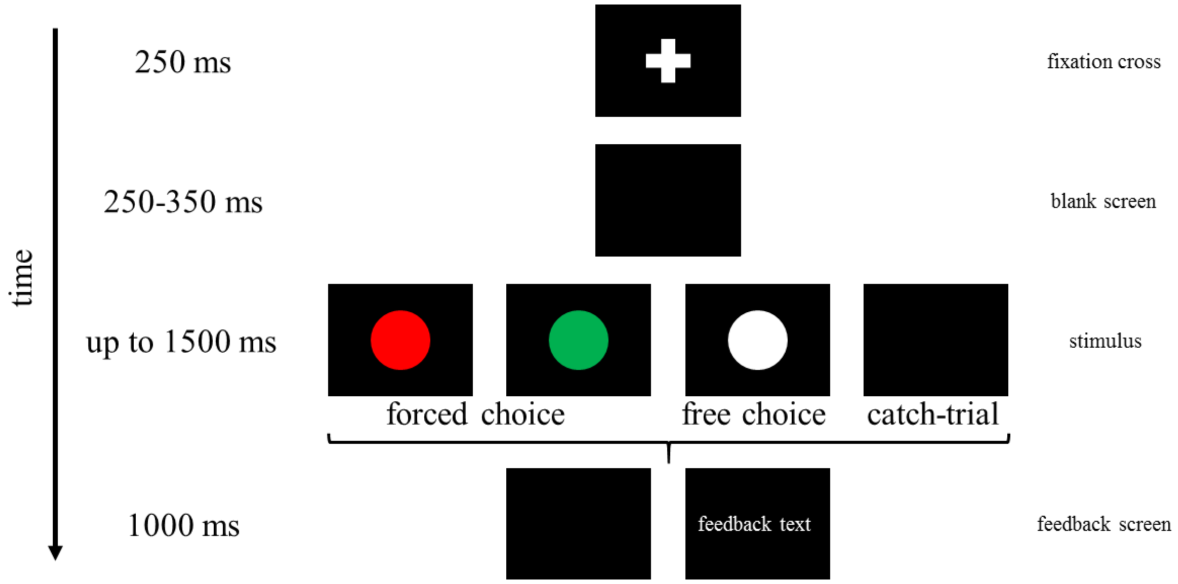

**Figure 2.** Time course of a trial. No feedback text was given if no error occurred. Feedback text was displayed in German and described the type of error made (“Wrong key!”, “Too slow!”, “No stimulus was given!”).

Participants were tested individually in one single session of about 45 minutes. Written instructions emphasized speed as well as accuracy and, for the free choice trials, an even distribution of left and right responses as well as the avoidance of patterns in maintaining this distribution. The mappings of stimuli and responses in the forced choice task and the order of blocks were counterbalanced across participants. The data of participants whose free choice responses showed a strong bias towards one response option (>80% of choices) were discarded and new data were collected from new participants with the same block sequence (three participants in this experiment).

***Design and analyses.*** The experimental manipulations resulted in two independent variables of interest, namely (1) task type (forced choice vs. free choice) and (2) block type (0% vs. 25% vs. 50% vs. 75% catch-trials). Trials with general errors were discarded. For RT analyses only correct responses were considered (note that no erroneous responses can be made in free choice tasks). Trials with RTs deviating more than 2.5 standard deviations from the participants' mean per condition were excluded. Data were then submitted to a 2 × 4 Analysis of Variance (ANOVA) with repeated-measures on task type and block type. Percentages of errors (PEs) were only analyzed for the forced choice task with an ANOVA with block type as repeated measures factor. The choice rates in the free choice task were analyzed similarly as a function of block type. *p*-values were Greenhouse-Geisser adjusted when the assumption of sphericity was violated. In these cases the respective ε is reported.

**Results.**

Participants chose the left response button in the free choice task about 48.7% of the time in the 0%, 48.0% in the 25%, 47.2% in the 50%, and 50.6% in the 75% catch-trials blocks. These differences were not significant, $F(3,93) = 1.13$, $p = .331$, $\eta_p^2 = .04$, $\varepsilon = .69$.

Mean correct RTs (2.5% excluded as outliers) are shown in Figure 3 and are summarized in Table 1. As expected, responses in the forced choice task were faster than in the free choice task, $F(1,31) = 55.50$, $p < .001$, $\eta_p^2 = .64$, and responses slowed down with an increasing amount of catch-trials in a block, $F(3,93) = 102.59$, $p < .001$, $\eta_p^2 = .77$, $\varepsilon = .63$. This latter result suggests that the manipulation worked as intended and increased the decision thresholds. Most importantly, there was a significant interaction between block type and task type, $F(3,93) = 3.15$, $p = .048$, $\eta_p^2 = .09$, $\varepsilon = .69$. A closer look at Figure 3, however, suggests that this interaction is driven by the smaller RT difference in the 0% catch-trials blocks compared to the other blocks, and arguably the 0% blocks differ in an important aspect from the other blocks: While in the 0% condition participants knew that a response is always required, in the other blocks the additional demand of distinguishing normal from catch-trials was imposed. The drift rate account, however, predicts an increasing RT difference across all levels of increasing decision thresholds (with growing differences, as the amount of catch-trials increases). Therefore, we re-analyzed the data but omitted the 0% catch-trial blocks. Again the two main effects were significant as expected, task type: $F(1,31) = 47.33$, $p < .001$, $\eta_p^2 = .60$, and block type: $F(2,62) = 90.47$, $p < .001$, $\eta_p^2 = .74$, $\varepsilon = .87$. Clearly, however, their interaction was not significant, $F(2,62) = 0.36$, $p = .697$, $\eta_p^2 = .01$.

PEs in the forced choice task (i.e., wrong response keys pressed) increased with the amount of catch-trials (see Table 1), $F(3,93) = 8.41$, $p = .001$, $\eta_p^2 = .21$, $\varepsilon = .57$. Finally, there was a negligible amount (< 0.1%) of catch-trials in which a response was given.

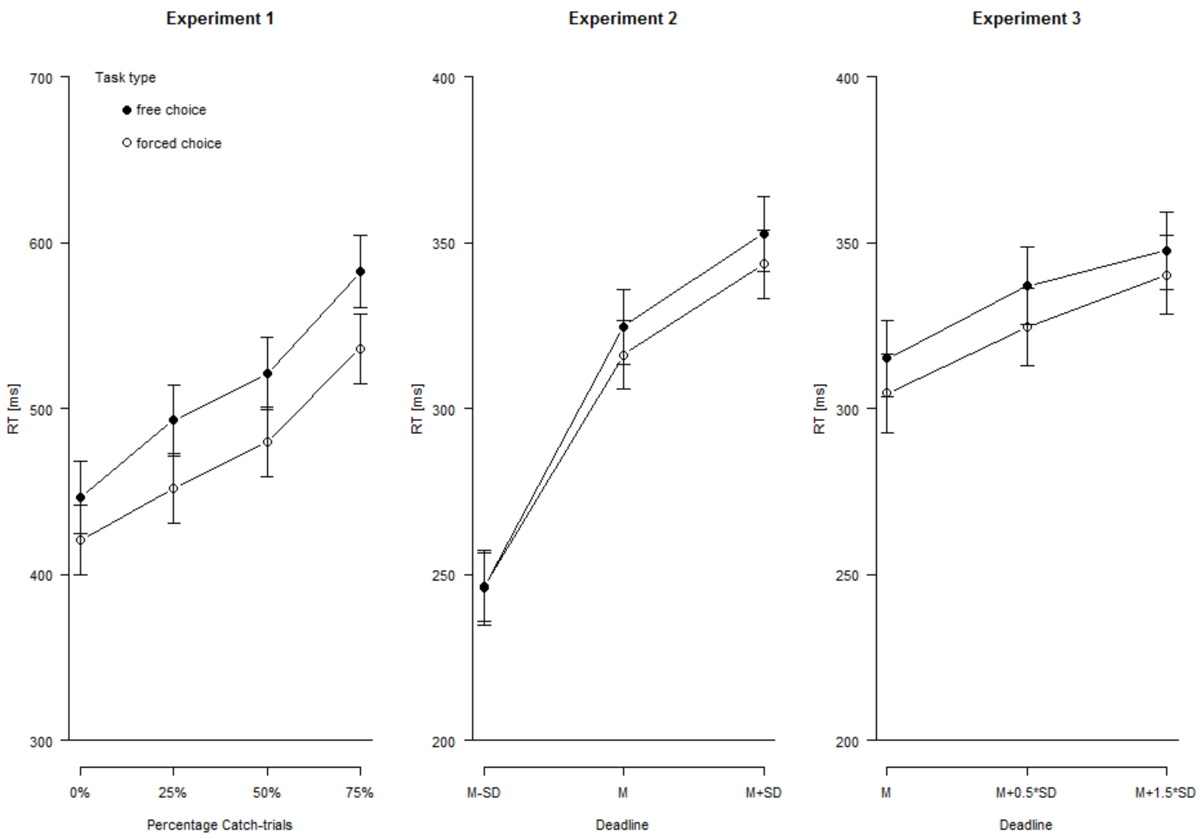

**Figure 3.** Mean correct RTs in milliseconds (ms) from all three experiments as a function of task type and block type. Error bars are 95% within-subject confidence intervals calculated for the difference between free and forced choice tasks collapsed across block types (see Pfister & Janczyk, 2013).

