MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„ Effects of the Spatial Separation of Response Keys on the Reaction Times and Accuracy Rates of the Familiarity-Related Inducer Reaction Time-Based Concealed Information Test "

verfasst von / submitted by

Daniel Lauber BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge

Introduction

Over the course of the evolution of mankind, the special ability of attributing mental states to oneself and others has developed in humans. This special ability is called theory of mind (ToM) and is crucial in everyday social interactions because it allows one to reason about people's actions in terms of their mental states (Gweon & Sachse, 2013). The ability to deceive others came with the capacity of ToM and is developed in early childhood (Chandler et al., 1989). Humans are tempted to lie or to deceive others in order to either achieve future rewards or to avoid punishment. However, due to the circumstance that humans live in complex social structures where laws create order, unlawful behaviors and transgressions such as lying in court are not tolerated. Detecting deception has a long tradition. Historic records from ancient China show that as early as 1,000 BCE it was attempted to identify liars by their decreased saliva production. The reasoning being that if one were to spit out more rice, one would appear to be more nervous and thus more likely to be deceptive (Kleinmuntz & Szucko, 1984). Nowadays polygraph methods, which measure peripheral, physiological reactions associated with increased arousal due to anxiety or stress, like heart rate, skin conductance and breathing rate (Steinbrook, 1992) are widely used to operationalize deceptive behavior. Other methods to detect deception are methods which measure central correlates of nervous system activity such as Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI) (Langleben & Moriarty, 2013; Rosenfeld et al., 1987). Besides the aforementioned peripheral, physiological and neural measuring methods, behavioral methods such as measuring Reaction Times (RTs) have been argued to be more advantageous for reasons of economy, easiness of use and quickness to measure (Verschuere & De Houwer, 2011; Verschuere et al., 2014). The RT-CIT is a variant of the Concealed Information Test (CIT) which is used to assess recognition of critical information. In general terms, the CIT measures the behavioral reaction differences between the item of critical information and the control items. Someone who is innocent would respond similarly to all items. However, a guilty perpetrator would react differently to the critical item due to its subjective relevance (e.g., the object the crime was committed with) and hence inadvertently reveal critical crime details (Verschuere et al., 2014). This different reaction to the critical item could be explained by the cognitive approach to deception which states that lying is cognitively more challenging and demanding than telling the truth (Suchotzki et al., 2017). The cognitive challenge is expressed in longer RTs in the RT-CIT. During the RT-CIT the examinee is presented with the critical detail (probe), which is intermixed with equally plausible control items (irrelevants). In case of a murder scenario, the probe would be the object which the crime was committed with and the irrelevants would be other potential weapons. These items, which are collectively called *non-target items*, must be responded to utilizing the same response key. A randomly selected additional target which represents another crime-unrelated item, however, must be reported using an alternative key command. Guilty examinees would semantically categorize the probe as one rare item among the irrelevants and the target as the other rare item, but, according to the instructions, they would have to classify the probe as just another irrelevant item. The probe is unique among the non-targets and similar to the rarely appearing target (Lukács et al., 2016; Lukács, Gula, et al., 2017). This induces a response conflict in the examinees (Seymour & Schumacher, 2009) which can be measured as generally slower RTs to the probe than to irrelevant items. According to these RT differences, guilty examinees can be effectively differentiated by the response patterns from the innocent participants.

By adding familiarity-related inducer items to the RT-CIT, a considerable improvement can be observed regarding the validity of the test method (Lukács, Kleinberg, et al., 2017). Inducer items aim for the increase of response conflict due to the increased cognitive load and attention in discriminating between target and non-target items. Through the application of more specific inducer items, an increased RT-CIT accuracy of distinguishing guilty examinees from innocent ones was found. The application of familiarity-related inducer items like “familiar”, “recognized”, and “mine” enhances the semantic processing of both the non-target (probes and irrelevants) and target items. Hence it increases the attention to the relevant dimension of the CIT stimuli (i.e., the self-relevance of specific objects). The implicit semantic wiring of the brain and the verbal association networks (Greenwald et al., 1998) result in the ease of categorizing familiarity-related inducer items as self-relevant due to similar meaning contexts and their neuronal wiring (Hebb, 1949). This neuronal connection of the semantic associations activated through familiarity-related inducer items enhances the response conflict at answering a self-relevant probe item as unfamiliar and therefore elicits longer RTs in corresponding key presses in comparison to irrelevant items.

