Judgments of frequency and duration: One or two underlying dimensions?

Johannes Titz[a]   Peter Sedlmeier[b]

Abstract

How *often* and for how *long* a certain object was encountered are two distinct questions, requiring two distinct answers. But when humans judge an object to have been presented often, they usually also judge it to have been presented for a long duration. Under certain conditions the reverse is also true: When an object has been presented for a long time, humans think it has been presented often. These mutual influences hint at a single underlying dimension on which judgments of frequency and duration are based. But the judgments also differ in certain respects, which instead speaks for two independent dimensions. In this article, from a novel formal perspective, we revisit arguments for and against unidimensional and two-dimensional models. To determine which of these models is supported empirically, we analyzed all available studies and tested an overlooked but critical condition that would allow us to draw a definitive conclusion. Our evidence speaks against a unidimensional model in all variants.

Keywords: judgment of frequency, judgment of duration, dual path, common path, total time hypothesis, state trace analysis

---

[a] Department of Psychology, Research Methods and Evaluation in Psychology, Chemnitz University of Technology, 09107 Chemnitz, Germany. E-mail: johannes@titz.science. https://orcid.org/0000-0002-1102-5719.

[b] Department of Psychology, Research Methods and Evaluation in Psychology, Chemnitz University of Technology. https://orcid.org/0000-0002-7998-6313

1 Introduction

In the natural environment the exposure frequency and exposure duration of homogeneous events are often correlated. For instance, when one has to wait at *many* traffic lights during a stroll through town, one has probably waited quite a *long* time in total. If one knows either the frequency or total duration of the waiting situations, one can estimate the other attribute fairly accurately.[1] In the laboratory, the exposure frequency and duration of stimuli can be varied orthogonally, and so accurate judgments of these attributes would be uncorrelated. But humans still produce correlated judgments: When they think an object has been presented often, they usually also judge it to have been presented for a long duration (e.g., Betsch et al., 2010; Hintzman, 1970; Hintzman et al., 1975; Smith et al., 2017; Winkler et al., 2015). Under certain conditions the reverse is also true: when an object has been presented for a long time, humans think it has been presented often (e.g., Betsch et al., 2010; Bonanno, 1996; Hintzman et al., 1975; Winkler et al., 2015; Zhao & Turk-Browne, 2011). These findings hint at a single underlying dimension on which judgments of frequency and duration are based.

But at the same time, it is known that judgments of frequency are extremely accurate, while judgments of duration are usually poor in comparison (e.g., Betsch et al., 2010; Hintzman, 1970; Hintzman et al., 1975; Winkler et al., 2015). If the same single dimension is used for both judgments, the difference in accuracy is difficult to explain. But does it really demand two separate dimensions?

Half a century ago, Hintzman (1970) answered in the affirmative and concluded that judgments of frequency and duration must be based on two dimensions because they are affected differently by exposure frequency and duration (subsequently this was also expressed in Hintzman, 2011; Hintzman et al., 1975). But in the last decade new findings have accumulated and a unidimensional model has resurfaced again (e.g., Smith et al., 2017; Zhao & Turk-Browne, 2011).[2] Specifically, Betsch et al. (2010), as well as Winkler (2009) and Winkler et al. (2015), have favored a unidimensional model. They have found that under certain conditions, the biasing[3] effects of duration on judgments of frequency and of frequency on judgments of duration are relatively symmetrical.[4] In several experiments they manipulated how much attention participants paid to the stimuli during encoding. Under high-attention conditions (e.g., interesting material, special instructions), frequency influenced judgments of duration and duration also influenced judgments of frequency. In

---

[1]Note that here we are referring to the total duration and not to the average duration. The total duration is the sum of the durations of all occurrences of an event. If the duration of each occurrence is constant, there is a positive correlation between exposure frequency and total exposure duration.

[2]Note that Smith et al. (2017) discussed the unidimensional model only as one of several potential models.

[3]A bias in this context is an influence of the irrelevant independent variable on the judgment: frequency on judgment of duration or duration on judgment of frequency (as initially described in Betsch et al., 2010; Winkler et al., 2015).

[4]The terms symmetry and asymmetry were introduced by Betsch et al. (2010) and we use them accordingly. Symmetry means that the corresponding effects are of similar size, which was the main hypothesis in Betsch et al. (2010). Hintzman (1970) preferred to speak of "different patterns".

some conditions, these effects were comparable in size and this symmetrical pattern the authors took as evidence of a single underlying dimension.

Our goal in this paper is to resolve which of these opposing positions is valid: Are judgments of frequency and duration based on one or two dimensions? We reconstruct the arguments for and against both models formally, which shows that previous attempts to clarify this issue were methodologically unsatisfactory. We propose a different, more stringent method that is related to state trace analysis (Dunn & Kalish, 2018; Dunn & Kirsner, 1988). Finally, we gather available data and apply our proposed method to test which of the models is supported empirically.

2   Applying the Total Time Hypothesis to Judgments of Frequency and Duration

In a typical experiment on judgments of frequency and duration, participants observe many stimuli that are presented with varying frequencies and presentation durations. After the encoding phase, participants are asked to provide a numerical estimate of frequency and/or duration for each stimulus in retrospect.

This paradigm of studying the joint effects of exposure frequency and exposure duration on memory-related judgments was introduced by Hintzman (1970). He borrowed the idea of the total time hypothesis (Cooper & Pantle, 1967) in memory research and applied it to judgments of frequency. The total time hypothesis states that the long-term memory strength for a stimulus depends on the total exposure time of the stimulus, independent of how time is distributed (Hintzman, 1970, p. 437). This can be translated into a more formal statement:

$$m(f, d) = g_m(f \cdot d)$$

where m is memory strength, f and d are exposure frequency and exposure duration[5], and g_m is a function. The statement means that several repetitions of an item are like one long repetition, for example, $m(10, 1) = m(1, 10)$. We cannot say much about the function g_m because the memory strength is not measurable (at least not within the presented framework) and theorists have so far not been interested in it.