**Table 1.** Means (and standard deviations; SD) of response times (RTs) and percentages of errors (PEs) in forced choice tasks of Experiments 1-3 as a function of block type and trial type. Deadline conditions are denoted in SD steps from M (i.e., M+x*SD with x denoting the block type in the table.)

| | | Experiment 1 | | | | Experiment 2 | | | Experiment 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | %Catch-trials | | | | Deadlines | | | Deadlines | | |
| Task type | | 0% | 25% | 50% | 75% | -1 | 0 | 1 | 0 | .5 | 1.5 |
| Free choice | RT | 446 | 493 | 521 | 582 | 246 | 325 | 353 | 315 | 337 | 348 |
| | | (66) | (77) | (83) | (106) | (39) | (39) | (36) | (36) | (39) | (40) |
| Forced choice | RT | 421 | 452 | 480 | 536 | 246 | 316 | 343 | 305 | 324 | 340 |
| | | (66) | (73) | (79) | (93) | (46) | (35) | (35) | (35) | (41) | (42) |
| | PE | 4.6 | 4.9 | 4.8 | 8.2 | 29.6 | 14.5 | 10.5 | 17.4 | 14.4 | 12.5 |
| | | (3.3) | (4.2) | (5.1) | (7.6) | (10.2) | (6.1) | (5.7) | (8.2) | (7.5) | (6.5) |

***Discussion.***

The results of Experiment 1 are not in line with the drift rate account, but more compatible with differences in the non-accumulation time. Yet, they entail several aspects that complicate a straightforward interpretation.

First, RTs increased with increasing amount of catch-trials, an observation that complies with the intended manipulation of increasing thresholds (see also Näätänen, 1972; Seibold et al., 2011). At the same time, though, more errors were made in the forced choice task as well. This is unexpected, since increasing thresholds should make errors *less* likely.[7] We will get back to this in the General Discussion.

Second, we replicated the common observation of longer RTs in the free than in the forced choice task, and task type interacted with the amount of catch-trials in the initial analyses,

---

[7] A similar observation with PEs increasing descriptively with the amount of catch-trials can be seen in the condition with low intensity stimuli in the study by Seibold et al. (2011; see their Fig. 4).

which included the 0% catch trials condition. Straightforwardly, this would argue against the account of equal drift rates with the differences arising from different non-accumulation times. Yet, different drift rates should result in increasing RT differences across all amounts of catch-trials, and clearly this was not the case. Rather, the interaction was driven by a smaller RT difference in the 0% catch-trial blocks, perhaps reflecting the absence of the additional demand of distinguishing normal from catch-trials in the other blocks. When considering only the comparable blocks with catch-trials, task type and block type combined additively.

Tentatively, we therefore take the results as support for the account of different non-accumulation times between the tasks (see Figure 1, right panel). At the same time, we wish to avoid pre-mature conclusions on this single experiment. Accordingly, in the following experiments, we sought for converging evidence and employed time-pressure as a different means of manipulating response thresholds.

## Experiment 2

Experiment 2 followed the same logic as Experiment 1, but time pressure was used to manipulate response thresholds. To individually adjust time limits, the mean and the standard deviation of participants' RTs in free and forced choice tasks were determined first. Subsequently, the same tasks were presented with three different levels of time pressure that were announced prior to each block and were varied block-wise.

### Methods.

***Participants.*** Thirty-six persons from the Tübingen area participated (mean age = 23 years; standard deviation = 4 years; 31 female) for monetary compensation or course credit. All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection.

***Stimuli and procedure.*** Stimuli were adopted from Experiment 1. The task was largely the same, except that there were no catch-trials, and the blank screen interval between the fixation cross and the stimulus' appearance was fixed to 250 ms. At the beginning, two pre-experimental blocks with a response window of 1500 ms assessed mean (*M*) RTs  of each participant separately for free and forced choices. The respective *M*s and their standard deviations (*SD*s) were then used to calculate three different response deadlines separately for the free and forced choice tasks: long (*M*+*SD*), medium (*M*), and short (*M*-*SD*). Then three experimental blocks, one of each deadline condition, followed. The order of these blocks and the S-R mapping within the forced choice task were fully counterbalanced. After these three blocks, another three blocks in reverse order followed. At the beginning of every block, the time limit of the task type with the shorter deadline (determined in the first two blocks, see above) was announced to the participants. After each block, participants were informed about how long their responses took on average (averaged across both free and forced choice trials). The same exclusion criterion as in Experiment 1 was used, and data from one participant were discarded and replaced by a new data set in the same condition.

***Design and analyses.*** The experimental manipulations resulted in two independent variables of interest, namely (1) task type (forced choice vs. free choice) and (2) block type (*M*-*SD* vs. *M* vs. *M*+*SD* response deadline). Trials with general errors were discarded. For RT analyses, only correct responses were considered (note that no erroneous responses can be made in free choice tasks), and trials with RTs deviating more than 2.5 *SD*s from the participants' mean per condition were excluded as outliers from analyses. Data from the experimental blocks were then submitted to a 2 × 3 repeated-measures ANOVA on task type and block type. Error data were only analyzed for the forced choice task by means of an ANOVA with repeated measures on block type. The choice rates in the free choice task were analyzed similarly, but included the pre-experimental blocks. *p*-values were Greenhouse-Geisser adjusted when the assumption of sphericity was violated. In these cases the respective $\varepsilon$ is reported.

## Results

Participants chose the left response button in the free choice task about 54.0% of the time in the *M+SD* blocks, 56.1% in the *M* blocks, 58.3% in the *M-SD* blocks, and 51.1% in the pre-experimental blocks, and the main effect of block type was significant, $F(3,105) = 5.65$, $p = .004$, $\eta_p^2 = .14$, $\varepsilon = .72$. In the pre-experimental blocks, mean RTs were 423 ms in the forced choice task and 444 ms in the free choice task, $F(1,35) = 18.16$, $p < .001$, $\eta_p^2 = .34$.

Mean correct RTs (1.3% excluded as outliers) are shown in Figure 3 (middle panel) and are summarized in Table 1. As expected, there was a main effect of block type on RTs, $F(2,70) = 363.32$, $p < .001$, $\eta_p^2 = .91$, $\varepsilon = .75$, with higher time pressure induced by shorter response deadline resulting in shorter RTs, as well as a main effect of task type, $F(1,35) = 5.72$, $p = .022$, $\eta_p^2 = .14$, with longer RTs in the free choice task compared to the forced choice task. The interaction between block type and task type was also significant, $F(2,70) = 4.50$, $p = .021$, $\eta_p^2 = .11$, $\varepsilon = .81$. Inspection of the RTs revealed virtually no RT difference between both tasks in the high time-pressure (*M-SD*) block, which may point to a large proportion of fast guesses in this condition. Indeed, the PEs in this block ranged from 9.6% to 46.6%, that is, close to chance level. Thus, we performed a median split based on error rates in this condition (with mean PEs in the *M-SD* condition of 22% and 38% for the below- and above median groups, respectively), and ran an ANOVA that included this grouping variable. This ANOVA yielded an almost significant interaction between block type, task type, and the grouping variable, $F(2,68) = 2.98$, $p = .068$, $\eta_p^2 = .08$, $\varepsilon = .84$, and we continued to analyze both groups separately. As expected, for the participants with the above-median PEs, the interaction of task type and block type was significant, $F(2,34) = 6.41$, $p = .004$, $\eta_p^2 = .27$. In contrast, for the other group of participants with lower PEs – and thus a performance not as close to chance level – the interaction was far from significance, $F(2,34) = 0.23$, $p = .799$, $\eta_p^2 = .01$.

The PEs in the forced choice task increased with shorter response deadline, $F(2,70) = 52.16$, $p < .001$, $\eta_p^2 = .60$, $\varepsilon = .63$ (see Table 1).