Another classical psychological experiment, the Stroop task (Stroop, 1935), has been widely used to investigate the underlying cognitive mechanisms in distinguishing task-relevant from task-irrelevant information. The Stroop task is a color-identification task, where participants identify the physical color (task-relevant) of a letter string (normally a color word), while actively ignoring the conveyed meaning by the string (task-irrelevant). Consequently, the Stroop effect encapsulates the latency of reaction induced by the dissonance between meaning and physical color of the letter string. As a result, a large Stroop effect indicates an insufficient performance in the discrimination of the incongruent word meaning and physical color (e.g., the word “green” written in red).

Originally, the Stroop effect describes the delay in reaction times between congruent and incongruent stimuli. The delay observed in an incongruent trial is due to the interference between automated reading on the one hand and being forced to determine the semantic meaning of the word on the other hand (e.g., the word “red” and so the color “red” is imagined) and the conscious perception of the actual color of the word (e.g., the word “red” written with black ink). Van Maanen et al. (2009) found that semantic interference is the phenomenon when the naming of an ink color of a neutral stimulus is faster than in incongruent conditions. Here, the root of interference is the relationship in meaning between ink color and word. Secondly, that semantic facilitation is observable when the naming the ink of a congruent stimulus is faster than when a neutral stimulus is presented (like the stimulus three in the original experiment by Stroop (1935) where only a colored square is presented). They found thirdly, that both semantic interference and facilitation disappear when only the word is to be read instead of naming the ink color (van Maanen et al., 2009). The underlying cognitive principles of the Stroop effect are based on the notion that both relevant and irrelevant information within the task are processed in parallel and therefore compete for the attention of the single central processing unit during response selection (Johnson & Proctor, 2004).

Lakens et al. (2011) demonstrated that the Stroop effect varies depending on the physical distance of the response hands using a two-choice Stroop task where the colors red and blue were assigned to two distinct keypresses. The Stroop effect was larger when the response hands were physically closer compared to being farther apart from each other. Lakens et al. (2011) argued that this response-distance effect is to be interpreted as being in line with anatomical discriminability. According to this concept, the distance between one's hands can help to tell things apart in the mind. Subsequently to their conclusion, this physical discriminability effect could be measured in other investigations using reactions on distinct keypresses, too.

In this study, we investigated another potential way to enhance the sensitivity of the RT-CIT. As reasoned by Lakens et al. (2011), the widened distance of response keys may enhance the cognitive categorization of distinct response options. This finding should be conceptually replicated in this experiment using the RT-CIT, which would in turn presumably result in larger accuracy rates and smaller probe-to-irrelevant RT effects in the wide key condition compared to the narrow key distance condition, which is expected to have larger probe-to-irrelevant RT differences. Particularly at wider key placement, analogous to the notion of "telling things apart" mentioned by Lakens et al. (2011), discriminating of specific responses may be facilitated resulting in smaller probe-to-irrelevant differences. While narrower key placement may impede telling things apart from each other, thus resulting in larger probe-to-irrelevant differences, wider key placement may support telling things apart.

Methods

Participants. Recruiting of participants for this experiment was done using the Laboratory Administration for Behavioral Sciences (LABS)-platform of the University of Vienna, where students of the Faculty of Psychology are registered to take part at different experiments. For the participation students are granted credits, the amount of which depends on the duration of the concrete assignment. This experiment was set to take approximately 40 min which translated to 3 laboratory-credits.

Each participant was randomly assigned to one of the four conditions. Participants in the first and third condition group started under the wide key condition, whereas subjects in the second and fourth conditions began under a narrow key condition. The first and second conditions started with categorizing the forenames first, whereas the third and fourth conditions started with surnames. Once assigned to one of the key distance conditions, participants completed both surname and forename categorizations before they automatically switched to the opposite key distance condition.

