Evaluating processes involved in recognition decisions using different model comparison techniques

Inaugural- Dissertation

zur
Erlangung der Doktorwürde
der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät
an der Albert-Ludwigs-Universität Freiburg

vorgelegt von
Anne Voormann
geboren am 30.05.1994 in Wuppertal

WS 2020 / 2021

Dekan der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät:
Prof. Dr. Olaf N. Rank

GutachterInnen:
Prof. Dr. Karl Christioph Klauer
Prof. Dr. Andrea Kiesel

Datum des Promotionsbeschlusses:
28.01.2021

To my family,
who always supported me in everything I did.

And especially to my grandfathers, who both would (have) like(d)
to see me pursuing an academic career.

Contents

Summary

An ongoing discussion within the field of recognition memory concerns the nature of information retrieved for recognition decisions. Recognition describes the ability to identify previous encountered events as such. And models describing recognition decisions argue either that recognition decisions are mediated via discrete states (threshold models) or that a continuous amount of evidence drives recognition decisions (continuous models). Generally, there exist two different ways to distinguish between proposed candidate models: 1) Investigation of model fit and model validity and 2) test of distinct predictions retrieved from model's structure.

With my dissertation I contribute to the above mentioned discussion, presenting three manuscripts that cover both techniques of model discrimination. My first manuscript sheds light upon a quite unrecognized recognition task, paired-word recognition, through examination of quantitative model fit. We extended two models, one from the threshold model class and one from the continuous model class. Next to observing that single-word and paired-word recognition differ in mnemonic processes, we found evidence that the assumption of discrete states underlying paired-word recognition decisions can best account for participants' behavior. As both models considered in my first manuscript were newly introduced for this specific task, I validated those via selective-influence studies assessing qualitative model fit in my second manuscript. Because, only if experimental manipulations, as strength and base-rate manipulations, map onto meaningful parameter(s) considered in the model, models can validly capture the decision mechanisms. And finally, in my third manuscript, I took the approach of testing distinct predictions by validating the existence of the error speed effect. While continuous models can naturally account for the error speed effect, only some threshold models are capable of explaining it. Thus, through our verification of its existence we were able to narrow the candidate models which have to be considered for the description of recognition decisions.

Based on these results I argue, in line with previous accounts, that the nature of information retrieved for recognition decisions might depend on task and stimulus features. Furthermore, I suggest a shift of the discussion for future research away from the nature of underlying information and towards an inspection of the processes underlying recognition decisions, as for example considering the necessity of evidence free guessing.

Zusammenfassung

Die Beschaffenheit der Informationen, die Wiedererkennungsentscheidungen zugrunde liegen, wird in der Rekognitionsforschung ausgiebig diskutiert. Rekognition beschreibt dabei die Fähigkeit von Personen, zuvor erlebte Ereignisse wiederzuerkennen. Modelle zur Beschreibung von Rekognitionsentscheidungen argumentieren, dass entweder diskrete Zustände (Schwellenwertmodelle) oder kontinuierliche Evidenz (kontinuierliche Modelle) die Grundlage für Rekognitionsentscheidungen bilden. Um zwischen diesen beiden Ansätzen zu diskriminiren, können zwei verschiedene Techniken verwendet werden: 1) ein Vergleich der Höhe der Modellpassung sowie die Inspektion der Modellvalidität und 2) die Untersuchung distinkter Vorhersagen, abgeleitet aus der jeweiligen Modellarchitektur.

Mit meiner Dissertation trage ich zu der oben genannten Debatte durch drei Manuskripte bei, die beide Techniken des Modellvergleichs abdecken. In meinem ersten Manuskript haben wir jeweils ein Modell aus der Klasse der Schwellenwertmodelle und ein Modell aus der Klasse der kontinuierlichen Modelle erweitert, um mithilfe von quantitativen Modellvergleichen eine eher unbekannte Rekognitionsaufgabe, das Wiedererkennen gepaarter Stimuli, genauer zu untersuchen. Neben Unterschieden zwischen einer klassischen Rekognitionsaufgabe und dem Wiedererkennen gepaarter Stimuli, die wir auf mnemonische Unterschiede zurückführen konnten, fanden wir, dass ein Modell mit der Annahme von zugrundeliegenden diskreten Zuständen das beobachtete Verhalten in dieser Aufgabe am besten erklären konnte. Da beide Modelle meines ersten Manuskriptes neu eingeführt wurden, widme ich mich in meinem zweiten Manuskript mithilfe von Tests zur selektiven Beeinflussbarkeit von Parametern deren Validierung. Denn nur, wenn experimentelle Manipulationen, wie eine Stärkenmanipulation oder eine Basisraten Manipulation, durch inhaltlich sinnvolle Parameter erfasst werden, können Modelle den zugrundeliegenden Entscheidungsprozess valide beschreiben. Abschließend befasse ich mich in meinem dritten Manuskript anhand eines Beispiels mit distinkten Vorhersagen der beiden Modellklassen, abgeleitet aus deren jeweiliger Modellarchitektur. Während kontinuierliche Modelle den *error speed* Effekt durch ihre Beschaffenheit vorhersagen, können nur spezifische Schwellenwertmodelle, solche die fälschliches Erinnern zulassen, diesen erklären. Durch die Verifizierung der Existenz des error speed Effektes, konnten wir die Modelle zur Beschreibung der Rekognition eingrenzen.

Ausgehend von den Ergebnissen meiner drei Manuskripte argumentiere ich, dass,

wie bereits andere Wissenschaftler nahelegten, die Beschaffenheit der Informationen, die Wiedererkennungsentscheidungen zugrunde liegen, von der Aufgabe, dem Stimulusmaterial und dem Kontext abhängig sind. Darüber hinaus plädiere ich für eine Verschiebung des Fokusses der Diskussion in zukünftigen Untersuchungen: weg von der reinen Beschaffenheit der Informationen und stärker zu den Entscheidungen zugrundeliegenden (Meta-)Prozessen, wie der Notwendigkeit von Raten.

Einmal kündigte der Physiker Leo Szilard seinem Freund Hans Bethe an, er wolle eventuell ein Tagebuch führen: “Ich habe nicht vor, etwas zu veröffentlichen. Ich möchte die Tatsachen nur festhalten, damit Gott Bescheind weiß.”
Daraufhin fragte Bethe: “Glauben Sie nicht, dass Gott die Tatsachen schon kennt?” – “Ja,”, erwiderte Szilard, “die Tatsachen kennt er. Aber diese Version der Tatsachen kennt er noch nicht.”

from *Das Atom in der Falle*
by Hans Christian Baeyer

[Once the physicist Leo Szilard announced to his friend Hans Bethe, he would possibly start writing a diary: “I do not aim in publishing anything. I just want to record the facts so that god is aware of them.”
Thereupon Bethe asked: “Don’t you belief that god already knows about the facts?” – “Yes,” Szilard responded, “he is aware of the facts. But he doesn’t know this particular version of them.”]

1 Introduction

The ability to recognize situations we have encountered previously as such is a fascinating and fundamental cognitive mechanism that is essential to nearly all aspects of everyday lives. It is fascinating, because people differ substantially in their individual ability to recognize situations, and fundamental, as it allows us to find one's way in social and natural environments. Situations refer to, for example, a voice previously heard and now recognized as being familiar or to a nature spot identified as already seen. Thus, it is not surprising that ongoing debates exist about the processes and mechanisms involved in recognition in general and about the type of information being available during response selection.

In the case of recognition memory, there are two major categories of cognitive models which capture processes involved in recognition decisions: process models and measurement models. While *process models* focus on the concrete structures and processes underlying recognition – including *encoding and retrieval* –, *measurement models* offer a framework to measure the conceptual processes involved in *memory retrieval* within certain situations and tasks (Brandt, 2007). Thus, process models aim to describe recognition on a more fine grained level compared to measurement models which focus more strongly on higher level mechanisms. Nevertheless, measurement and process models do not contradict one another. Rather, investigating the basic mechanisms of retrieval using measurement models helps to specify the retrieval processes within process models (Malmberg, 2008).

One open and often discussed question in the framework of memory retrieval addresses the quality of information being available during recognition decisions. This debate takes mostly place on the level of measurement models, as they offer an easy statistical framework to determine the differences between processes across different tasks and conditions. On the one hand, some models assume people to have access to a graded familiarity signal driving recognition responses. Within those theories, elicited familiarity is compared to a criterion determining the recognition response (continuous models; e.g., Banks, 1970; Pazzaglia, Dubé, & Rotello, 2013; Ratcliff, 1978). Whereas, other theories assume recognition decisions to be mediated by discrete states. In such a case, people only have access to the concrete decision state (e.g., detection or uncertainty) driving the response but no information about the processing path leading to the respective state (threshold models; e.g., Batchelder & Alexander, 2013; Batchelder & Riefer, 1999).

In the present dissertation I aimed to investigate the qualitative nature of information underlying recognition decisions using different tasks and methods. In the first manuscript (Voormann, Spektor, & Klauer, in press), we evaluated absolute model fit in the case of paired-word recognition. Therefore, we extended models from both theoretical positions and evaluated which model could numerically account best for this task. In the second manuscript (Voormann, Spektor, & Klauer, unsubmitted manuscript), a comparison of those two models is achieved through an assessment of both model's general ability to capture manipulations in a psychological meaningful way. More precisely, departing from well investigated manipulations of a traditional recognition task, we assessed whether both models reflect those manipulations in the expected manner (known as selective-influence studies). Finally, the third manuscript (Voormann, Rothe-Wulf, Starns, & Klauer, 2021) examines the question about the type of retrieval information is addressed by distinct behavioral predictions deduced from the two model frameworks, continuous and threshold models. Specifically, we investigated whether the speed of recognition errors represents the (continuous) amount of misleading memory evidence.