**Discussion.**

In this experiment, we manipulated the thresholds by inducing time pressure with a response deadline. First, and as expected, RTs were shorter the more time-pressure was induced in a block, and also the PEs (in the forced choice task) increased accordingly. This pattern suggests that the time pressure manipulation worked as intended. Secondly, the initial analysis revealed a significant interaction of task and block type. Taking into account PEs, however, post-hoc analyses indicated that this interaction likely resulted from a substantial proportion of fast guesses in the high time-pressure condition, which undermines the validity of the measured performance. When considering only the half of participants with below-median PEs, the interaction vanished, and results are compatible with our tentative proposal from Experiment 1, favoring an account in terms of comparable drift rates but different non-accumulation times. Also, the RT difference in the other two blocks remained constant, whereas the drift rate account would predict an increase of the RT difference in the longer deadline.

To further validate our conclusion that longer RTs in free than in forced choices are due to differences in non-accumulation times rather than in drift rates, we ran Experiment 3. This experiment was essentially a repetition of Experiment 2, but with less severe time pressure to avoid the high error rates that supposedly resulted from fast guesses.

## Experiment 3

Experiment 3 used the same setup as Experiment 2 except that we used response deadlines of $M$, $M+0.5*SD$, and $M+1.5*SD$ to avoid fast guesses as in the very short time limit in Experiment 2. We expected an additive combination of task type and block type in the present experiment.

### Methods.

Thirty-six persons from the Tübingen area participated (mean age = 23 years, standard deviation = 4 years; 29 female) for monetary compensation or course credit. All participants reported normal or corrected-to-normal vision, were naïve regarding the underlying hypotheses, and provided written informed consent prior to data collection. This experiment was identical to Experiment 2 in all regards with the exception of the time limits, which were set at *M*, *M*+0.5**SD*, and *M*+1.5**SD*.

### Results.

Participants chose the left response button in the free choice task in about 55.6% of the time in the *M* blocks, 55.0% in the *M*+0.5**SD* blocks, 55.6% in the *M*+1.5**SD* blocks, and 52.3% in the pre-experimental blocks. These differences were not significant, $F(3,105) = 1.80$, $p = .164$, $\eta_p^2 = .05$, $\varepsilon = .82$. In the pre-experimental blocks, mean RTs were 398 ms in the forced choice condition and 417 ms in the free choice condition, $F(1,35) = 15.00$, $p < .001$, $\eta_p^2 = .30$.

Mean correct RTs (1.3% excluded as outliers) are shown in Figure 3 (right panel) and are summarized in Table 1. As expected, there was a main effect of block type, $F(2,70) = 98.41$, $p < .001$, $\eta_p^2 = .74$, $\varepsilon = .96$, with shorter response deadlines resulting in shorter RTs, as well as a main effect of task type, $F(1,35) = 12.92$, $p = .001$, $\eta_p^2 = .27$, with longer RTs in free choice tasks compared with the forced choice tasks. The interaction between block type and task type was not significant, $F(2,70) = 2.07$, $p = .133$, $\eta_p^2 = .06$.

The PEs in the forced choice task decreased with increasing response deadlines, $F(2,70) = 11.93$, $p < .001$, $\eta_p^2 = .25$, $\varepsilon = .88$ (see Table 1).

### Discussion.

In Experiment 3 we observed no significant interaction between block type and task type and, if anything, the numerical decrease of the RT effect with longer deadlines was in a direction incompatible with the drift rate account (see Figure 1, left panel). Rather, the results are in line with predictions of different non-accumulation times between the tasks.

## General Discussion

Three experiments were run to elucidate the source of the RT difference between forced and free choice tasks. We used the sequential sampling framework to derive two hypotheses (see Figure 1): First, the difference can arise from differences in the speed of evidence accumulation with drift rates being smaller for free choice tasks (see Figure 1, left panel). Second, the difference can arise from differences in the non-accumulation time with a later onset (but a similar rate) of accumulation in the case of free choice tasks (see Figure 1, right panel). To distinguish these two accounts we manipulated the response thresholds by varying the amount of catch-trials per block in Experiment 1 (Näätänen, 1972; Seibold et al., 2011) and by inducing time pressure via response deadlines in Experiments 2 and 3 (Ratcliff & McKoon, 2008).

***Summary of results.*** First, in all experiments, forced choice stimuli were responded to faster than free choice stimuli. Second, the manipulations of catch-trials and response deadlines effectively changed the overall level of RTs as expected. However, evidence about the nature of interactions between these manipulations and task type was rather mixed. In Experiments 1 and 2 we observed significant interactions, which seemingly argue against non-accumulation differences. In Experiment 1, though, this interaction was attributable to the block without any catch-trials, thus without an additional demand of distinguishing normal and catch-trials. In Experiment 2, no RT difference between forced and free choice tasks was evident in the high time-pressure condition, and RTs were only about 250 ms. We suspected a large proportion of fast guesses in this case, and indeed only the participants with above-median PEs yielded a significant interaction. In the group with below-median PEs, and thus a performance not as close to chance level, the interaction vanished. Admittedly, excluding trials or reducing the number of participants lowers the statistical power for detecting an interaction. However, as a further aspect the drift rate account predicts increasing RT differences with increasing thresholds, and this was not even descriptively the case. The clearest evidence against the drift rate account, however, comes from Experiment 3. This experiment was a repetition of Experiment 2 without a very high level of time-pressure. In this experiment, no interaction was observed and the results are compatible with the predictions derived from assuming differences in the non-accumulation time.

Overall, it seems that the drift rate account received little if any support from these results. In contrast, we did not observe evidence against the idea that there is a difference in non-accumulation times between free and forced choice tasks. Therefore, we suggest that the RT difference between free and forced choice tasks is at least partly caused by additional processes subsumed in the non-accumulation times of free choice tasks.

***Limitations.*** One odd result in Experiment 1 is that PEs increased with increasing proportions of catch-trials. Because a higher PE, especially together with a longer RT, is compatible with a lower drift rate, a possible explanation would be that the manipulation in Experiment 1 targeted the drift rates instead of decision thresholds. This would have broader implications for every argumentation that requires the assumption or concludes that the amount of catch-trials influences (only) the decision thresholds (e.g., in Brysbaert, 1994; Grice et al., 1982; Seibold et al., 2011). Should the manipulation through catch-trials target the drift rates instead of or additionally to the decision thresholds, this of course complicates the interpretation of the results of Experiment 1. The mean RTs and PEs in Experiments 2 and 3, though, were in line with our assumptions about the manipulation of decision thresholds as PEs increased while RTs became shorter with shorter response deadlines and thus increasing time pressure. It should be noted, though, that also for time pressure manipulations and speed-accuracy tradeoff instructions, concerns have been raised that not (only) decision thresholds but also other parameters such as the drift rate change (e.g., Arnold, Bröder, & Bayen, 2015; Dambacher & Hübner, 2015; Rae, Heathcote, Donkin, Averall, & Brown, 2014; Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004).

To check which parameters our manipulations affected, we extracted the parameters for the forced choice trials with EZ (Wagenmakers, Van Der Maas, & Grasman, 2007) and compared the parameter estimates across the block types for each experiment (see the Appendix for a summary of these analyses). Notably, in Experiments 1 and 2 the drift rates were indeed affected by the manipulations (i.e., smaller drift rates for conditions which should only have higher decision thresholds in Experiment 1 and larger drift rates for conditions which should only have higher decision thresholds in Experiment 2). No such effect was observed for Experiment 3, which was also the one with the most straightforward and clear data pattern in RTs and PEs. Further, in Experiment 1, there was no significant effect on the extracted response thresholds, while for Experiments 2 and 3 the threshold pattern matched our theoretical assumptions. Finally, in all three experiments there was a significant influence on the non-accumulation times, which increased with the decision thresholds.

While the results from Experiments 1 and 2 must be interpreted with some caution, it is unclear whether the effects in drift rates are due to trade-offs in parameter estimation itself. If not, previous studies using similar manipulations may suffer from the same limitations, which potentially has broader implications for other research fields. Importantly, the results of Experiment 3 revealed no drift rate effect.