Data exclusions. We will exclude all participants with an accuracy rate for the main items (probes and irrelevants merged) below 75% and accuracy rate for target items below 50%. For all further analysis, responses below 150 ms are excluded. For RT analysis, only correctly answered trials will be used (i.e., we exclude incorrect trials and trials deemed too slow, i.e., those with an RT above 800 ms).

We excluded participant 20 due to too much errors. Thus, the final sample size consisted of 59 participants (age $M = 20.9$ years, $SD = 2.2$; 10 male), who had been randomly assigned to one of the four groups (C1, C2, C3 and C4): 15 participants in the *start wide with forename condition group C1* (age $M = 20.5$ years, $SD = 1.8$; 2 male), 15 in the *start narrow with forename condition group C2* (age $M = 20.6$ years, $SD = 2.1$; 2 male), 15 in the *start wide with surname condition group C3* (age $M = 21.6$ years, $SD = 2.3$; 2 male) and 14 in the *start narrow with surname condition group C4* (age $M = 20.9$ years, $SD = 2.4$; 4 male).

For the main outcomes, the dependent variable was the probe-to-irrelevant RT mean (i.e., probe RT mean minus irrelevant RT mean, per each participant), which was compared between the wide and narrow key distance conditions with two statistical tests: (1) a simple *t*-test to test the potential difference, (2) Bayesian likelihood ratio to test the null hypothesis.

While in the RT-CIT usually only RT means are used as predictors (for guilty-innocent classifications), certain extents of probe-to-irrelevant differences are also often observed in accuracy rates, and therefore may be of specific interest (in particular, see Lukács & Ansorge, 2019; Lukács, Kleinberg, et al., 2017). Consequently, the tests above were repeated with probe-to-irrelevant accuracy rate differences (i.e., probe accuracy rate minus irrelevant accuracy rate, per each participant), in place of RT means, as dependent variables.

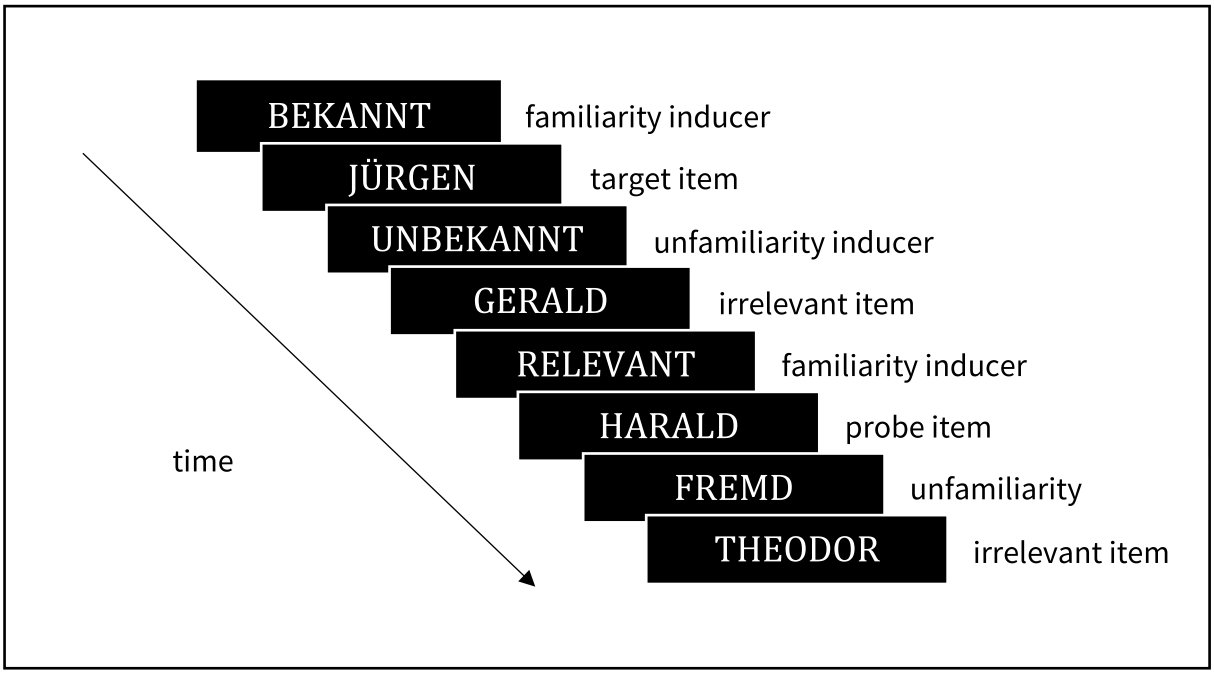
Effect sizes. We report Welch-corrected t-tests (Delacre et al., 2017), with corresponding Cohen´s *d* values as standardized mean differences and their CIs (Lakens, 2013). We used the conventional alpha level of .05 for all statistical significance tests.