1.1 Processes involved in recognition memory

Recognition describes the ability to identify previously encountered situations as such. Consequently, in a typical recognition experiment participants first have to learn stimuli in a study phase and are later asked to categorize learned (targets) and new stimuli (lures) as previously studied, thus as being "old", or not studied, thus being "new". Given the two possible types of stimuli, targets and lures, and the two possible responses, "old" and "new", four types of responses can be distinguished: *hits*, *misses*, *false alarms* and *correct rejections* (Green & Swets, 1966). A correct "old" response to a target is called a hit while an incorrect "new" response to a target is called miss. Similarly, a correct "new" response to a lure is called correct rejection while an incorrect "old" response to a lure is called false alarm. As the probability of misses can be defined as $P(\text{miss}) = 1 - P(\text{hit})$ and correspondingly the probability of correct rejections can be defined as $P(\text{correct rejection}) = 1 - P(\text{false alarm})$, most models for recognition memory use only the pattern of hits and false alarms to discuss processes involved in recognition decisions (e.g., Batchelder & Riefer, 1999; Wickens, 2002; Yonelinas, 1994).

A fundamental process underlying recognition decisions that is often considered in process models is the match of a probe (presented stimulus) to the information stored in memory. The degree of similarity between the probe and memory is then compared to a criterion to determine the response (Humphreys, Pike, Bain, & Tehan, 1989). However, process models differ in their types of information stored in memory as well as their computation of the global activation during retrieval.

For example in the case of MINVERVA 2 (Hintzman, 1988), it is assumed that each situation during study is coded and stored in a separate memory trace. If the same stimulus appears twice during study, two traces are stored. The precision with which a trace is stored in memory is determined by the learning rate. A learning rate of 1 means a perfect match between the stimulus presented and its trace stored in memory. During test, memory (all traces stored) is probed with a presented stimulus. To evaluate the overall familiarity (called intensity within MINERVA 2) that the respective probe evokes, the probe is matched to each trace separately and the resulting activations are summed across all traces. Thus, the learning rate already determines to a certain extend how well a stimulus will match the corresponding trace during test. Final responses are selected by a comparison of the elicited overall familiarity to a global criterion. Each time familiarity is higher than the criterion the probe is categorized as "old" otherwise the probe is categorized as being "new". As retrieval is often not simply based on a single probe but on additional conditional information, an extension, MINERVA-dm (MINERVA for decision making), allows to evaluate the familiarity on subsets of traces matching certain conditions (Dougherty, Gettys, & Ogden, 1999).

In comparison, the model SAM (search of associative memory; Gillund & Shiffrin, 1984) assumes that each image (comparable to MINERVA 2's traces) stored in memory contains three different types of information: context information, inter-item relations, and a self-coding part containing the item information. Within SAM, the strength and correctness of encoding depends on the amount of and time spent for rehearsal as well as on the coding of the information itself. With regard to the retrieval process, SAM is similar to MINERVA 2. Again, familiarity is computed as the sum of all activations resulting from the match between the probe and each image and a categorization as "old" is performed when the familiarity is higher than the subject's criterion. Moreover, SAM allows to weigh the information used during retrieval and thus captures the possibility of attending and considering certain probe information with differing strength during recognition decisions.

As can be seen from the two examples mentioned above, process models aim to describe the processes involved during encoding and retrieval quite precisely for a number of different tasks and phenomena (Gillund & Shiffrin, 1984; Hintzman, 1988). By simulating the processes involved using meaningful parameter constellations, they model the behavioral patterns of different memory phenomena and offer therewith an explanation of the mechanisms involved.

In the light of their high explanatory potential, process models, however, have two major limitations: On the one hand, process models are often only implemented to simulate behavior within certain tasks and contexts. Thus, although it is possible to estimate relevant parameters in some cases applying maximum likelihood estimation, such a procedure has rarely been investigated (Brandt, 2007). In most cases, parameter constellations are chosen to replicate a certain data pattern, but they are seldom tested based on observed data patterns. On the other hand, process models make no explicit assumptions about the information being available to the individual during retrieval. Thus, although the parameters depict the involved processes very detailed, process models do not indicate which of these processes and information can be accessed and controlled intentionally.

Measurement models try to fill those gaps. On the one hand, they are explicitly designed to capture the processes involved during retrieval through parameter estimation that is based on the observed data (see, for example, Banks, 1970; Batchelder & Riefer, 1999; Van Zandt, 2000). On the other hand, measurement models are embedded in an ongoing debate about which type of information is available to individuals during response selection (see, for example, Batchelder & Alexander, 2013; McAdoo, Key, & Gronlund, 2018; Pazzaglia et al., 2013).

1.2 Measurement models for recognition decisions

The aim of measurement models is to describe the underlying processes based on the observed data. Thus, although measurement models do not make any assumptions about the exact processes leading to the amount of evoked memory evidence, they try to quantify the processes involved during retrieval. Based on their parameter estimates, they allow to compare the different mechanisms involved across different tasks and conditions.

There exist two major classes of measurement models, *continuous* and *threshold models*. Both make different assumptions about the quality of information being available to the individual during recognition decisions. While continuous models assume access to a graded familiarity signal that guides recognition decisions (Pazzaglia et al., 2013), threshold models assume that recognition decisions are mediated by discrete states (Batchelder & Alexander, 2013).

In the following, I will elaborate on the core ideas of continuous and threshold models in general. Additionally, I selected two models of each class to explain their core assumptions in more detail. The selected models form either the basis for modeling the simultaneous recognition of paired stimuli investigated in Manuscript 1 and 2 or are used as candidates within model comparison conducted in Manuscript 3.

1.2.1 Continuous evidence accumulation

A multitude of models assume continuous evidence accumulation as the basis for recognition decisions. These include, for example, signal detection theory (SDT; Banks, 1970), the diffusion model (Ratcliff, 1978), or the race model of recognition memory (Van Zandt, 2000). All models share the assumption that evidence is accumulated (over time) and that individuals have access to the exact amount of this accumulated evidence during response selection. However, continuous models implement two different mechanisms for the response selection itself. The first type of models, like SDT, maps the familiarity signal retrieved from the stimulus information directly to a certain response by comparing it to a response criterion (Wickens, 2002). The second type of models assumes that stimulus information only inform a decision process which finally determines the response, like the diffusion model (Ratcliff, 2014).

Signal detection theory

A very popular model, to capture decision processes based solely on the hits and false alarms in various fields is SDT (e.g., Banks, 2000; Tanner & Swets, 1954). SDT was originally implemented to describe decision processes in perception (Tanner & Swets, 1954). The fundamental idea is that each stimulus elicits a certain amount of (neural) activation and that it is categorized based on this amount of activation relative to a criterion, c, either as noise or as signal (Swets, Tanner, & Birdsall, 1961). If the perceived activation exceeds the criterion, the stimulus is categorized as signal otherwise as noise. As the information obtained from noise differs randomly on a trial-by-trial basis, this information can be represented as a random variable, whose distribution is commonly assumed to be normal. Correspondingly, signal activation also follows a normal distribution because information obtained from signals combines the elicited random noise with the activation by the signal. Thus, the mean activation difference between signal and noise distributions represents signal sensitivity, d' (Green & Swets, 1966).

SDT can easily be adapted to recognition memory when assuming that noise trials represent lures and signal trials represent targets. The amount of activation elicited corresponds to the familiarity perceived by a certain stimulus while signal sensitivity can be understood as memory sensitivity capturing individuals' general ability to differentiate between targets and lures. Moreover in the case of recognition memory, additional encoding variance exists for targets due to, e.g., attention fluctuation during study. Thus, the variance for familiarity elicited by targets, σ_{old}, should exceed the variance of lures (e.g., Kellen & Klauer, 2018). As the response criterion determines the categorization of a stimulus as "old" or "new", all lures eliciting familiarity values lower than the criterion will be classified as correct rejections while lures activating familiarity higher than the criterion will be classified as false alarms. Correspondingly, targets with a familiarity

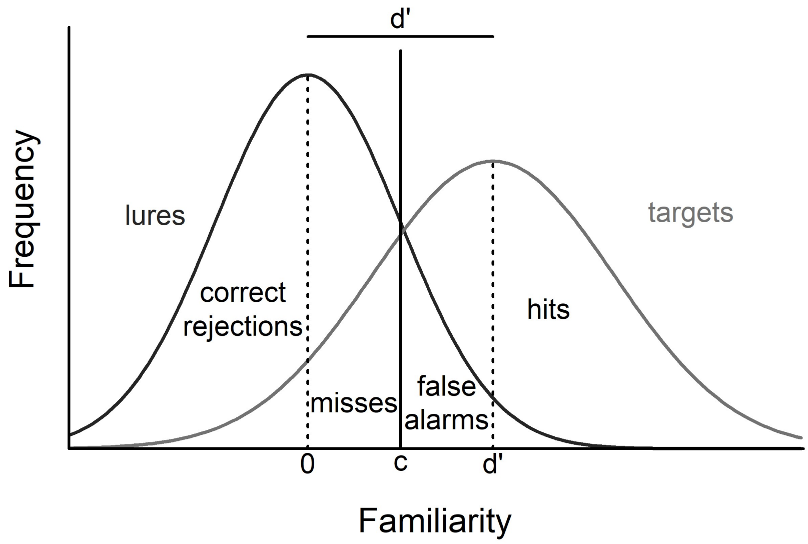

Figure 1.1: Illustration of signal detection theory for a recognition task. Dashed lines represent distribution means for new and old words with d' describing the memory sensitivity for targets. The response criterion c represents the critical familiarity value to categorize stimuli as "old" or "new".

higher than the criterion will be classified as hits and targets with familiarities lower than the criterion as misses (see Figure 1.1; Wickens, 2002).