Another potential limitation is that our conclusion is based on retaining the null-hypotheses of the critical (2×3) interaction effects. To facilitate interpretation of the results, we ran a power analysis using GPower (Faul, Erdfelder, Lang, & Buchner, 2007). To achieve a power of $1\text{-}\beta = .8$ with $\alpha = .05$ and $\rho = .3$ to detect a medium sized effect, the calculated required sample size was $n = 20$.[8]

***Potential additional processes.*** If we accept that RT differences in free and forced choices are due to differences in the non-accumulation time and assume that there are one or more additional cognitive processes involved in free choice task performance: what is known about them? They are most likely not or only minimally influenced in their duration by manipulations of stimulus features, because in Experiment 3 of Janczyk, Dambacher et al. (2015) stimulus brightness only affected forced choice RTs but not free choice RTs (both task types were intermixed in the same blocks). As there is an alternative explanation of the latter result (that participants only ruled out the presence of a forced choice stimulus instead of identifying free choice stimuli), this should be seen as a tentative conclusion. We discuss candidates for the additional processes in the following.

(1) *Memory processes triggered one or the other response (trial history bias):* Part of the premise of free choice tasks is that participants are asked to respond roughly with the same amount with each response option and without a clear pattern, essentially asking the participants to act as (pseudo-)random number generators for the experiment. The breadth of the literature on random number generation alone suggests that this task is not trivial and can be approached in many different ways. Participants either really generate random numbers or they try to generate patterns that ‘feel’ random but are, in fact, not. Various biases in human random number generation are known see also Heuer, Janczyk, & Kunde, 2010, for an overview).

---

[8] To correct the effect size entered into GPower, we used the method described by Rasch, Friese, Hofmann, and Naumann (2010).

Examples are a lack of symmetrical response sequences, a lack of long runs of the same response, or a balancing of responses across short sequences (Bar-Hillel, & Wagenaar, 1991). Both negative and positive recency effects (i.e., lowered and heightened chances of repetitions) can be observed under different circumstances (Ayton, & Fischer, 2004). We suggest that investigating what strategies, if any, are used to generate the pattern of decisions in free choice tasks could provide insight into the processes that are subsumed in the non-accumulation time. To shed some light on whether a free choice is affected by the immediate history of responses in the preceding trials, we ran a post-hoc analysis on choice frequencies. In particular, when comparing the ratios of left to right free choice responses with the ratios of the same type conditional on the previous response (left or right) and type of task (free or forced choice), there were significant differences for all three experiments (see Table 2)[9], Experiment 1: $F(4,124) = 6.11$, $p = .005$, $\eta_p^2 = .16$, $\varepsilon = .451$; Experiment 2: $F(4,140) = 8.16$, $p = .002$, $\eta_p^2 = .19$, $\varepsilon = .404$; Experiment 3: $F(4,140) = 16.39$, $p < .001$, $\eta_p^2 = .32$, $\varepsilon = .438$. Interesting, the resulting pattern of choice frequencies bears similarities to reports in the task switching literature where a response repetition benefit (in RTs) is only observed when the task repeats but not when the task switches (e.g., Kleinsorge, 1999; Rogers & Monsell, 1995). Whether or not the present result of fewer response repetitions following a switch from a forced choice to a free choice extends this effect is open to future research.

In sum, these observations point to the idea that the responses in the immediately preceding trials were considered on a current free choice trial. In other words, participants seem to use systematic strategies to decide what response to give, which takes time and adds to the RTs in free choice tasks.

---

[9] We thank one of the reviewers for this suggestion.

**Table 2.** Percentage of left responses in free choice trials, both unconditional (column overall) and conditional on the previous trial being a free or forced choice task and a left or right response.

| | Overall | Trial *n*-1 forced choice, left | Trial *n*-1 forced choice, right | Trial *n*-1 free choice, left | Trial *n*-1 free choice, right |
|---|---|---|---|---|---|
| Experiment 1 | 51.2% | 43.5% | 58.1% | 58.6% | 41.8% |
| Experiment 2 | 45.5% | 39.8% | 48.1% | 57.6% | 44.2% |
| Experiment 3 | 45.6% | 42.0% | 47.1% | 61.3% | 37.8% |

(2) *Endogenous generation of stimulus/effect representations:* Free choice tasks as used in this study are usually intermixed with forced choice tasks. In the introduction we stated that the exact basis of evidence accumulation is not fully specified in diffusion models. First, after having realized to be in a free choice trial, participants may endogenously generate a representation of one of the two forced choice stimuli and evidence is then accumulated for internal representations of these stimuli that are associated with one or the other response. Second, according to Ideomotor Theory (e.g., Greenwald, 1970; Harleß, 1861; Shin, Proctor, & Capaldi, 2010) bodily movements are always addressed via an anticipation of the sensorial consequences of these movements, that is, their action effects (see also Janczyk, 2016; Janczyk, Durst, & Ulrich, 2017; Kunde, 2001). Importantly, the possible action effects (depressed left/right response keys, visual and proprioceptive feedback from moving a left/right finger) are the same in forced and free choice tasks. A difference, however, is that for forced choice tasks the stimulus determines the desired action effect, while in free choice tasks this state must be generated again endogenously. Either way, such processes take time and would therefore be compatible with the results of this study.

The present data do not allow to distinguish between these two possibilities, and we do not claim that our list is exhaustive. For example, it is also possible that motor execution takes longer in free choice compared to forced choice tasks. Furthermore, these accounts are not mutually exclusive. It may well be that the choice of an effect is first driven by response history and then the action effect is endogenously generated, and thus both processes contribute to the non-accumulation time.

***Modeling free choice data.*** The present results can be used as constraints for future formal models of free choice behavior that assess their parameters more directly. We are currently aware of only one direct application of a sequential sampling model to data from priming experiments in free and forced choice tasks (Mattler & Palmer, 2012). The most important outcome of the experiments in this study was that the response in the free choice task was biased by a (subliminal) stimulus-preceding prime. In their model, the activity of two accumulator nodes mutually inhibit the response unit of the other accumulator. This inhibition may account for potential response-response conflicts in free choices, which slow down responses (Berlyne, 1957). Furthermore, when a free choice stimulus appears, an exponential drop of the decision thresholds is assumed. While this model fits the priming data, the threshold drop can generally be seen critically, because it is not assumed for the forced choice task because it is assumed to start immediately after the appearance of the stimulus, implying some sort of stimulus identification. Nevertheless, we believe that the approach by Mattler and Palmer (2012) is a valuable step towards the identification of similarities and differences between forced and free choice tasks.

**Conclusion**. Applying a framework borrowed from diffusion models, we observed no evidence that the mean RT difference between free and forced choice tasks is attributable to a higher drift rate in forced than in free choice tasks. Our results are rather compatible with a delay of the information accumulation process in free compared to forced choice tasks. Future work should aim at identifying the nature of this delay and the concurrent processes in more detail.

## References

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research, 79*(5), 882-898. doi: 10.1007/s00426-014-0608-y

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*(8), 1369–1378. http://doi.org/10.3758/BF03206327

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*(4), 428–454. http://doi.org/10.1016/0196-8858(91)90029-I

Bausenhart, K. M., Rolke, B., Seibold, V. C., & Ulrich, R. (2010). Temporal preparation influences the dynamics of information processing: Evidence for early onset of information accumulation. *Vision Research, 50*(11)*,* 1025–1034. http://doi.org/10.1016/j.visres.2010.03.011

Berlyne, D. E. (1957). Conflict and choice time. *British Journal of Psychology*, *48*(2), 106–118. http://doi.org/10.1111/j.2044-8295.1957.tb00606.x

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist*, *14*(4), 319–325. http://doi.org/10.1177/1073858408317417

Brysbaert, M. (1994). Behavioral estimates of interhemispheric transmission time and the signal detection method: A reappraisal. *Perception & Psychophysics, 56*(4)*,* 479–490. http://doi.org/10.3758/BF03206739

Dambacher, M., & Hübner, R. (2015). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological Research*, *79*(1), 83-94.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41*(3), 260–274. http://doi.org/10.1006/jmps.1997.1167

Dror, I. E., Basola, B., & Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & Cognition, 27*(4)*,* 713–725. http://doi.org/10.3758/BF03211564

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Cramon, D. Y. von, Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, 105*(45), 17538–17542. http://doi.org/10.1073/pnas.0805903105

Gaschler, R., & Nattkemper, D. (2012). Instructed task demands and utilization of action effect anticipation. *Frontiers in Cognition*, *3*, 578. http://doi.org/10.3389/fpsyg.2012.00578

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493-503.