Bayesian analysis. We report Bayes factors using the defaults *r*-scale of .707 (Morey & Rouder, 2018; Rouder et a., 2009). The Bayes factor is a ratio between the likelihood of the data fitting under the null hypothesis and the likelihood of fitting under the alternative hypothesis (Jarosz & Wiley, 2014; Wagenmakers, 2007). For example, a Bayes factor (BF) of 3 means that the obtained data is three times as likely to be observed if the alternative hypothesis is true, while a BF of 0.5 means that the obtained data is twice as likely to be observed if the null hypothesis is true. Here, for more readily interpretable numbers, we denote Bayesian factors as BF10 for supporting the alternative hypothesis, and as BF01 for supporting the null hypothesis. Thus, for example, BF01 = 2 means that the obtained data is twice as likely under the null hypothesis than under the alternative hypothesis. Typically, BF = 3 is interpreted as the minimum likelihood ratio for a substantial evidence for either the null or the alternative hypothesis (Jeffreys, 1961; Wagenmakers et al., 2011).

Procedure. The experiment was written by Gáspár Lukács in Python and conducted with the PsychoPy3 (Peirce et al., 2019) application on a personal computer (PC). Before the RT-CIT was conducted, participants gave their informed consent to start the task on the PC. The informed consent included information about the aim of the present study, namely that the RT-CIT simulates a lie detection scenario where they should hide their true identities. Initially, they filled out a sheet containing their age, their left- or right-handedness, state of origin, forenames, and surnames. After that, they had to confirm if their full name was written correctly. At this sheet, they were also assigned randomly to one of the four groups. After the participants provided their demographic information, they had to choose two names which were personally meaningful to them from a list containing names not identical to their own names. Participants had to choose one forename and surname each from the list and afterwards obtained two names which were randomly selected from the non-chosen items to assure that the two names were truly irrelevant for the participants. Subsequently, they were tested if they had internalized them correctly. If they could not remember them, they had to repeat this step of the procedure until they memorized the two items correctly. In the next, step the participants entered the first of the three practice phases. The first practice task did not have a time limit to answer the question if the item was familiar or relevant to them. In the RT-CIT, the items were presented successively in the center of the screen and participants had to press one of two distinct keys on the keyboard. These keys varied depending on the condition participants were assigned to. In the wide condition, participants had to press the number “5” on the number pad to say that the displayed item is “FAMILIAR” to them. On the contrary, if they pressed “S” with their left hand’s index finger, they responded that the displayed item was “UNFAMILIAR” to them. In the narrow condition, the key “L” meant that the displayed item was “FAMILIAR”, and the key “K” meant that the displayed item was “UNFAMILIAR” to them. The two obtained names functioning as the target items should be categorized as “FAMILIAR” as well as the filler items (interchangeable with the word inducer item) referring to familiarity (“FAMILIAR”, “MINE” and “RELEVANT”). Whereas the probes which are actually the real names of the participants and the other irrelevant names together with the filler items referring to unfamiliarity (“UNFAMILIAR”, “FOREIGN”, “UNKNOWN”, “OTHER”, “ELSE” and “IRRELEVANT”) had to be categorized as “UNFAMILIAR”, as shown schematically in Figure 1.