For recognition memory in the SDT framework, two main mechanisms affect the components of the decision process: On the one hand, encoding and retrieval strength can be manipulated by, e.g., the number of presentations, presentation time, or the depth of processing which impacts the memory sensitivity, d' (see, for example, Gillund & Shiffrin, 1984; Morrell, Gaitan, & Wixted, 2002; Snodgrass & Corwin, 1988). As mentioned above, the retrieval process is solely based on perceptual information. Adapting a multidimensional version of the SDT to the simultaneous recognition of paired stimuli in Manuscript 1, we assumed that the perceptual information of paired stimuli presented simultaneously interact with each other leading to mnemonic dependencies.

On the other hand, strategic decision adaptations manipulated, for example, by varying pay-off or base-rates influence the height of the criterion, c (Bröder & Schütz, 2009; Snodgrass & Corwin, 1988). As strategic adaptations can also occur on a trial-by-trial basis depending on the context in which the stimuli appear (Rhodes & Jacoby, 2007), we assumed in Manuscript 1 and 2 that this response criterion can also be adapted according to the cognitive context, more precisely to the amount of familiarity elicited by the pair member.

Diffusion model

In the diffusion model, the evidence elicited by the stimulus also plays an important role, but unlike in SDT where it is mapped directly to a response, evidence first accumulates over time and therewith only informs the decision process (Ratcliff, 2014). The diffusion model is based on the theory of memory retrieval (Ratcliff, 1978) but was successfully adapted to decision processes across a broad range of two-choice decision tasks including perception, recognition memory, interference tasks, and social categorization (e.g., Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; Ratcliff, 1978; Ratcliff & McKoon, 2008). The model owes its popularity to the fact that it uses not only response accuracies for the description of the underlying decision process, like SDT, but that it additionally allows to make predictions about mean response times (RTs) and RT distributions.

Within the diffusion model, evidence from the presented stimuli is sampled over time in a noisy accumulation process until one of two response boundaries is reached (see Figure 1.2; Ratcliff, 1978). The mean increase in evidence in favor of the one decision compared to the other, is captured by the drift rate, v, which is comparable to the amount of familiarity in SDT (Ratcliff, 2014). The drift rate thus captures the difficulty to discriminate between the two response options (Ratcliff & McKoon, 2008) and in the case of recognition, it captures the memory evidence for the stimulus. The better a stimulus is remembered the higher its drift rate towards the "old" response boundary. The better the stimulus is detected as a lure the lower (more negative) its drift rate and the accumulation process tends towards the "new" boundary. The zero point of the drift rate, the point in which the amount of evidence is indecisive regarding the two response options, is called drift criterion and corresponds to the point of the criterion, c, within SDT. As the accrual of information is a noisy process, within-trial variability exists during the accumulation process, leading to different RTs of stimuli with the same amount of memory evidence and leading from time to time to erroneous responses (Ratcliff, Smith, & McKoon, 2015). Additionally, between-trial variability, η, captures the possibility that targets (and lures) can differ in the specific amount of evidence they elicit for the respective response boundaries. This variability allows that targets are not only responded to incorrectly by accident, but can be misremembered. In addition, the amount of misleading evidence determines the speed of errors with a higher misleading evidence resulting into faster errors than low misleading evidence (Starns, Dubé, & Frelinger, 2018; Voormann et al., 2021). This reasoning is fundamental to the argumentation in my third manuscript.

Next to those processes accounting for the nature of stimulus information, there are processes capturing more task specific decision adaptation. The response boundary separation, a, describes, for example, the speed-accuracy settings during a task, with narrow boundaries representing a focus on response speed and wider boundaries depicting a focus on response accuracy. Additionally, the starting point for the accumulation process, z,

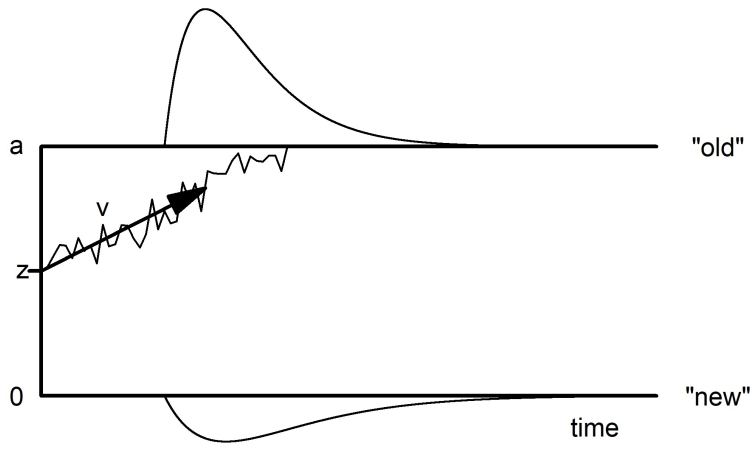

Figure 1.2: Schematic illustration of the diffusion model for a target trial with v representing the drift rate towards the “old” boundary, z the starting point, and a the boundary separation. The distribution on the “old” boundary represents the RT distribution of correct responses and the distribution on the “new” boundary represents the RT distribution of error responses.

can vary on a trial-by-trial basis within the uniform range, s_z. The position of the starting point indicates the bias towards a certain response. It can be manipulated, for example, by the proportion of specific responses during the experiment, a base-rate manipulation (Ratcliff & McKoon, 2008). For completeness, the non-decision time, T_{er}, captures the time needed for stimulus encoding and response execution. Its mean hight depends on the respective task but it can also vary uniformly on a trial-by-trial basis within range s_t. Thus, strictly speaking this parameter is no parameter describing the decision process itself but expresses the time of additional processes contributing to the RT.

1.2.2 Recognition via discrete states

Just as continuous models, threshold models or multinomial processing trees in general, assume that there exists an underlying continuous memory evidence signal leading into certain discrete cognitive states (Bröder & Schütz, 2009). However, once a discrete decision state is entered, for example a detection state for targets, the underlying information is lost and participants responses are finally mediated via those discrete states (Batchelder & Riefer, 1999). This information loss assumption is crucial for all thresholds models while in the concrete conception of cognitive states the models differ (Province & Rouder, 2012). Examples for threshold models are the one-high threshold model (Blackwell, 1963), the two-high threshold model (2HTM; Snodgrass & Corwin, 1988), the one-low threshold model (Luce, 1963; but see also Kellen, Erdfelder, Malmberg, Dubé, &

Criss, 2016), and the two-low threshold model (2LTM; Starns et al., 2018). The number of thresholds indicates the existence of a detection state for lures with one threshold models allowing for no such state. The hight of the threshold (high vs. low) indicates the existence of an exclusiveness criterion for detection states. With low threshold models allowing erroneous responses to result from mistaken detection and guessing while high threshold models only allow errors to result from incorrect guesses.

Two-high threshold model

The 2HTM is arguably the most famous threshold model to describe recognition decisions (Bröder & Schütz, 2009; Kellen, Singmann, Vogt, & Klauer, 2015; Province & Rouder, 2012). It considers three discrete cognitive states: a detection state for targets, a detection state for lures, and an uncertainty state out of which responses arise from guessing (see Figure 1.3; Snodgrass & Corwin, 1988). As for each threshold model, parameter estimates represent the conditional probabilities for certain stimulus types to result in the respective states. Thus, with probability d_o a target results in the detect old state out of which the participants provide an "old" response. With probability $1 - d_o$ detection of a target fails and participants enter a state of uncertainty out of which they guess with probability g the stimulus to be "old" and with probability $1 - g$ to be "new". Correspondingly, lures are correctly detected with probability d_n to be new and therefore categorized as "new" and with probability $1 - d_n$ participants again enter the exact same uncertainty state as for targets.

As also depicted in Figure 1.3, targets and lures share the same uncertainty state. In a state of uncertainty, not enough evidence exists for the stimuli to be detected, however, contextual information can be used to strategically guess the response. Thus, g represents the bias towards a certain response category and can be manipulated by pay-offs or base-

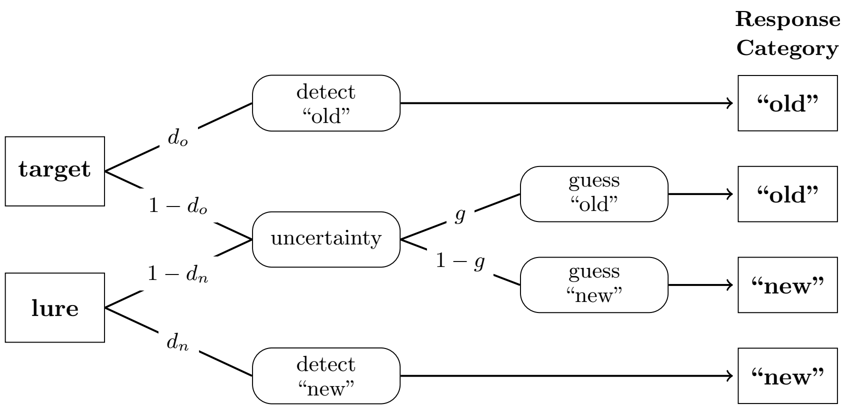

Figure 1.3: Illustration of the two-high threshold model.

rates just as the criterion, c, in SDT (Bröder & Schütz, 2009; Snodgrass & Corwin, 1988). Importantly, contextual information include the mental context as basis for guessing. This allowed us to consider dependencies in guessing for the simultaneous recognition of paired stimuli in our extended version of the 2HTM presented in Manuscript 1 and 2.