Gordon, I. E. (1967). Stimulus probability and simple reaction time. *Nature, 215*(5103), 895–896. http://doi.org/10.1038/215895a0

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review, 77*(2), 73–99. http://doi.org/10.1037/h0028689

Grice, R. (1968). Stimulus intensity and response evocation. *Psychological Review*, *75*(5), 359–373. http://doi.org/10.1037/h0026287

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982). Human reaction time: Toward a general theory. *Journal of Experimental Psychology: General, 111*(1), 135–153. http://doi.org/10.1037/0096-3445.111.1.135

Hanoch, Y., Wood, S., Barnes, A., Liu, P.-J., & Rice, T. (2011). Choosing the right medicare prescription drug plan: The effect of age, strategy selection, and choice set size. *Health Psychology*, *30*(6), 719–727. http://doi.org/10.1037/a0023951

Harleß, E. (1861). Der Apparat des Willens. *Zeitschrift für Philosophie und philosophische Kritik*, *38*, 50-73.

Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1540–1554. http://doi.org/10.1080/17470210601119134

Herwig, A., & Waszak, F. (2009). Intention and attention in ideomotor learning. *The Quarterly Journal of Experimental Psychology*, *62*(2), 219–227. http://doi.org/10.1080/17470210802373290

Herwig, A., & Waszak, F. (2012). Action-effect bindings and ideomotor learning in intention- and stimulus-based actions. *Frontiers in Psychology*, *3,* 444. http://doi.org/10.3389/fpsyg.2012.00444

Heuer, H., Janczyk, M., & Kunde, W. (2010). Random noun generation in younger and older adults. *The Quarterly Journal of Experimental Psychology, 63*(3), 465-478. http://doi.org10.1080/17470210902974138:

Janczyk, M. (2013). Level-2 perspective taking entails two processes: Evidence from PRP experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1878-1887. http://doi.org/10.1037/a0033336

Janczyk, M. (2016). Die Rolle von Handlungszielen bei der Entstehung von Doppelaufgabenkosten. *Psychologische Rundschau, 67,* 237-249. http://doi.org/10.1026/0033-3042/a000324

Janczyk, M. (2017). A common capacity limitation for response and item selection in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* http://doi.org/10.1037/xlm0000408

Janczyk, M., Dambacher, M., Bieleke, M., & Gollwitzer, P. M. (2015). The benefit of no choice: goal-directed plans enhance perceptual processing. *Psychological Research*, *79*(2), 206–220. http://doi.org/10.1007/s00426-014-0549-5

Janczyk, M., Durst, M., & Ulrich, R. (2017). Action selection by temporally distal goal states. *Psychonomic Bulletin & Review, 24,* 467-473. http://doi.org/10.3758/s13423-016-1096-4

Janczyk, M., Nolden, S., & Jolicoeur, P. (2015). No differences in dual-task costs between forced- and free-choice tasks. *Psychological Research*, *79*(3), 463–477. http://doi.org/10.1007/s00426-014-0580-6

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General*, *141*(3), 489–501. http://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*(1), 30–43. http://doi.org/10.1016/j.cognition.2014.03.001

Janczyk, M., Pfister, R., & Kunde, W. (2012). On the persistence of tool-based compatibility effects. *Journal of Psychology*, *220*(1), 16–22. http://doi.org/10.1027/2151-2604/a000086

Janczyk, M., Skirde, S., Weigelt, M., & Kunde, W. (2009). Visual and tactile action effects determine bimanual coordination performance. *Human Movement Science*, *28*(4), 437–449. http://doi.org/10.1016/j.humov.2009.02.006

Keller, P. E., Wascher, E., Prinz, W., Waszak, F., Koch, I., & Rosenbaum, D. A. (2006). Differences between intention-based and stimulus-based actions. *Journal of Psychophysiology*, *20*(1), 9-20. https://doi.org/10.1027/0269-8803.20.1.9

Kleinsorge, T. (1999). Response repetition benefits and costs. *Acta Psychologica*, *103*(3), 295–310. https://doi.org/10.1016/S0001-6918(99)00047-5

Kühn, S., Elsner, B., Prinz, W., & Brass, M. (2009). Busy doing nothing: Evidence for nonaction-effect binding. *Psychonomic Bulletin & Review*, *16*(3), 542–549. http://doi.org/10.3758/PBR.16.3.542

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 387–394. http://doi.org/10.1037/0096-1523.27.2.387

Kunde, W., Pfister, R., & Janczyk, M. (2012). The locus of tool-transformation costs. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 703–714. http://doi.org/10.1037/a0026315

Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition*, *123*(3), 347–360. http://doi.org/10.1016/j.cognition.2012.03.002

Merkel, J. (1885). Die zeitlichen Verhältnisse der Willensthätigkeit. *Philosophische Studien*, *2*, 73–127.

Miller, J., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1126-1142. http://doi.org/10.1037/0096-1523.29.6.1126

Näätänen, R. (1972). Time uncertainty and occurence uncertainty of the stimulus in a simple reaction time task. *Acta Psychologica*, *36*(6), 492–503. http://doi.org/10.1016/0001-6918(72)90029-7

Passingham, R. E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: From self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14*(1), 16–21. http://doi.org/10.1016/j.tics.2009.11.001

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74–80. http://doi.org/10.2478/v10053-008-0133-x

Pfister, R., Kiesel, A., & Hoffmann, J. (2011). Learning at any rate: Action–effect learning for stimulus-based actions. *Psychological Research*, *75*(1), 61–65. http://doi.org/10.1007/s00426-010-0288-1

Pfister, R., Kiesel, A., & Melcher, T. (2010). Adaptive control of ideomotor effect anticipations. *Acta Psychologica*, *135*(3), 316–322. http://doi.org/10.1016/j.actpsy.2010.08.006

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The Hare and the Tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226-1243. http://dx.doi.org/10.1037/a0036801

Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2010). *G*Power-Ergänzungen*. Retrieved from http://quantitative-methoden.de/Dateien/Auflage3/Band_II/Kapitel_7_GPower_Ergaenzungen_A3.pdf

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. http://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. http://doi.org/10.1037/0033-295X.85.2.59

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*(2), 261-282. http://doi.org/10.1037/0096-3445.133.2.261

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*(2), 261-282. http://doi.org/10.1037/0096-3445.133.2.261

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology*, *18*(2), 105-139. https://doi.org/10.1016/0022-2496(78)90059-7

Seibold, V. C., Bausenhart, K. M., Rolke, B., & Ulrich, R. (2011). Does temporal preparation increase the rate of sensory information accumulation? *Acta Psychologica*, *137*(1), 56–64. http://doi.org/10.1016/j.actpsy.2011.02.006

Shin, Y. K., Proctor, R. W., & Capaldi, EJ. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136*(6), 943–974. http://doi.org/10.1037/a0020541

Stock, A., & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research, 68*(2–3), 176–188. http://doi.org/10.1007/s00426-003-0154-5

Wagenmakers, E. J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3-22. http://doi.org/10.3758/BF03194023

Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, *162*(3), 346-356. http://doi.org/10.1007/s00221-004-2183-8

Wolfensteller, U., & Ruge, H. (2011). On the timescale of stimulus-based action–effect learning. *The Quarterly Journal of Experimental Psychology*, *64*(7), 1273–1289. http://doi.org/10.1080/17470218.2010.546417

## Acknowledgements

Work of MJ is supported by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft/German Research Foundation ZUK 63). Further, this work was supported by Grant JA2307/1-2 from the Deutsche Forschungsgemeinschaft/German Research Council awarded to MJ. We thank Dirk Vorberg and an anonymous reviewer for very helpful and constructive comments on a previous version of this manuscript.

## Appendix

In this appendix we report the results from a diffusion model analysis on the forced choice data from Experiments 1-3. EZ (Wagenmakers, Van Der Maas, & Grasman, 2007) was used to extract parameters for every participant and relevant experimental condition (i.e., excluding the 0% catch-trial condition of Experiment 1 and excluding the pre-experimental blocks of Experiments 2 and 3) for each experiment. Tables A1-A3 summarize the resulting means and standard deviations for drift rates, response thresholds, and non-accumulation times. The parameters were submitted to ANOVAs with block type as a repeated measures factor. Greenhouse-Geisser corrected *p*-values are reported when the sphericity assumption was violated (in this case the respective ε is reported as well). In case a participant made no mistakes in a given condition, the edge correction proposed by Wagenmakers et al. (2007) was performed, in which, essentially, the sum of errors is changed from zero errors to half an error.