Figure 1

Successive presenting protocol of probe, target, irrelevant and inducer items



In the first practice task, these items (probe, target and irrelevant items) were presented intermixed with the filler items and participants had no time pressure to respond to them correctly. If they made a mistake in categorizing the items, the word “Inkorrekt!”, the German word for *incorrect*, appeared below the stimulus in red color for 500 ms. In case of too much wrong responses (over 50%) or too many suspiciously fast RTs (more than 20% faster than 150ms), participant had to repeat the first practice phase and before that, the task instruction was presented again. After passing the first practice phase, participants entered the second practice task, where a response deadline was introduced. If the participants did not respond within 1000 ms, the feedback “Zu langsam!”, the German expression for *too slow*, preceded the following trial for 500 ms. Time was counted from the onset of the stimulus and if their average RT was slower than 800 ms, they had to repeat the second practice phase. The third practice phase had the same passing and repeating criteria as the second one. Nevertheless, participants received feedback for too slow responses. If they have not given any response within 800 ms, the message “Zu langsam!” appeared in red color again above the stimulus for 500 ms. The stimulus remained displayed until a key was pressed or the response time of 1000 ms was exceeded. After completing the three practice phases, the actual experiment, the main task, started. It had the same parameters as the third practice task except that it contained 18 times as many trials. This resulted in a total of 648 trials in the main task (72 probes, 72 targets, 288 irrelevant items, 72 familiarity-related inducer items, and 144 unfamiliarity-related inducer items were presented). The four conditions (group C1-C4) differed only in the order of which response pattern (narrow or wide key distance) and which probe items (forename or surname) was completed the task with first. Namely, each of the four groups started with one specific combination of the narrow or wide key distance response condition with either their forenames or surnames as first probe items. Within one condition, the task was divided in two blocks containing each 324 trials. Between the blocks, there was a pause and a practice task to repeat the obtained target names and to exercise the modified key distance response pattern (narrow vs. wide). The two blocks differed mainly in the key distances. Those participants who started with the narrow condition (C2 and C4) switched to wide key distance response patterns and those who started with the wide condition (C1 and C3) switched to the narrow key distance response patterns. Within the blocks, the order of the questions of the forenames and surnames was changed only resulting in 162 trials per forename and surname condition. The filler items were subsequently intermingled among the randomly displayed probe, target and irrelevant items, with the restriction that filler trials never followed another filler trial and each of the nine filler items preceded the probe, target, and irrelevant items exactly one time.

Statistical analyses were conducted using the free software R. For the simple *t*-test, Bayesian likelihood ratio test and general statistical output, specific R packages were needed ("neatStats", "data.table", "ez", "pROC", "ggplot2", "sizeMat", "aod", "plyr", "BayesFactor", "TOSTER", "Hmisc", "MBESS").

Results

Mean RTs (with *SDs*) for probe and irrelevant items for both narrow and wide key distance conditions are given in Table 1.

Table 1

Mean RTs and SDs for probe and irrelevant items for each group (C1, C2, C3 and C4) and key distance condition

	Probe Items		Irrelevant Items	
	Narrow	Wide	Narrow	Wide
C1	546±60	576±50	468±57	505±56
C2	561±46	514±59	497±35	445±35
C3	537±54	572±46	470±45	520±54
C4	578±58	531±47	517±55	456±47

We conducted a Welch corrected *t*-test with within-subject factors Key Condition (narrow vs. wide). Table 2 shows the results of the Welch corrected *t*-test with corresponding effect sizes (*d* within) and *p*-values for the differences of probe mean RTs and irrelevant mean RTs in both key distance conditions.

Table 2

Welch corrected t-test for Mean RTs and Accuracy rates for both key distance conditions

Key distance	Mean RT			Accuracy		
	$t(58)$	p	Cohen's d	$t(58)$	p	Cohens's d
Narrow	14.90	.00	1.94 [1.50, 2.37]	-5.27	.00	-0.69 [-0.97, -0.40]
Wide	13.54	.00	1.76 [1.35, 2.17]	-5.73	.00	-0.75 [-1.03, -0.45]

The *t*-test examining the main hypothesis which predicted shorter mean RTs at probe opposing to irrelevant items in the wide condition compared to the narrow key distance condition due to assumingly better mental distinctiveness resulting from the bigger physical distance of the participants two hands revealed no significant difference between the two conditions at the specified $p < .05$ level, $t(58) = 0.24$, $p = 0.813$, $d$within = 0.03, 95% CI [-0.22, 0.29], BF01 = 6.87.