The probability to detect targets, d_o, captures the memory strength and the successful retrieval of targets. Larger values of d_o indicate a higher memory evidence across all targets which is the case for, e.g., a higher number of presentations during study or high imaginable vs. low imaginable words (Snodgrass & Corwin, 1988). As the underlying mechanism to reach certain states is still based on memory evidence, it is just as natural as for the SDT to assume that the two detection processes describing two simultaneous recognition decisions interact. We considered these interactions as mnemonic dependencies within the extended version of the 2HTM in Manuscripts 1 and 2.

Refined versions of multinomial processing trees, the RT-MPT, include additionally a description of the respective process times of the underlying cognitive processes (Klauer & Kellen, 2018). As particular process time distributions are tied to certain decisional states and in the case of the 2HTM recognition errors can only result from the same underlying uncertainty state, the speed of recognition errors is, in contrast to the diffusion model, not representative of the amount of misleading evidence (Starns et al., 2018). We challenged the assertion about error speed being informative of misleading evidence in Manuscript 3.

Two-low threshold model

The 2LTM extends the 2HTM by allowing for mistaken detection for targets and lures. In addition to the processes considered in the 2HTM, it includes a path (and parameter) that links targets to the detect new state with probability d_{nt} and lures to the detect old state with probability d_{ol} (see Figure 1.4; Starns et al., 2018). Due to its additional paths, RTs from recognition decisions hold information about the underlying decision path for a specific item in the framework of RT-MPTs because errors no longer result exclusively from the uncertainty state. Nevertheless, as only two discrete states underly error decisions (mistaken detection and uncertainty), information about the specific path are limited as discussed within the third manuscript.

The 2LTM is the youngest and least tested model of the presented measurement models for recognition decisions. Until now, this model has only been developed on a conceptual level by Starns et al. (2018) and is discussed in Manuscript 3, but its parameters have never been estimated so far. In fact, because of its flexibility and the amount of parameters needed, the 2LTM is not identified for standard recognition tasks even with reasonable restrictions. Thus, future studies should investigate whether the parameter manipulations described above also hold for the 2LTM.

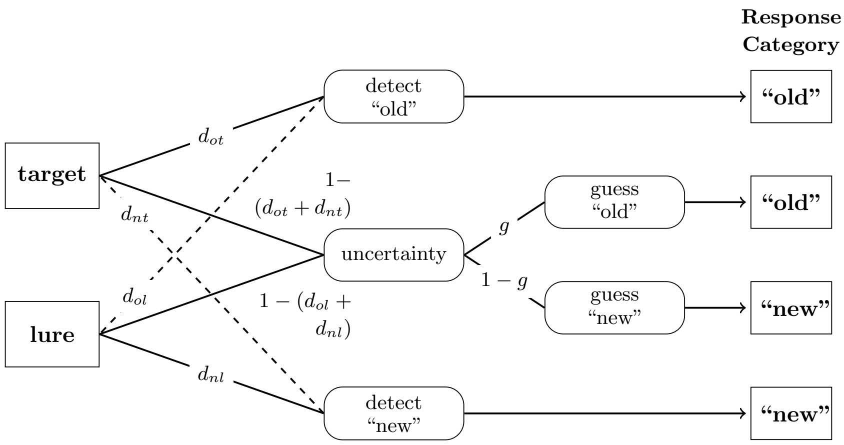

Figure 1.4: Illustration of the two-low threshold model. Dashed lines indicate the paths characterizing the low threshold assumption.

1.3 Discriminating retrieval mechanisms for recognition decisions

Attached to the question about which model describes recognition decisions best is a long and ongoing debate whether recognition decisions are based on a continuous familiarity signal or are mediated via discrete states. Whereas, some authors even argue that the quality of the decision process might depend on instructions or task and stimulus features (Kellen & Klauer, 2015; McAdoo et al., 2018, 2019). A third model class, dual-process models (Yonelinas, 1994), which assumes two independent processes of which one is continuous and the other one discrete, contributes to the debate as well. However, as dual-process models only have a minor role within my dissertation, I will not explain them in detail.

Generally, there exist two disparate techniques to compare models' adequacy to account for observed data. One possibility is to assess the absolute model fit via fit indices as information criteria like AIC (Akaike, 1974) or BIC (Raftery, 1986) or Bayes factor (Jeffreys, 1961). The absolute model fit captures the quantitative misfit between predicted data, based on the estimated parameters, and the observed data while most indices penalize additionally for model complexity (e.g., as the number of free parameters considered). By comparing the absolute fit of all models analyzed, that model will be extracted that describes the data numerically best. The other way to distinguish between models is via distinct behavioral predictions based on the respective model structures. As each model has its own structure, it predicts some behavioral patterns that other models would not predict. Thus, testing those predictions experimentally helps to discriminate different models and additionally to validate certain model assumptions.

A procedure frequently used to distinguish underlying recognition processes is the analysis of the shape of and the model fit to receiver operating characteristic (ROC) curves. ROCs relate the false alarm rate to the hit rate. Based on their varying model assumptions, models predict different shapes of the ROC curve. For example, while SDT predicts an inversed U-shaped ROC curve for rates and for z-transformed rates a linear shape, the 2HTM on the other hand predicts a linear ROC curve for rates and a U-shape for z-standardized rates. And finally, dual-process models predict an inverted U-shaped ROC curve for rates and a U-shaped ROC curve for z-transformed rates (Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). Yonelinas et al. (1996) capitalized on those diverting predictions and evaluated the shape of confidence based ROC curves in two experiments. The shape of the resulting ROCs seemed to be in concert with the predictions of dual-process models and contradicted the 2HTM and SDT. Additionally, a second task, a remember-know paradigm, was used to generate predicted ROC curves based on parameter estimates. Those predicted curves matched the observed confidence based ROC curves very well, fostering the result that a dual-process model underlies recognition decisions. The remember-know paradigm aims to distinguish responses based on event recollection from those based on familiarity and thus maps the assumptions of the dual-process model. However, Erdfelder and Buchner (1998) argued that the considered models were not adapted to confidence ratings. Adapting the evaluated threshold model to confidence ratings, it is well able to account for curved ROCs and even outperformed the dual-process model regarding model fit.

Bröder and Schütz (2009, 2011) tied up on that argument and therefore included in their meta-analysis only studies considering ROCs from binary-response recognition tasks. Within their meta-analysis and three additional experiments, they found no support for neither the 2HTM nor SDT. However, Dubé and Rotello (2012) criticized their meta-analysis as it contained a number of studies involving only two point ROCs which cannot be used to discriminate between curved or linear shapes. Elimination of those studies from the meta-analysis let to a better fit of SDT compared to the 2HTM. This result was even supported by two additional studies conducted by Dubé and Rotello (2012). Both studies considered an instructed base-rate manipulation (in contrast to an implemented base-rate manipulation) and the ROC curves from SDT fitted the observed data better than the ROCs generated by the 2HTM for both the aggregated data (Experiment 1 and 2) and the individual data (Experiment 2). To conclude, the shape of ROC curves do not allow to distinguish between the discrete or continuous processes underlying recognition decisions very well.

With a series of studies, Klauer and Kellen (Kellen & Klauer, 2011; Kellen, Klauer, & Bröder, 2013; Klauer & Kellen, 2015) contributed to the discussion using different tasks and a more sophisticated method to evaluate model fit, the minimal description length

(NML). Next to model flexibility based on the number of considered parameters, NML additionally accounts for the functional flexibility of models, penalizing them for the space of possible observations the respective model can account for (Kellen & Klauer, 2011). The results were again mixed. While dual-process models were favored evaluating first and second choice responses in a four-alternative forced-choice task (Kellen & Klauer, 2011) and ROC curves based on confidence ratings (Klauer & Kellen, 2015), the 2HTM on the other hand described binary-response ROCs best (Kellen et al., 2013).

Next to those studies assessing primarily model fit, there are a number of studies investigating distinct behavioral predictions. For example, Kellen and Klauer (2014) tested the conditional probabilities for strong and weak targets in a ranking task. While SDT assumes the conditional probability for targets being assigned rank 2 given that it was not ranked first (c_2) to be larger for strong than that for weak targets, the 2HTM assumes c_2 to be of the same size as the underlying guessing probability should not be affected by target strength. Nevertheless, in two experiments c_2 turned out to be larger for strong items than for weak items. Thus, the results matched the predictions of SDT and hint to the necessity of participants having access to a continuous memory signal.

Additionally, Starns et al. (2018) found evidence for the *error speed effect*. The error speed effect terms the finding that previously made errors of a single-recognition task predicts the probability of those errors to be correctly categorized in a subsequent two-alternative forced-choice task, in which participants have to select the target among two presented stimuli, a target and lure. As mentioned earlier, this effect is predicted by the diffusion-model. However, the 2HTM cannot account for it's occurrence because the same uncertainty state underlies all memory errors. Taken together, the occurrence of the error speed effect requires, just as the findings from ranking tasks, participants to have access to continuous memory evidence information while responding.