**Table A1.** Extracted parameter means, standard deviations in parentheses per experimental condition for Experiment 1. In nine cases, edge corrections were applied because there were no errors in one or more conditions. *n* = 32.

| | Response thresholds | Drift rates | Non-accumulation times |
|---|---|---|---|
| Block type | F(2,62) = 2.40, p = .109, $\eta p^2 = .07$, ε = .846 | F(2,62) = 9.90, p < .001, $\eta p^2 = .24$ | F(2,62) = 33.21, p < .001, $\eta p^2 = .52$, ε = .811 |
| 25% catch-trials | 3.45 (0.69) | 0.0097 (0.0024) | 294 (47) |
| 50% catch-trials | 3.71 (0.85) | 0.0092 (0.0024) | 300 (56) |
| 75% catch-trials | 3.47 (0.90) | 0.0081 (0.0027) | 352 (51) |

**Table A2.** Extracted parameter means, standard deviations in parentheses per experimental condition for Experiment 2. $n = 36$.

| | Response thresholds | Drift rates | Non-accumulation times |
|---|---|---|---|
| Block type | F(2,70) = 123.37, $p < .001$, $\eta p^2 = .78$, $\varepsilon = .755$ | F(2,70) = 62.90, $p < .001$, $\eta p^2 = .64$, $\varepsilon = .660$ | F(2,70) = 146.79, $p < .001$, $\eta p^2 = .81$, $\varepsilon = .760$ |
| $M-SD$ | 1.67 (0.20) | 0.0057 (0.0035) | 172 (50) |
| $M$ | 1.94 (0.22) | 0.0096 (0.0024) | 238 (38) |
| $M+SD$ | 2.25 (0.27) | 0.0100 (0.0021) | 253 (30) |

**Table A3.** Extracted parameter means, standard deviations in parentheses per experimental condition for Experiment 3. $n = 36$.

| | Response thresholds | Drift rates | Non-accumulation times |
|---|---|---|---|
| Block type | F(2,70) = 73.10, $p < .001$, $\eta p^2 = .68$ | F(2,70) = 2.20, $p = .118$, $\eta p^2 = .06$ | F(2,70) = 24.46, $p < .001$, $\eta p^2 = .41$ |
| $M$ | 1.83 (0.25) | 0.0091 (0.0030) | 231 (32) |
| $M+0.5*SD$ | 1.98 (0.33) | 0.0097 (0.0027) | 245 (33) |
| $M+1.5*SD$ | 2.15 (0.30) | 0.0095 (0.0023) | 252 (35) |

# Appendix C – Study 3

The following represents the article described in Chapter 7. Springer Nature granted the license to reproduce the final author's accepted manuscript here.

Naefgen, C., & Janczyk, M. (2018a). Free choice tasks as random generation tasks: an investigation through working memory manipulations. *Experimental Brain Research*, *236*(8), 2263–2275. doi:10.1007/s00221-018-5295-2

The published version can be found under: https://link.springer.com/article/10.1007/s00221-018-5295-2

RUNNING HEAD: Working Memory and Free Choice

# Free Choice Tasks as Random Generation Tasks: An Investigation through Working Memory Manipulations

Christoph Naefgen & Markus Janczyk

Department of Psychology

Eberhard Karls University of Tübingen, Tübingen, Germany

**Address correspondence to:**
Christoph Naefgen
Eberhard Karls University of Tübingen
Department of Psychology
Schleichstraße 4
72076 Tübingen
Germany
Phone: +49 (0)7071 29 76769
Email: christoph.naefgen@uni-tuebingen.de

## 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 as well. These result suggest that free choice tasks have much in common with random generation tasks.

**Keywords:** free choice ; forced choice ; action selection ; working memory ; random generation

## 1. 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[10], 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.

**1.1 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 Section 1.2, 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, Janczyk, & Kunde 2018 for an exception). This RT difference might be taken to indicate qualitative differences with regard to response

---

[10] 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).

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) actions (e.g., Brass & Haggard 2008; Herwig, Prinz, & Waszak 2007; Passingham, Bengtsson, & Lau 2010; Keller, Wascher, Prinz, Waszak, Koch, & Rosenbaum 2006; Waszak, Wascher, Keller, Koch, Aschersleben, Rosenbaum, & Prinz 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 & Nattkemper 2012; Herwig & Waszak 2009, 2012; Pfister, Kiesel, & Melcher 2010). However, Pfister, Kiesel, and Hoffmann (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, Huffman, & Pratt 2016; Huffman, Gozli, Hommel & Pratt 2018; Janczyk, Durst, & Ulrich 2017; Janczyk, Pfister, Crognale, & Kunde 2012; Janczyk, Pfister, Hommel, & Kunde 2014; Janczyk, Pfister, & Kunde 2012; Kühn, Elsner, Prinz, & Brass 2009, Exp. 3; Kunde 2001; Kunde, Pfister, & Janczyk 2012; Pfister & Kunde 2013; Wolfensteller & 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, Nolden, and Jolicoeur (2015) compared both task types with regards 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, Dambacher, Bieleke, & Gollwitzer 2015). 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, Dambacher, and Janczyk (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, Rolke, Seibold, & Ulrich 2010) and time pressure (e.g., Dror, Basola, & Busemeyer 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.

**1.2 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 & Wegner 2011) and on neuroimaging evidence that random choice tasks and free choice tasks activate similar brain regions (Jahanshahi, Dirnberger, Fuller, & Frith 2000; Jenkins, Jahanshahi, Jueptner, Passingham, & Brooks 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[11]), 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 between random generation tasks and how they are affected by working memory manipulations.

---

[11] 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 4294967296 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, Rushworth, Passingham, Jahanshahi, & Rothwell 2001, see also Waszak et al. 2005; Elsner & Hommel 2001) |
| Similar to random response instructions | Overlap between instructions and definitions of randomness<br>**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 & Kunde 2013, see also Linser & Goschke 2007) |
| Emphasizing spontaneity/freedom of choice | No mention of randomness<br>No particular overlap in instructions with definitions of randomness<br>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, Kiesel, & Melcher 2010; see also Herwig, Prinz, & Waszak 2007) |
| None reported | No explication of instructions present in the text | “When lights of both colours appeared, either response, but not both, was to be performed.” (Berlyne 1957) |
| 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, Jokic, Linden, Marlier, & Bussel 1996) |

**1.3 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 Baddeley 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, Wutke, and Davelaar (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, Friedman, Emerson, Witzki, and Howerter (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.

**1.4 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 & 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, Janczyk & Kunde 2010; Heuer, Kohlisch & Klein 2005). 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 divided into all possible sequential sub-sequences of a predefined length and the formula is applied to all of these sub-sequences. For an illustration of what this looks like, see Figure 1. The formula for the LU in each segment is as follows:

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

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:

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

The range of values for the LU lies between 0 and .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 .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, Figure 1 gives an example sequence of choices and the resulting LUs, for four different window sizes of 2, 4, 6, and 8 for each window, as well as the mean LU for the sequence.

Window Size 2

```
LRLLRRLRRRLL
LR            0
 RL           0
  LL         .5
   LR         0
    RR       .5
     RL       0
      LR      0
       RR    .5
        RR   .5
         RL   0
          LL .5 Mean .227
```

Window Size 4

```
LRLLRRLRRRLL
LRLL         .25
 RLLR         0
  LLRR        0
   LRRL       0
    RRLR     .25
     RLRR    .25
      LRRR   .25
       RRRL  .25
        RRLL  0 Mean .156
```

Window Size 6

```
LRLLRRLRRRLL
LRLLRR        0
 RLLRRL       0
  LLRRLR      0
   LRRLRR    .166
    RRLRRR   .333
     RLRRRL  .166
      LRRRLL  0 Mean .095
```

Window Size 8

```
LRLLRRLRRRLL
LRLLRRLR      0
 RLLRRLRR    .125
  LLRRLRRR   .125
   LRRLRRRL  .125
    RRLRRRLL .125 Mean .1
```

*Figure 1*. Examples of (average) LUs in an example sequence for window sizes 2, 4, 6, and 8.

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 e.g. 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.

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)

*Figure 2*. All sequences that can occur for window size 4 and the respective LU. The resulting ideal LU value is then .1875.

This results in an ideal LU of .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 .25, .1875, .15625, and .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.

**1.6 The present study.** Our prediction is that, if free choice tasks are random generation tasks, then 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 an 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 tradeoffs, where faster responses come with committing more errors). Thus, additionally analyzing RTs makes it possible to rule out such phenomena.

## 2. 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, Rushworth, Passingham, Jahanshahi, & Rothwell 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.

## 2.1 Methods

***2.1.1 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.

***2.1.2 Apparatus and stimuli.*** Stimulus presentation and response collection happened on a PC connected to a 17-inch 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.

***2.1.3 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 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 Figure 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.