The accuracy rate measured as correct responses divided through incorrect responses comparing both key distance conditions yielded also no significance at the specified $p < .05$ level, $t(58) = -0.40$, $p = 0.690$, $d$within = -0.05, CI [-0.31, 0.20], BF01 = 6.50.

Analysis of differences between the mean probe to irrelevant RTs in both narrow and wide key stance conditions found a substantially significant positive correlation, $r(116) = .784$, CI [.702, .845], $p < .001$. The total mean RT differences between probe and irrelevant items were $M = 551.90$, $SD = 55.72$ vs. $M = 484.71$, $SD = 54.38$ (raw mean difference: 67.19, CI [60.58, 73.80]), respectively.

The comparison of both probe and irrelevant items' RTs revealed a substantial significance at the specified $p < .05$ level, $t(117) = 20.14$, $p < .001$, $d = 1.85$, CI [1.55, 2.15], BF10 = 2.95 × 10^36. The mean RT differences in ms within the narrow and wide key distance condition were at narrow probe-minus-irrelevant: $M = 67.67$, $SD = 34.88$ and at wide probe-minus-irrelevant: $M = 66.71$, $SD = 37.85$.

Discussion

Main purpose of the present study was to validate the key distance effect found by Lakens et al. (2011) by using the RT-CIT. Unfortunately, this key distance effect could not be transferred from the Stroop-task to the CIT and neither could it be replicated in general. The Bayesian Factor analysis showed through comparing the RT mean and accuracy rate probe-to-irrelevant differences between the narrow and wide key distance condition no statistical significance ($BF01 = 6.87$ and $BF01 = 6.50$, respectively). Both Bayesian Factors indicate that the obtained data is approximately seven times as likely under the null hypothesis than under the alternative hypothesis.

However, the general validity and effectiveness of the familiarity-related inducer RT-CIT in revealing that participants try to hide their true identity while having longer reaction times at probe items (their own forenames and surnames) compared to irrelevant items could be shown according to the found substantial statistical significance ($BF10 = 2.95 \times 10^{\wedge}36$).

It seems possible that the spatial dislocation of both hands requires more attention, thus being more cognitively demanding and causing different responses on probe vs. irrelevant items in the wide key distance condition. This conclusion is based on the notion of the underlying cognitive principle of the single processing unit, mentioned in the introductory chapter on page four. However, this runs contrary to Lakens et al. (2011), who argue that wider key distance helps telling things apart. The notion of wider key distances may be cognitively more effortful could be plausible in the context of the *Automaticity theory* of the Stroop effect, where automatic reading does not need controlled attention, whereas the spatial dislocation of the required response pattern may interfere with the automatic answering on stimuli in the wide key distance condition.

A possible explanation for the findings of the present study could be the limited transferability of the key distance effect (KDE) from the Stroop task (Lakens et al., 2011) to the RT-CIT. Whereas in the Stroop task there are normally three different sources of interference: motor (due to different responses required for or implied by color vs. color word), meaning (different meanings of color and color word), and task (respond to color vs. respond to word), in the RT-CIT there are only two sources of interference: motor (different types of required responses by left vs. right hand), and meaning (responses according to familiarity vs. unfamiliarity). In the RT-CIT, the presence of motor and meaning interference is concurrently active and there may be no hierarchical processing structure unlike in the Stroop task.

Another possible explanation of the findings is that Lakens et al.'s (2011) proposal that perceived space between response keys automatically gives rise to spatial representations of color categories is not fully reasonable (Jonas et al., 2014). Spatial response discriminability contributes to the KDE (Jonas et al., 2014; Proctor & Chen, 2012), as opposed to the anatomical discriminability effect stated by Lakens et al. (2011).