However, three studies assessing the *conditional independence* (Kellen & Klauer, 2015; Kellen et al., 2015; Province & Rouder, 2012) found exactly the opposite result namely discrete mediation of recognition decisions. Conditional independence describes the assumption of threshold models, that the conditional response probability and response distribution of a state should be invariant across different experimental conditions once that state was entered. Thus, this condition should hold for guessing responses out of a state of uncertainty for different strength manipulations. Regardless of the mean memory strength for items, guessing probabilities and response distributions should be constant. Province and Rouder (2012) tested the assumptions for confidence responses, Kellen et al. (2015) replicated those results and further evaluated them for response times, and finally Kellen and Klauer (2015) used a similar approach, addressing the confidence distribution of misses. In all studies, the assumption of conditional independence held which supports threshold models and the assumption os discrete states mediating recognition decisions but contradicts SDT which is not compatible with the finding of conditional independence. As can be seen from the literature reviewed above, the debate about the nature of information underlying recognition decisions is far from being solved.

1.3.1 Contribution of the present dissertation

Based on the variety of experimental and analytical approaches, my dissertation contributes to the still intense discussion on the continuous and discrete nature of memory information underlying recognition decisions by taking two distinct approaches. Within Manuscript 1 and 2, I assessed the model fit and validity within a currently neglected recognition paradigm. More precisely, in my first manuscript (Voormann et al., in press) we extended two models to *paired-word recognition*: general recognition theory (Ashby & Perrin, 1988), a multidimensional SDT, and the 2HTM. In a paired-word recognition task (Greene & Klein, 2004), participants study single words sequentially. However, during test two words are randomly paired and participants are asked to categorize these pairs as being either of type NN, two new words paired, NO, a new word on the left side paired with an old word on the right side, ON, an old word on the left side bing paired with a new word on the right side, or OO, two old words paired. In our studies, we selected the model describing paired-word recognition decisions best within and across the two model classes. Additionally, as the paradigm is quite new, we investigated the differences of paired-word recognition to single-word recognition to get a broader understanding of processes involved in paired-word recognition decisions.

As the models we developed for the paired-word recognition paradigm seemed to capture the fundamental mechanisms quite well, in my second manuscript (Voormann et al., unsubmitted manuscript) my co-authors and I went one step further and assessed the validity of those models testing for selective influence. Selective-influence studies implement a manipulation that is believed to affect only one or a few of the involved processes and test whether the effect of the manipulation is captured in the parameters describing exactly these cognitive processes (convergent validity) while leaving other parameters unaffected (discriminant validity; see, e.g., Erdfelder & Buchner, 1998; Jacoby, 1991; Jacoby, Lindsay, & Hessels, 2003; Snodgrass & Corwin, 1988). Thus, implementing a strength manipulation and a base-rate manipulation, we assessed whether both models developed for paired-word recognition covered the manipulation in a psychologically meaningful way.

My third manuscript (Voormann et al., 2021) employs distinct behavioral predictions to differentiate between threshold and continuous recognition models. In a preregistered adversarial collaboration, I investigated the mechanisms behind the error speed effect. Because the original study by Starns et al. (2018) contained some potentially confounding experimental procedures, we assessed whether the error speed effect was truly driven by misleading evidence or by an error-correction strategy that uses response times as a heuristic cue. Through a direct replication paired with an extension condition eliminating possible confoundings, this study helped to validate the conclusions drawn by Starns et al. (2018).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124–150. doi:10.1037/0033-295X.95.1.124

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81–99. doi:10.1037/h0029531

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*(4), 267–273. doi:10.1111/1467-9280.00254

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*(6), 1204–1212. doi:10.1037/a0033894

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi:10.3758/BF03210812

Blackwell, H. R. (1963). Neural theories of simple visual discriminations. *Journal of the Optical Society of America*, *53*(1), 129–160. doi:10.1364/JOSA.53.000129

Brandt, M. (2007). Bridging the gap between measurement models and theories of human memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*(1), 72–85. doi:10.1027/0044-3409.215.1.72

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*(3), 587–606. doi:10.1037/a0015279

Bröder, A., & Schütz, J. (2011). 'Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition': Correction to Bröder and Schütz (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1301–1301. Publisher: American Psychological Association. doi:10.1037/a0024783

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. doi:10.1037/0033-295X.106.1.180

Dubé, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 130–151. doi:10.1037/a0024957

Erdfelder, E., & Buchner, A. (1998). Comment: Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General*, *127*(1), 83–96. doi:10.1037/0096-3445.127.1.83

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. doi:10.1037/0033-295X.91.1.1

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Greene, R. L., & Klein, A. A. (2004). Does recognition of single words predict recognition of two? *The American Journal of Psychology*, *117*(2), 215–227. doi:10.2307/4149023

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. doi:10.1037/0033-295X.95.4.528

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*(1), 36–67. doi:10.1016/0022-2496(89)90003-5

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. doi:10.1016/0749-596X(91)90025-F

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638–644. doi:10.3758/BF03196526

Jeffreys, H. (1961). *Theory of probability* (3. ed.). Oxford: Clarendon Pr.

Kellen, D., Erdfelder, E., Malmberg, K. J., Dubé, C., & Criss, A. H. (2016). The ignored alternative: An application of Luce's low-threshold model to recognition memory. *Journal of Mathematical Psychology*. Special Issue in Honor of R. Duncan Luce, *75*, 86–95. doi:10.1016/j.jmp.2016.03.001

Kellen, D., & Klauer, K. C. (2011). Evaluating models of recognition memory using first- and second-choice responses. *Journal of Mathematical Psychology*, *55*(3), 251–266. doi:10.1016/j.jmp.2010.11.004

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1795–1804. doi:10.1037/xlm0000016

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*(3), 542–557. Publisher: American Psychological Association. doi:10.1037/a0039251

Kellen, D., & Klauer, K. C. (2018). Elementary signal detection and threshold theory. In J. T. Wixted & E. J. Wagenmakers (Eds.), *V. methodology* (4th ed., Vol. 5, pp. 161–200). The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience. doi:10.1002/9781119170174

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review*, *20*(4), 693–719. doi:10.3758/s13423-013-0407-2

Kellen, D., Singmann, H., Vogt, J., & Klauer, K. C. (2015). Further evidence for discrete-state mediation in recognition memory. *Experimental Psychology*, *62*(1), 40–53. doi:10.1027/1618-3169/a000272

Klauer, K. C., & Kellen, D. (2015). The flexibility of models of recognition memory: The case of confidence ratings. *Journal of Mathematical Psychology*, *67*, 8–25. doi:10.1016/j.jmp.2015.05.002

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111–130. doi:10.1016/j.jmp.2017.12.003

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*(3), 353–368. doi:10.1037/0022-3514.93.3.353

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review*, *70*(1), 61–79. doi:10.1037/h0039723

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. doi:10.1016/j.cogpsych.2008.02.004

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2018). Stimulus effects and the mediation of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(11), 1814–1823. doi:10.1037/xlm0000550

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2019). Task effects determine whether recognition memory is mediated discretely or continuously. *Memory & Cognition*, *47*(4), 683–695. doi:10.3758/s13421-019-00894-9

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1095–1110. doi:10.1037/0278-7393.28.6.1095

Pazzaglia, A. M., Dubé, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*(6), 1173–1203. doi:10.1037/a0033044

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, *109*(36), 14357–14362. doi:10.1073/pnas.1103880109

Raftery, A. E. (1986). Choosing models for cross-classifications. *American Sociological Review*, *51*(1), 145–146. doi:10.2307/2095483

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

Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 870–888. doi:10.1037/a0034954

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

Ratcliff, R., Smith, P. L., & McKoon, G. (2015). Modeling regularities in response time and accuracy data with the diffusion model. *Current Directions in Psychological Science*, *24*(6), 458–470. doi:10.1177/0963721415596228

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 305–320. doi:10.1037/0278-7393.33.2.305

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. doi:10.1037/0096-3445.117.1.34

Starns, J. J., Dubé, C., & Frelinger, M. E. (2018). The speed of memory errors shows the influence of misleading information: Testing the diffusion model and discrete-state models. *Cognitive Psychology*, *102*, 21–40. doi:10.1016/j.cogpsych.2018.01.001

Swets, J. A., Tanner, W. P. J., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301–340. doi:10.1037/h0040547

Tanner, W. P. J., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401–409. doi:10.1037/h0058700

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 582–600. doi:10.1037/0278-7393.26.3.582

Voormann, A., Rothe-Wulf, A., Starns, J. J., & Klauer, K. C. (2021). Does speed of recognition predict two-alternative forced-choice performance? Replicating and extending Starns, Dubé, and Frelinger (2018). *Quarterly Journal of Experimental Psychology*, *74*(1), 122–134. doi:0.1177/1747021820963033

Voormann, A., Spektor, M. S., & Klauer, K. C. (unsubmitted manuscript). Do models of paired-word recognition capture experimental manipulations the way they are supposed to do? A validation study.

Voormann, A., Spektor, M. S., & Klauer, K. C. (in press). The simultaneous recognition of multiple words: A process analysis. *Memory & Cognition*.

Wickens, T. D. (2002). *Elementary signal detection theory.* Oxford: Oxford University Press.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. doi:10.1037/0278-7393.20.6.1341

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition*, *5*(4), 418–441. doi:10.1006/ccog.1996.0026

5 General discussion

> Dear Hilde,
> if the human brain was simple enough for us to understand, we would still be so stupid that we couldn't understand it.
>
> (from *Sophie's world* by *Jostein Gaarder)*

Resting upon the results and implications of all three manuscripts, my dissertation perfectly reflects and extends the shifting discussion on the discrete or continuous nature of memory information in recognition decisions. To be more precise, the evidence provided by the reported experiments is mixed with some suggesting the necessity of continuous information under certain conditions and in specific tasks while others allow for discretization of the underlying information.