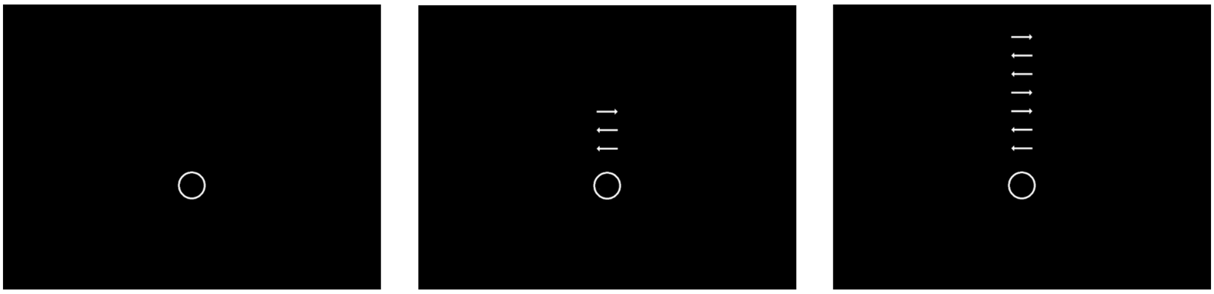

*Figure 3*. Examples of the different WM support conditions. In the left panel, no WM support is given, in the middle panel three previous choices are displayed, and in the right panel seven previous choices are displayed. Not visible here are the arrows that appear within the circle for 50 ms after a response.

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 minutes.

***2.1.4 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.

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 *SD*s from the respective cell mean (calculated separately for each participant).

### 2.3. Results

The LUDs and average RTs (1.79% outliers) are visualized in Figure 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 *p*s ≥ .217 (for more details, please see Appendix).

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: $t(29) = -0.14$, $p = .886$.

**2.4. 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 condition. 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 & Lerche 2018; Janczyk, Durst & Ulrich 2017; Koch & Kunde 2014). 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.[12]

---

[12] 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 others results, the experiment is not reported here.

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

<table>
<tr><td></td><td colspan="5">WM support: number of displayed choices</td></tr>
<tr><td>Dependent variable</td><td>0</td><td>3</td><td colspan="3">7</td></tr>
<tr><td>LUDs</td><td>.045 (.220)</td><td>.027 (.213)</td><td colspan="3">.030 (.215)</td></tr>
<tr><td colspan="3">RTs</td><td>481 (289)</td><td>410 (208)</td><td>414 (266)</td></tr>
</table>

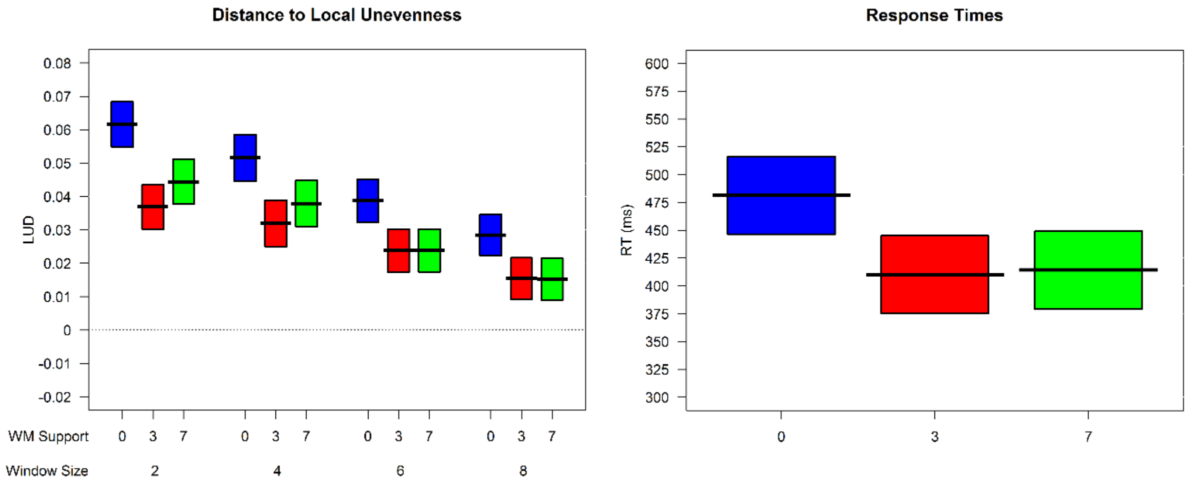

*Figure 4.* Mean LUDs (for the window sizes 2, 4, 6, and 8) and RTs in Experiment 1 for each level of WM support. Error bars are 95% within confidence intervals (separate for all window sizes in case of the LUDs) (Loftus & Masson 1994).

## 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 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.

### 3.1 Methods

***3.1.1 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.

***3.1.2 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).

***3.1.3 Tasks and procedure.*** Participants performed two tasks in alternation (for an illustration, see Figure 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 four times, with the task either requiring attention to the color or the location of the stimulus. A participant performed all 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.

Participants performed in 16 blocks 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 are illustrated in Figure 5.

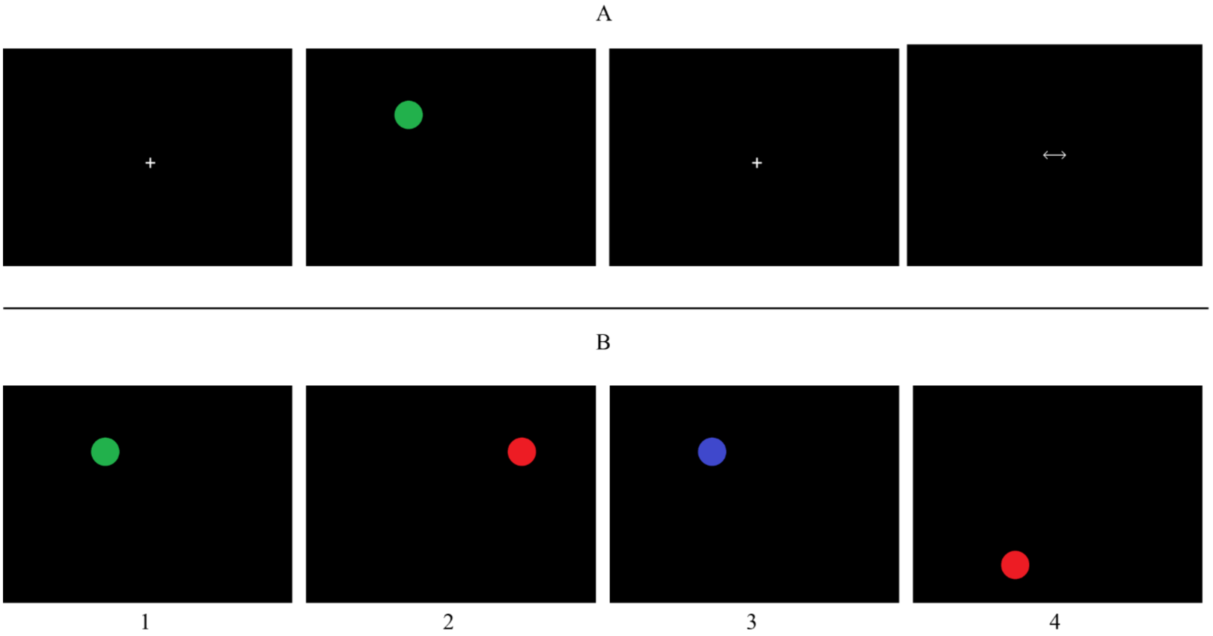

*Figure 5.* (A) Example of a sequence of displayed stimuli and fixation crosses on the screen. (B) Example of a sequence of *n*-back stimuli (free choice stimuli not displayed). In the color-based 2-back condition, only panel (4) would require a response, while in the location-based 2-back condition, panel (3) would require a response.

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 minutes. 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).

***3.1.4 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 size (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 the window size and the Helmert contrast, separate Helmert contrasts for each window size were calculated and are reported in the Appendix.

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 *SD*s from the respective cell mean (calculated separately for each participant).

### 3.3. 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, $F(1, 31) = 3.94$, $p = .056$, $\eta_p^2 = .11$, and the three-way interaction WM load × window size × stimulus feature was also not significant, $F(9, 279) = 0.96$, $p = .421$, $\eta_p^2 = .03$, $\varepsilon = .36$ (Greenhouse-Geisser estimate). With RTs as the dependent variable, also no significant main effect was observed, $F(1, 31) = 3.75$, $p = .062$, $\eta_p^2 = .11$, and the interaction WM load × stimulus feature was also not significant, $F(3, 93) = 2.74$,

$p = .063$, $\eta_p^2 = .08$, $\varepsilon = .78$. To simplify the main analyses, we thus dropped this variable from further analyses.