Chen and Proctor (2014) showed that the Stroop effect is influenced by the conceptual distance and the intervening keys between response keys rather than the physical key distance according to Lakens et al.'s (2011) hand separation or anatomical discriminability account. According to the *action goal account* (Buhlmann et al., 2007), response keys are coded as action goals and due to farther distance in these action goals, a lesser interference and consequently a smaller Stroop effect occurs. Hypothetically, when the action goals are less distinguishable, the performance will be worse and therefore, applied to the RT-CIT, longer reaction times should be measurable at probe compared to irrelevant items.

Whether the distance effect is applicable at all, remains an open question. And if it is, which effect would it have on the ease of answering "no" on the cognitive dissonant denial of one's own identity referring probe item in dependence of the amount of qualitative and quantitative distance of the response keys.

Possible limitations of the present study may be the rather small study sample resulting in approximately 15 participants per group condition and the unequal gender distribution containing mainly women. Another limitation could be that the study sample is consisting exclusively of psychology students. As a further matter, the conceptual similarities between the RT-CIT and the Stroop task and the comparability of the observable KDE in both experiments may be rather restricted. Potentially, through making the RT-CIT more similar to the Stroop task, a KDE could be measurable. More similarity could be reached by implementing another source of interference at the RT-CIT, maybe through introducing a color task, which would increase the cognitive demandingness. But that further task could interfere with the general aim of the RT-CIT in validly detecting deception.

Conclusively, the KDE found by Lakens et al. (2011) at the Stroop task could not be transferred to the familiarity-related inducer RT-CIT. However, the familiarity-related inducer RT-CIT is a highly effective method of detecting deception, which could be empirically proven in the present study, too. Finally, assumptions and later research following the Lakens et al. (2011) paper suggest that the underlying cognitive mechanisms of the KDE are rather complex and for the transference to the RT-CIT a more sophisticated test method is required.

References

Buhlmann, I., Umiltà, C., & Wascher, E. (2007). Response coding and visuomotor transformation in the Simon task: The role of action goals. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1269-1287. https://doi/10.1037/0096-1523.33.6.1269

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 69*(6), 1263-1277. https://doi.org/10.2307/1130919

Chen, J., & Proctor, R. W. (2014). Conceptual response distance and intervening keys distinguish action goals in the Stroop color-identification task. *Psychonomic Bulletin & Review, 21*(5), 1238-1243. https://doi.org/10.3758/s13423-014-0605-6

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of student's *t*-test. *International Review of Social Psychology*, *30*(1), 92–101. https://doi.org/10.5334/irsp.82

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. https://doi.org/10.1037//0022-3514.74.6.1464

Gweon, H., & Saxe, R. (2013). Developmental cognitive neuroscience of theory of mind. In J. Rubinstein, & P. Rakic (Eds.), *Neural circuit development and function in the brain* (Vol. 3, pp. 367-377). Academic Press. https://doi.org/10.1016/B978-0-12-397267-5.00057-1

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* John Wiley & Sons Inc.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2-9. https://doi.org/10.7771/1932-6246.1167

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Johnson, A., & Proctor, R. W. (2004). *Attention: Theory and practice.* Sage Publications, Inc.

Jonas, M., Eloka, O., Stephan, J., & Franz, V. H. (2014). Need for space: The key distance effect depends on spatial stimulus configurations. *PLoS One, 9*(3), 1-12. https://doi.org/10.1371/journal.pone.0091432

Kleinmuntz, B., & Szucko, J. J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist*, *39*(7), 766-776. https://doi.org/10.1037/0003-066X.39.7.766

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D., Schneider, I. K., Jostmann, N. B., & Schubert, T. W. (2011). Telling things apart: The distance between response keys influences categorization times. *Psychological Science*, *22*(7), 887-890. https://doi.org/10.1177/0956797611412391

Langleben, D. D., & Moriarty, J. C. (2013). Using brain imaging for lie detection: Where science, law, and policy collide. *Psychology, Public Policy, and Law, 19,* 222-234. https://doi.org/10.1037/a0028841

Lukács, G., & Ansorge, U. (2019). Information leakage in the response time-based concealed information test. *Applied Cognitive Psychology, 33*(6), 1178-1196. https://doi.org/10.1002/acp.3565