The first and second manuscript show that even in complex situations, as it is the case for multiple simultaneous recognition decisions, threshold models can account for the observed behavior in a meaningful and parsimonious way. For the paired-word recognition task, the threshold model was able to identify the differences to a simple recognition task, to capture the interactions between recognition decisions consistently on the decisional level, and acted as expected, at least to some extend, in selective influence studies. Overall in the paired-word recognition task, the threshold model outperformed the continuous model, which attributed interactions less consistently to either mnemonic or decisional processes and seemed to be less parsimonious compared to the threshold model.

Conversely, in the case of the error speed effect considered within my third manuscript, it seems to be the other way around. Although we eliminated strategically confounds of error correction instructions and feedback in an extended condition, the error speed effect still occurred. This supports the idea that participants use continuous information, because models assuming access to a continuously graded memory signal can naturally account for the occurrence of error speed effects whereas only some threshold models are capable of explaining them without additional assumptions. The occurrence of the error speed effect also reflects the necessity of a mechanism that accounts for misremembering (as a result of either misleading evidence or mistaken detection). All continuous models incorporate such a mechanism because memory evidence informs a decision process or maps directly on response decisions so that memory evidence can inform the incorrect decision. However, only certain types of threshold models, namely low threshold models, include a discrete state for misremembering and therefore are able to account for the error speed effect.

5.1 Discrete and continuous information underlying recognition decisions

These results are probably not overly surprising as they are perfectly in line with the mixed results of previous studies summarized in the Introduction. Thus, these and previous results foster the necessity to put aside the question of *whether* recognition decisions base on either discrete or continuous information, but rather to question *when* and *under which conditions* certain types of information are required to enable optimal performance (Kellen & Klauer, 2015; McAdoo, Key, & Gronlund, 2018, 2019).

Looking at the question of *under which conditions*, the results of my three manuscripts allow for two possible sources: 1) the response components and 2) the type of information required for an adequate response. While my first and second manuscript analyzed recognition performance on the level of response frequencies, the third manuscript considered response latencies as well when investigating the error speed effect. Thus, when taking into account only these three manuscripts, one might jump to the premature conclusion that RTs reflect the continuous amount of memory information while examining only response categories might favor a discrete mechanism. However, evidence from ranking tasks, which only ask for a ranking of items and do not consider response latencies, also favor continuous models (Kellen & Klauer, 2014). This contradicts the assumption that the response components reflect different paths of processing, continuous vs. discrete.

When examining the type of information necessary for participants to solve the task as accurate as possible, the models that fit the data best perfectly match the type of information needed. Consider first the case of paired-word recognition. Here participants are asked to provide a simple joint categorization judgment. Thus, there seems to be no need for participants to base their responses on more complex continuous information and accordingly threshold models have a better fit compared to continuous models. However, this looks differently for the error speed effect. To evaluate the occurrence of the effect, only such trials are investigated in the subsequent two-alternative forced-choice task in which both words are hard to discriminate as they have previously elicited the same response (either "studied" or "not studied"). Consequently, to solve the task accurately more fine grained information is necessary to discriminate the target from the lure. Therefore, it seems natural that individuals base their responses on continuous information in such situations.

The idea is not new that the type of information required for an adequate response determines whether threshold or continuous models can account better for the behavioral findings. Based on previous findings in the field of recognition memory, Kellen and Klauer (2015) already discussed that depending on task characteristics individuals might act differently upon the available mnemonic information. McAdoo et al. (2019) fostered this conclusion showing within participants that depending on the task, a confidence rating task or a ranking task, either threshold models or continuous models account for the observed data. Additionally, McAdoo et al. (2018) found that even within the same task memory mechanisms can depend on stimulus composition. Considering related lures, more fine grained knowledge about memory evidence is necessary to respond accurately than when words are easily discriminable as in case of unrelated lures. Hence, the results argue for continuous models in the first case (related lures) but for threshold models in the second case (unrelated lures).

Building upon those results, McAdoo et al. (2019) proposed a control mechanism that regulates whether recognition decisions are mediated continuously or discretely based on internal (participant-specific) and external (task-specific) information in order to maintain efficient responding. They assume that discrete mediation of memory evidence is preferred over continuous mediation as it is less demanding on cognitive resources. Consequently, once participants evaluate their responding based on metacognitive judgments to be inadequate, control processes interrupt and shift from a discrete processing mode to a continuous mediation of memory evidence.

Grouping the literature according to the demands of information needed to solve the task, the previously mixed evidence seems to resolve itself into a plausible pattern. Binary response tasks (Bröder & Schütz, 2009, 2011; Kellen, Klauer, & Bröder, 2013) and the paired-word recognition task (Voormann, Spektor, & Klauer, unsubmitted manuscript, in press) require on most of their trials only basic information and thus threshold models are favored. For confidence ratings (Klauer & Kellen, 2015) and the four-alternative forced-choice task (Kellen & Klauer, 2011) however, more fine grained information is needed. Possibly even the type of information used, either a discrete recollection or a continuous familiarity signal, might depend on trial specific information, thus dual-process models can account for these tasks best. Finally, to solve ranking tasks (Kellen & Klauer, 2014) and the two-alternative forced-choice trials considered for the error speed effect (Starns, Dubé, & Frelinger, 2018; Voormann, Rothe-Wulf, Starns, & Klauer, 2021), continuous information is necessary and consequently continuous models are favored.

5.2 The role of guessing

What the accounts mentioned above neglect, however, is a bunch of research hinting towards a second question that is necessary to raise and crucial to evaluate separately from the discussion about the nature of memory information. This question addresses the necessity of evidence free guessing for recognition decisions. While neither continuous models nor process models of recognition memory incorporate a guessing mechanism, most threshold models assume an uncertainty state out of which responses are guessed entirely free of memory evidence and based solely on contextual information.

Evidence in favor of the existence of feature-unrelated guessing exists for example in the area of working memory. For the demonstration of its existence, Rouder, Thiele, Province, and Cusumano (unpublished manuscript) used a working memory task in which participants were asked to remember the position of an arrow angle in a circle. Thus, the deviance of the response and the actual position was defined continuously. They showed that additionally to finely encoded representations in which memory matched the actual stimulus well and coarsley encoded representations in which only the correct side but not angle was memorized there exist trials in which responses are best characterized by feature-unrelated guessing. In those guessing trials, responses spread uniformly distributed across all angles of the circle indicating no knowledge about any stimulus features.

Additionally, the line of research indicating conditional independence within confidence ratings (Kellen & Klauer, 2015; Kellen, Singmann, Vogt, & Klauer, 2015; Province & Rouder, 2012) basically demonstrates the same: the existence of information free guessing within recognition decisions. As a reminder, conditional independence describes a central assumption of threshold models asserting that state-response mappings are independent from the conditional probability to enter a certain state. Kellen and Klauer (2015) investigated it by inspecting the distribution of incorrect responses on a confidence scale. They proved that SDT predicts fewer errors for strong targets and, more importantly, less extreme errors compared to weak targets. Whereas, within the 2HTM response distributions for errors should not depend on a strength manipulation. Assuming however, a general existence of evidence free guessing the assumptions of the two models become quite similar. Because response distributions from guessing should not vary for a strength manipulation, as no memory information is available, the predictions for error distributions only deviate to the amount of misdetection predicted by SDT. As the same line of reasoning basically holds for the other studies investigating conditional independence (Kellen et al., 2015; Province & Rouder, 2012), future research should address the necessity of guessing without memory evidence in more detail.

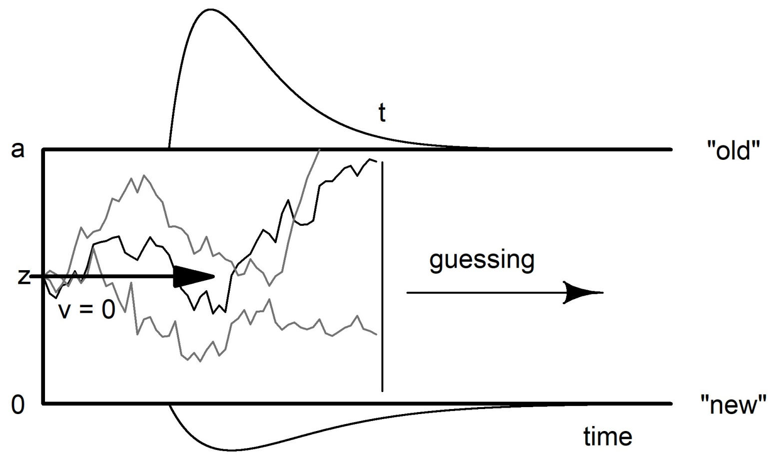

Figure 5.1: Schematic illustration of a diffusion model with a guessing mechanism at a certain point in time for a trial with drift rate equal to the drift criterion and a point in time for guessing. Note however, that response distributions still represent those of all targets.

5.2.1 Guessing within continuous models

The introduction of a guessing mechanism to continuous models of recognition memory does not necessarily contradict the general structure of those models. Rather it expands on their structure and specifies behavior in boundary conditions.