*3.3.1 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.

*3.3.2. Free choice task.* Average LUDs and RTs (2.81% outliers) for the free choice task are visualized in Figure 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 Appendix). 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.

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

*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 | .025 (.197) | .027 (.199) | .048 (.206) | .069 (.204) |
| | RT | 340 (94) | 329 (93) | 358 (98) | 386 (118) |

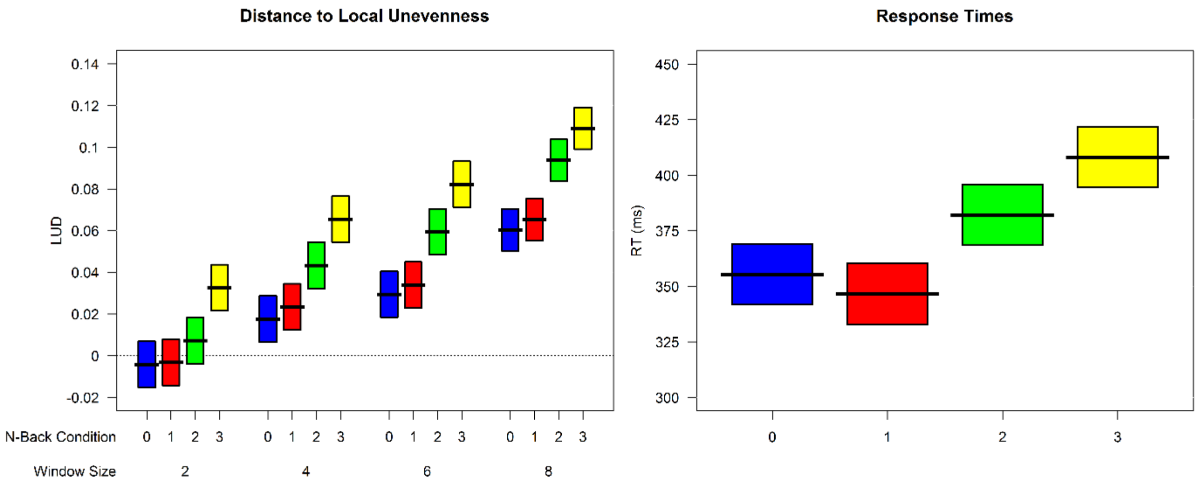

*Figure 6.* Mean LUDs (for the window sizes 2, 4, 6, and 8) and RTs in the free choice task in Experiment 2 for each level of WM load. Error bars are 95% within confidence intervals (separate for all window sizes in case of the LUDs) (Loftus & Masson 1994).

## 4. 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.

### 4.1. 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 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).

### 4.2. 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 random-ness 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.[13] This assumption seems not be universal among researchers though. Passingham, Bengtsson, and Lau (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.

The present study may also speak to the results reported in Naefgen, Dambacher et al. (2018). There, in Experiments 2 and 3, a time pressure manipulation was used to induce changes in threshold separation in a sequential sampling framework. However, in Experiment 2, this manipulation also affected drift rate, in addition to the (intended) effect on threshold separation. Thus, given that time pressure is known to affect random generation tasks and reduce randomness (e.g., Baddeley 1966), the effects in drift rate may in fact not be an issue solely of parameter estimation but rather reflect differences in the random generation process. The same might apply to Experiment 1 of Naefgen, Dambacher et al., but it is less clear how the frequency of catch-trials would affect random generation tasks.

### 4.3 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-

---

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

L: all result in LU = .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 regards 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 & 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).

### 4.4. Conclusion

We investigated whether LU, as a measure of randomness, based on responses from free choice tasks and RTs in this task are 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.

## Acknowledgements

This research was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), grant JA 2307/1-2 awarded to Markus Janczyk. Work of MJ is further supported by the Institutional Strategy of the University of Tübingen (DFG ZUK 63). We thank Davood Gozli for helpful comments on a previous version of this manuscript. In addition, Cosima Schneider and Moritz Durst provided valuable feedback that improved this manuscript.

## Compliance with ethical standards

### 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.

### Conflicts of interest

The authors declare that they have no conflict of interest.

## 5. 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. doi: 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. doi: 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. Vision Res 50:1025–1034. doi: 10.1016/j.visres.2010.03.011

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

Brass M, Haggard P (2008) The what, when, whether model of intentional action. The Neuroscientist 14:319–325. doi: 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

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

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

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

Gaschler R, Nattkemper D (2012) Instructed task demands and utilization of action effect anticipation. Front Psychol 3:. doi: 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. doi: 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. doi: 10.1162/089892901753294392

Henmon VAC (1911) The relation of the time of a judgment to its accuracy. Psychol Rev 18:186–201. doi: 10.1037/h0074579

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. doi: 10.1080/17470210601119134

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

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

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

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. doi: 10.1080/02724980343000855

Huffman G, Gozli DG, Hommel B, Pratt J (2018) Response preparation, response selection difficulty, and response-outcome learning. Psychol Res 1–11. doi: 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. doi: 10.1006/nimg.2000.0647

Janczyk M, Dambacher M, Bieleke M, Gollwitzer PM (2015) The benefit of no choice: goal-directed plans enhance perceptual processing. Psychol Res 79:206–220. doi: 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. doi: 10.3758/s13423-016-1096-4

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

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

Janczyk M, Pfister R, Crognale MA, Kunde W (2012) Effective rotations: Action effects determine the interplay of mental and manual rotations. J Exp Psychol Gen 141:489–501. doi: 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. doi: 10.1016/j.cognition.2014.03.001

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

Janczyk M, Skirde S, Weigelt M, Kunde W (2009) Visual and tactile action effects determine bimanual coordination performance. Hum Mov Sci 28:437–449. doi: 10.1016/j.humov.2009.02.006

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. doi: 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. doi: 10.1037/h0043688

Kühn S, Elsner B, Prinz W, Brass M (2009) Busy doing nothing: Evidence for nonaction-effect binding. Psychon Bull Rev 16:542–549. doi: 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. doi: 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. doi: 10.1037/a0026315

Linser K, Goschke T (2007) Unconscious modulation of the conscious experience of voluntary control. Cognition 104:459–475. doi: 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. doi: 10.3758/BF03210951

Martin LJ, Müller GE (1899) Zur analyse der unterschiedsempfindlichkeit: Experimentelle beiträge. J. A. Barth

Martin-Löf P (1966) The definition of random sequences. Inf Control 9:602–619. doi: 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. doi: 10.1006/cogp.1999.0734

Ginsburg N, Karpiuk P (1994) Random generation: Analysis of the responses. Percept Mot Skills 79:1059–1067. doi: 10.2466/pms.1994.79.3.1059

Naefgen C, Dambacher M, Janczyk M (2017) Why free choices take longer than forced choices: evidence from response threshold manipulations. Psychol Res 1–14. doi: 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. doi: 10.1016/j.tics.2009.11.001

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

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

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

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. doi: 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. doi: 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. doi: 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. doi: 10.1037/xlm0000474

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

Woodworth RS (1899) Accuracy of voluntary movement. Psychol Rev Monogr Suppl 3:i-114. doi: 10.1037/h0092992

## 6. Appendix

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

**Experiment 1.** The descriptive results of the following analyses are summarized in Table A1.

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

*Table A1*. Means (and *SDs*) of LUDs for window sizes 2, 4, 6, and 8 for Experiment 1 for each WM support condition.

| | WM support: Number of displayed choices | | |
|---|---|---|---|
| Window size | 0 | 3 | 7 |
| 2 | .062 (.306) | .037 (.066) | .044 (.067) |
| 4 | .052 (.213) | .032 (.057) | .038 (.059) |
| 6 | .039 (.174) | .024 (.053) | .024 (.056) |
| 8 | .028 (.152) | .015 (.051) | .015 (.052) |

**Experiment 2.** The descriptive results of the following analyses are summarized in Table A2.

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: $t(31) = 2.81$, $p = .008$

*Table A2*. Means (and *SD*s) LUDs for window sizes 2, 4, 6, and 8 for Experiment 2 for each WM load condition.

| | *n*-back | | | |
|---|---|---|---|---|
| Window size | 0 | 1 | 2 | 3 |
| 2 | -.003 (.116) | -.004 (.123) | .007 (.131) | .031 (.123) |
| 4 | .018 (.124) | .023 (.126) | .043 (.135) | .064 (.128) |
| 6 | .030 (.127) | .034 (.130) | .059 (.138) | .081 (130) |
| 8 | .060 (.118) | .065 (.119) | .094 (.125) | .109 (.121) |