Lukács, G., Gula, B., Szegedi-Hallgató, E., & Csifcsák, G. (2017). Association-based concealed information test: A novel reaction time-based deception detection method. *Journal of Applied Research in Memory and Cognition*, *6*(3), 283-294. https://doi.org/10.1016/j.jarmac.2017.06.001

Lukács, G., Kleinberg, B., & Verschuere, B. (2017). Familiarity-related fillers improve the validity of reaction time-based memory detection. *Journal of Applied Research in Memory and Cognition*, *6*(3), 295-305. https://doi.org/10.1016/j.jarmac.2017.01.013

Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsak, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56–65. https://doi.org/10.1016/j.ijpsycho.2016.10.010

Morey, R. D., & Rouder, J. N. (2018). *Bayes factor: Computation of bayes factors for common designs.* https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* https://doi.org/10.3758/s13428-018-01193-y

Proctor, R. W., & Chen, J. (2012). Dissociating influences of key and hand separation on the stroop color-identification effect. *Acta Psychologica, 141*(1), 39-47. https://doi.org/10.1016/j.actpsy.2012.06.005

Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience*, *34*(1-2), 125-129. https://doi.org/10.3109/00207458708985947

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis*. Psychonomic Bulletin & Review, 16*(2), 225-237. https://doi:10.3758/PBR.16.2.225

Seymour, T. L., & Schumacher, E. H. (2009). Electromyographic evidence for response conflict in the exclude recognition task. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 71-82. https://doi:10.3758/CABN.9.1.71

Steinbrook, R. (1992). The polygraph test: A flawed diagnostic method. *The New England Journal of Medicine, 327,* 122-123. https://doi.org/10.1056/NEJM199207093270212

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428-453. https://doi.org/10.1037/bul0000087

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture—word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16*(6), 987-999. https://doi.org/10.3758/PBR.16.6.987

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: Response latency-based measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 46–62). Cambridge University Press. https://doi.org/10.1017/ CBO9780511975196.004

Verschuere, B., Suchotzki, K., & Debey, E. (2014). Detecting deception through reaction times. *Detecting Deception: Current Challenges and Cognitive Approaches*, 269-291. https://doi.org/10.1002/9781118510001.ch12

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779-804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790

Abstract

The Concealed Information Test (CIT) is a deception detection method, which measures the recognition of a relevant stimulus amongst irrelevant ones. Besides many ways of operationalizing the recognition of stimuli, the measuring of reaction times is the most economic one. Based on theoretical assumptions of an empirical study concerning key distance effects at the Stroop task, which is a widely known psychological experiment and examines the delay in reaction time between congruent and incongruent stimuli, it was predicted that the answer key distance influences the reaction times (RT) and accuracy rates of responding to items during the RT-CIT. This could not be found. Thus, the general validity and effectiveness of the familiarity-related inducer RT-CIT could be demonstrated. Further, potential limitations of the applicability of the key distance effect on the RT-CIT are discussed.

Zusammenfassung

Der Concealed Information Test (CIT; *concealed* bedeutet auf Deutsch verborgen) ist eine Täuschungsdetektionsmethode, die das Erkennen eines relevanten Stimulus unter irrelevanten Stimuli misst. Es gibt viele Möglichkeiten das Erkennen von Stimuli zu operationalisieren. Die Messung von Reaktionszeiten ist die ökonomischste Methode unter diesen Möglichkeiten. Beruhend auf theoretischen Annahmen einer empirischen Untersuchung, die Antworttasten-Entfernungseffekte bei dem Stroop Test ermittelte, wurde vermutet, dass die Reaktionszeiten (RT) und die Genauigkeitsraten bei dem Antworten auf die Stimuli des RT-CIT durch die Entfernung der Antworttasten beeinflusst werden. Dies konnte nicht belegt werden. Jedoch konnte die generelle Gültigkeit und Effektivität des Vertrautheits-verbundenen Auslöser RT-CITs bestätigt werden. Ferner werden mögliche Einschränkungen der Anwendbarkeit des Antworttasten-Entfernungseffekts auf den RT-CIT erörtert.