Taking for example the diffusion model, it is generally possible that for certain stimuli there exists an equal amount of evidence towards both decisions. Mathematically, this occurs for a drift rate being equal to the drift criterion. However, in such a case no response criterion can be reached based on the memory evidence and decisions are simply based on noise.[1] There exist some workarounds and model adaptations to enforce responses with increasing time as, e.g., a model with collapsing boundaries (but see Voskuilen, Ratcliff, & Smith, 2016; Voss et al., 2019). Collapsing boundaries represent the assumption that with increasing time, less evidence is required for a decision and the speed-accuracy trade-off shifts away from accuracy. However, it would also be reasonable to assume that in such a situation at a certain point in time *t* participants simply guess their response (see Figure 5.1).

Incorporating a guessing mechanism into SDT differs slightly from the logic of the diffusion model. Because familiarity maps directly onto a decision dimension and not just informs a decision process, guessing has to be depended on a certain amount of

[1]Please note that even for a drift rate equal to zero, evidence will reach at a certain time one of the two criteria due to the within-trial variability in the accumulation process. For a starting point being located exactly in the middle between the two criteria, each decision will occur equally often (Voss, Lerche, Mertens, & Voss, 2019).

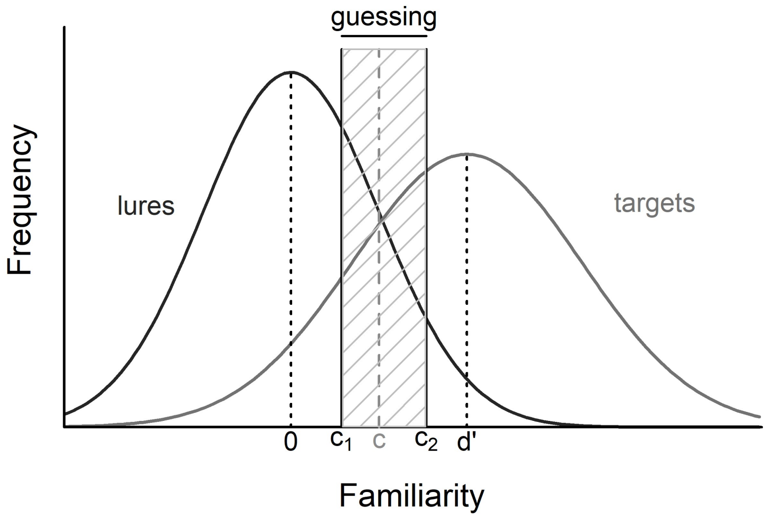

Figure 5.2: Illustration of signal detection theory for a recognition task with an area of uncertainty in which participants simply guess instead of relying on memory evidence

familiarity. For SDT, an amount of familiarity equal to the criterion represents evidence to neither decision. Therefore, in such a (although somewhat unlikely) case guessing would be necessary. In addition, research using confidence criteria suggests that even an area around the criterion might comprise uncertain responses. Thus, it seems natural to define a familiarity range in which participants might just simply guess instead of relying on the familiarity to retrieve a decision. For modeling, this would basically imply two criteria, familiarity values below the first criterion c_1 map to a "new" decision while familiarity values above the second criterion c_2 elicit an "old" response (see Figure 5.2). However for all stimuli with familiarities between those criteria responses are guessed with a certain probability.

Please note for all cases that guessing does not mean to discard the information available, but simply that responses are not driven by the amount of memory evidence. Thus, guessing can still depend on contextual information as base-rates and response incentives just like the responses out of a state of uncertainty within threshold models. Additionally, incorporating guessing still allows for misremembering of items within continuous models. For example, a target can lead towards the incorrect response by being smaller than the first criterion in the case of SDT or by having a drift rate leading towards the incorrect response boundary within the diffusion model. Misremembering has been shown to be necessary for the error speed effect (Starns et al., 2018; Voormann et al., 2021), and thus, needs to be accounted for by recognition memory models. However, evidence free guessing seems to be another essential response mode within recognition decisions and thus future models should consider it in their model structure as well.

5.3 Implications for process models

Based on the insights of the Introduction, the previous discussion of the various measurement models provides important implications for the specification of retrieval mechanisms in process models. Taken together, four fundamental mechanisms are vital for the extension of process models in recognition memory:

1) dependencies within recognition decisions
2) a mechanism for misremembering
3) context dependent adaptation of information available for responding
4) incorporation of guessing

First, my first two manuscripts (Voormann et al., unsubmitted manuscript, in press) indicated that most process models only incorporate mechanisms for basic recognition paradigms. Although there exist already adaptations for associated word pairs (Humphreys, Pike, Bain, & Tehan, 1989) and adaptations for conditional hypothesis testing (Dougherty, Gettys, & Ogden, 1999), more complex mechanisms as dependencies within decisions to paired-word recognition trials are not incorporated so far. However, based on the extended measurement models selected for the paired-word recognition paradigm, the specific mechanisms of such dependencies are known and thus can be established in process models. For example, mnemonic spill-over effects of activation can be included as a relative spread of activation from one recognition process to the other in case of two simultaneous recognition decisions.

Second, a mechanism of misremembering, derived from my third manuscript (Voormann et al., 2021) and recent studies by Starns et al. (2018) and Starns and Ma (2018), is actually unproblematic as all process models already incorporate the possibility of misremembering. Contrarily to the high threshold assumption within some threshold models but in line with continuous models, process models typically assume that target probes can lead into an activation which is below the criterion and thus elicits the incorrect response based on memory evidence (cf. Gillund & Shiffrin, 1984; Hintzman, 1988). This occurs for example when context changes between study and test or encoding has not been successful. Thus, just as continuous models, process models can naturally account for misremembering due to their structure.

Third, the type of information available for responding depends on the features of the stimulus material and the selected task. Thus, it seems to be crucial for process models to adapt their decision mechanisms in such a way that depending on task features either discrete or continuous information about memory activation are retrieved for responding. As continuous information can be mapped easily onto discrete states and threshold theories do not contradict a continuous evidence accumulation that results into certain states, only a mechanisms controlling the type of information available for responding needs to be considered.

And finally, process models currently do not incorporate the possibility of guessing. However, as most process models resemble SDT in that they map the amount of evidence directly to a response, the same mechanism suggested for SDT can be applied. Incorporating two criteria and an area of uncertainty, they can easily be adapted to describe evidence free guessing.

Thus, as can be seen by those four examples, investigating recognition decisions via measurement models yields a valuable simplification to assess retrieval mechanisms within recognition decisions. Nevertheless, it only offers insights into certain mechanisms while adaptations of the process models are still essential to account for the whole memory process.

5.4 Final Thoughts

The debate about the nature of information underlying recognition decisions and decisions in general has a long tradition. Given its length, it is fascinating that scientists still find new creative ways of exploring, testing, and challenging it. However, recently the discussions seems to shift away from the fundamental discussion of the discrete and continuous nature of cognitive information towards both a more context driven account and a more process oriented way of thinking. This development starts with questioning *under which conditions* certain information is gathered and acquired for responses (Kellen & Klauer, 2015; McAdoo et al., 2018, 2019) and it continues with the query of the necessity of misremembering (Starns et al., 2018; Starns & Ma, 2018; Voormann et al., 2021) or, as I argued above, evidence free guessing. To open up the discussion about the nature of recognition decisions towards, on the one hand, context depended properties and, on the other hand, more global processes involved in recognition decisions will bear interesting new insights about recognition memory in future research.

References

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*(3), 587–606. doi:10.1037/a0015279

Bröder, A., & Schütz, J. (2011). 'Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition': Correction to Bröder and Schütz (2009). *Journal of Experimental Psychology: Learning, Mem-*

ory, and Cognition, *37*(5), 1301–1301. Publisher: American Psychological Association. doi:10.1037/a0024783

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. doi:10.1037/0033-295X.106.1.180

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. doi:10.1037/0033-295X.91.1.1

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. doi:10.1037/0033-295X.95.4.528

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*(1), 36–67. doi:10.1016/0022-2496(89)90003-5

Kellen, D., & Klauer, K. C. (2011). Evaluating models of recognition memory using first- and second-choice responses. *Journal of Mathematical Psychology*, *55*(3), 251–266. doi:10.1016/j.jmp.2010.11.004

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1795–1804. doi:10.1037/xlm0000016

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*(3), 542–557. Publisher: American Psychological Association. doi:10.1037/a0039251

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review*, *20*(4), 693–719. doi:10.3758/s13423-013-0407-2

Kellen, D., Singmann, H., Vogt, J., & Klauer, K. C. (2015). Further evidence for discrete-state mediation in recognition memory. *Experimental Psychology*, *62*(1), 40–53. doi:10.1027/1618-3169/a000272

Klauer, K. C., & Kellen, D. (2015). The flexibility of models of recognition memory: The case of confidence ratings. *Journal of Mathematical Psychology*, *67*, 8–25. doi:10.1016/j.jmp.2015.05.002

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2018). Stimulus effects and the mediation of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(11), 1814–1823. doi:10.1037/xlm0000550

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2019). Task effects determine whether recognition memory is mediated discretely or continuously. *Memory & Cognition*, *47*(4), 683–695. doi:10.3758/s13421-019-00894-9

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, *109*(36), 14357–14362. doi:10.1073/pnas.1103880109

Rouder, J. N., Thiele, J. E., Province, J. M., & Cusumano, M. (unpublished manuscript). The evidence for a guessing state in working-memory judgments. Retrieved from http://pcl.missouri.edu/sites/default/files/p_0.pdf

Starns, J. J., Dubé, C., & Frelinger, M. E. (2018). The speed of memory errors shows the influence of misleading information: Testing the diffusion model and discrete-state models. *Cognitive Psychology*, *102*, 21–40. doi:10.1016/j.cogpsych.2018.01.001

Starns, J. J., & Ma, Q. (2018). Guessing versus misremembering in recognition: A comparison of continuous, two-high-threshold, and low-threshold models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 527–539. doi:10.1037/xlm0000461

Voormann, A., Rothe-Wulf, A., Starns, J. J., & Klauer, K. C. (2021). Does speed of recognition predict two-alternative forced-choice performance? Replicating and extending Starns, Dubé, and Frelinger (2018). *Quarterly Journal of Experimental Psychology*, *74*(1), 122–134. doi:0.1177/1747021820963033

Voormann, A., Spektor, M. S., & Klauer, K. C. (unsubmitted manuscript). Do models of paired-word recognition capture experimental manipulations the way they are supposed to do? A validation study.

Voormann, A., Spektor, M. S., & Klauer, K. C. (in press). The simultaneous recognition of multiple words: A process analysis. *Memory & Cognition*.

Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. doi:10.1016/j.jmp.2016.04.008

Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin & Review*, *26*(3), 813–832. doi:10.3758/s13423-018-1560-4

Acknowledgments

Freundschaft ist, wenn dich einer für gutes Schwimmen lobt, nachdem du beim Segeln gekentert bist.

(Werner Schneyder)

Danken möchte ich auf Deutsch, um sicher zu gehen, dass keine sprachlichen Hürden dem Dank im Wege stehen...

An erster Stelle möchte ich meinem Betreuer *Karl Christoph Klauer* danken. Danken, für Ihre immerzu offene Tür bei Fragen und Problemen, Ihre Geduld, mir Methoden, Modelle und Skripte genau zu erklären, Ihre Unterstützung bei all meinen Projekten, sowohl durch inhaltliche Diskussionen als auch durch organisatorische Anleitung, und Ihr stetes Interesse am Stand und Vorankommen unserer gemeinsamen Projekte. Ohne Sie wäre diese Forschung und Promotion nie möglich gewesen.

Viele neue inhaltliche Impulse, die mir stets halfen meinen Blickwinkel auf meine Untersuchungen zu ändern und Befunde gänzlich neu zu betrachten, habe ich durch *Andrea Kiesel*, meiner Zweitbetreuerin in SMiP, und *Arndt Bröder*, meinem Drittbetreuer, erfahren. Dafür möchte ich euch meinen herzlichen Dank aussprechen.

Ein großer Dank gilt ebenso meinen Kolleginnen und Kollegen: *Annelie Rothe-Wulf*, dir möchte ich besonders für deine stets unterstützende, offene und aufmunternde Art danken, die mir immer wieder Mut zum Weitermachen gab. Vielen Dank auch für deine Anmerkungen zu dem ersten Entwurf dieser Arbeit. *Mikhail S. Spektor* danke ich vor allem für seine Unterstützung bei methodischen Anliegen sowie sein sprachliches Geschick, mit dem er ungemein zur guten Lesbarkeit und Verständlichkeit meiner Manuskripte beigetragen hat. Bei *Lea Johannsen* und *Danielle Pessach* möchte ich mich für die vielen Diskussionen und Ratschläge über Wissenschaft, Gesellschaft und das Leben bedanken, die immer eine willkommene Abwechslung im Arbeitsalltag waren. *Raphael Hartmann* und *Constantin Meyer-Grant* danke ich für die Treue beim gemeinsamen Mittagessen (pünktlich um 12!) und *Constantin*, dir ebenso für die Anmerkungen zur ersten Version dieser Dissertation sowie zum inhaltlichen Austausch über meine Projekte; *Lydia Puffe*, *Manuel Becker*, *Rainer Leonhart*, *André Aßfalg* und *Jimmi Calanchini*, für die immer offenen Türen bei Fragen und Problemen bezüglich Lehre, Technik und Organisation. Und zu guter Letzt möchte ich den vielen *Hiwis* dieser Abteilung danken, die meine Experimente getestet und alle Versuchspersonen erhoben haben.

Dafür, dass ich mich egal wo im "SMiPLand" immer wie zu Hause gefühlt habe, möchte ich den anderen "SMiPstern" danken. Da es zu viele sind, um sie alle namentlich zu nennen, geht ein gemeinsamer Dank für eure Offenheit, den steten Austausch und die viele gemeinsame Zeit, die ich mit euch bei Workshops, Konferenzen und am Abend verbringen durfte, an euch. Auch bin ich dankbar für die Möglichkeiten, die ich durch SMiP erhalten habe, für die Vernetzung zu unzähligen WissenschaftlerInnen innerhalb von Deutschland aber auch international, die Anregungen durch viele spannende Workshops und die Möglichkeit immer wieder neue Dinge zu lernen.

Nach der Arbeit in ein Zuhause kehren zu dürfen, in dem man sich wohl fühlt und zu jeder Zeit wilkommen geheißen wird, half mir, auch nach stressigen Tagen den nötigen Abstand zu gewinnen. Dafür, dass ihr mit eurer fürsorglichen geselligen Art es geschafft habt, dass ihr mehr ein Familienersatz als eine Wohngemeinschaft für mich seid, bin ich euch, *Kerstin* und *Sabrina*, aus vollem Herzen dankbar.

Für ein gutes Maß an Ablenkung von all den wissenschaftlichen Problemen und erneutes Kraftschöpfen für neue Herausforderungen haben auch stets meine Freundinnen und Freunde gesorgt. Allen voran *Kathrin*, die in meiner ganzen Zeit in Freiburg eine beste Freundin für mich war und ist. Danke, dass du einfach bist wie du bist und mich so akzeptierst wie ich bin. Aber natürlich auch *Basti*, der mir seit dem Bachelor ein treuer Freund ist, *Doro* und *Susanne*, die mir immer wieder zu neuen Luftsprüngen verhalfen und *Larissa*, *Johanna*, *Philipp* und *Alina*, die mir unzählige Male halfen mein Segelboot in den Wind zu stellen und den Kurs neu zu suchen, danke euch allen. Und natürlich möchte ich dir, lieber *Philipp*, auch für die vielen gemeinsamen Stunden, die ich mit dir in deiner Hängematte verbringen durfte (und hoffentlich darf), danken, die Zeit, in der Worte nicht notwendig waren um sich zu verstehen, und für deine ruhige, geduldige Art, mit der du mir immer wieder begegnest und mich erdest, besonders jetzt in dieser Endphase.

Ohne eine *Familie* im Rücken zu wissen, die mich stets unterstützt in allem was ich tue, die immer für mich da ist, wenn ich sie brauche, und die mich bedingungslos liebt, hätte ich diese Reise nie begonnen. Dafür, dass ihr mir das Fliegen beigebracht habt, sodass ich mich gut gerüstet auf dieses Abenteuer einlassen konnte, immer mit dem Wissen um einen sicheren Hafen, der mich in eurer Mitte erwartet, wenn ich doch mal vom Weg abkommen sollte oder einfach nur des Reisens müde bin, dafür bin ich euch aus tiefstem Herzen dankbar.

Author and Co-author contribution

Manuscript I

Voormann, A., Spektor, M.S., & Klauer, K.C. (in press). The simultaneous recognition of multiple words: A process analysis. *Memory & Cognition*.

For this manuscript, I devised the theoretical background, programmed the second experiment, supervised data collection, helped developing the models and implemented all nested models, performed the analyses, wrote the first draft, and contributed to improving and revising the manuscript. Mikhail S. Spektor provided suggestions for the model comparisons, supervised data analysis, provided recommendations for structuring and improving the manuscript. Christoph Klauer contributed to the manuscript through developing and refining the theoretical background, programming the first experiment, developing and implementing the parent models, supervising model selection and data analyses, and improving and revising the manuscript.

Manuscript II:

Voormann, A., Spektor, M.S., & Klauer, K.C. (unsubmitted manuscript). Do models of paired-word recognition capture experimental manipulations the way they are supposed to do? A validation study.

I contributed to this manuscript by conceiving the theoretical background, planning and programming both experiments, supervising data collection, helping to develop and implementing all models, performing the analyses, writing the first draft, and contributing to the improvement and revision the manuscript. Mikhail S. Spektor provided suggestions for the experimental design and model comparison strategies, supervised data analysis, provided recommendations for structuring and improving the manuscript. Christoph Klauer contributed to this manuscript through suggestions on the experimental design, supervising model specification, model selection, and data analyses, and improving and revising the manuscript.

Manuscript III:

Voormann, A., Rothe-Wulf, A., Starns, J. J., & Klauer, K.C. (2021). Does speed of recognition predict two-alternative forced-choice performance? Replicating and extending Starns, Dubé, and Frelinger (2018). *Quarterly Journal of Experimental Psychology*, *74*(1), 122–134. doi:10.1177/1747021820963033

For this manuscript, I prepared the theoretical background, provided suggestions for the experimental design, performed the analyses, wrote the first draft, and contributed to improving and revising the manuscript. Annelie Rothe-Wulf provided suggestions for the experimental design, programmed the experiment, supervised data collection, prepared the data, and provided recommendations for improving the manuscript. Jeffrey J. Starns contributed to the manuscript by providing part of the data analysis, drafting the respective result section, and improving the manuscript. Christoph Klauer developed and refined the theoretical background, provided suggestions for the experimental design, supervised data analyses, and improved and revised the manuscript.

Funding:

All three manuscripts within this dissertation were supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to the Research Training Group "Statistical Modeling in Psychology" (GRK 2277).

Preparation of the third manuscript was additionally supported by grant no. 1454868 from the National Science Foundation awarded to Jeffrey J. Starns.