Judgments of frequency were conceptualized as memory-related judgments because they showed a spacing effect (Hintzman, 1969; Underwood, 1969). Thus, the transfer of the total time hypothesis from memory to judgment of frequency seems obvious in retrospect. Hintzman postulated that frequency judgments are based on the memory strength, which can be translated into

$$F(f, d) = g_F(g_m(f \cdot d))$$

where F is judgment of frequency and g_F is some unknown function. Again there is usually no explicit discussion of the shape of this function in the literature. Independent of the nature

---

[5]Note that d is the duration of a single occurrence, while $f \cdot d$ is the total duration of all f occurrences.

of the two functions (g_F and g_m), judgments of frequency should depend on the interaction between f and d, which Hintzman tested in his first experiment. His choice of analysis was an analysis of variance (ANOVA) with linear trends and a correlation analysis, which both constrain the relationships to be linear.[6] With this assumption we can equivalently state the expected relationship in regression form:

$$F(f, d) = b_{Ff \cdot d} f \cdot d + e$$

where $b_{Ff \cdot d}$ is the slope for the interaction term and e is the error term.[7] Note that in a model with the main effects and interaction specified, the interaction should be the best predictor, and the main effects b_{Ff} and b_{Fd} are expected to be 0 and thus are excluded from the above formula.

To his surprise, Hintzman found the interaction term was negligible in size, and the effect of frequency (b_{Ff}) almost completely explained the variation in the data, while the effect of duration (b_{Fd}) was essentially 0. His verbal summary corresponds to

$$F(f, d) = b_{Ff} f + e$$

He had two explanations for this finding. The first was that the total time hypothesis did not hold: Duration had no effect on the memory strength and thus no effect on judgment of frequency:

$$m(f, d) = g_m(f)$$

and

$$F(f, d) = g_F(g_m(f))$$

But if duration did not influence the memory strength, the total time hypothesis had to be fundamentally questioned, at a time when it was popular (Cooper & Pantle, 1967). Furthermore, it is not intuitive that seeing something for a longer period of time should have no effect on memory strength at all (e.g., Potter & Levy, 1969). Thus, Hintzman's alternative explanation was that the total time hypothesis was correct but not applicable to judgments of frequency. To be more specific, he assumed that judgments of frequency are not based on the memory strength. This idea required introducing the two-dimensional model: Exposure frequency and duration do not contribute to a single dimension of the memory trace, but to two independent dimensions (m_1, m_2):

$$m_1(f, d) = g_{m_1}(f)$$
$$m_2(f, d) = g_{m_2}(d)$$

---

[6]This assumption is inherent in Hintzman's arguments and in most arguments in the literature. We believe this is justified because Hintzman (1970) found that a logarithmic relationship could not explain additional variance. Other studies also showed that the relationship is fairly linear in the studied range (Betsch et al., 2010; Bonanno, 1996; Winkler et al., 2015). We take a closer look at the linearity assumption in the Discussion.

[7]For regression coefficients or correlations we use the coefficient with a subscript that first states the dependent and then the independent variable.

When judgments of frequency are made, the duration information is filtered out, so exposure duration has no effect on these judgments:

$$F(f, d) = g_F(g_{m_1}(f))$$

To test which of the two explanations was supported, Hintzman derived predictions of what would happen if participants had to make judgments of duration. If the total time hypothesis was incorrect, but all other assumptions valid, judgments of duration should be based on the same memory strength as judgments of frequency. Thus, they should be affected by frequency, but not by duration:

$$D(f, d) = g_D(g_m(f))$$

where D is judgment of duration and g_D is some unknown function. In this case the regression (or ANOVA with linear trends) should look like the one for judgments of frequency:

$$D(f, d) = b_{Df} f + e_D$$

with b_{Df} being the slope.[8] The crucial prediction here is that judgments of duration depend only on exposure frequency and not on exposure duration.[9]

The alternative was that the total time hypothesis was correct but not applicable to judgments of frequency. In this case, information about frequency and duration should be stored independently and therefore frequency information could be ignored when making judgments of duration:

$$m_1(f, d) = g_{m_1}(f)$$
$$m_2(f, d) = g_{m_2}(d)$$
$$D(f, d) = g_D(g_{m_2}(d))$$

and so a regression would result in

$$D(f, d) = b_{Dd}(d) + e_D$$

The crucial prediction here is that judgments of duration depend only on exposure duration and not on exposure frequency. This is the exact opposite of the first explanation, where judgments of duration depend only on exposure frequency and not on exposure duration. Thus, Hintzman's thought process leads to a decision experiment.

---

[8]Note that at this point Hintzman did not discuss whether the functions g_D and g_F are the same and so b_{Df} does not have to be identical to b_{Ff}. This makes sense because the range and scale of judgment of frequency and judgment of duration differ.

[9]Hintzman also suggested that judgments of duration might consist only of random noise. If participants know they have no information about duration, only about frequency, they would need to guess.

2.1   Interpreting the Results of Hintzman's Crucial Experiment

To his surprise, Hintzman (1970) found something in between the two opposing models: that b_{Dd} was not 0, but that neither was b_{Df}. Although one might think this result allows no definitive conclusion, Hintzman confidently rejected the unidimensional hypothesis. He studied r^2 for the effects of frequency/duration on judgment of frequency (.97/.003) and judgment of duration (.85/.123) and in the Discussion stated (p. 442): "Since the JOD data reveal a pattern that differs from that of JOF, the two cannot be based on identical information."

While the effect sizes evidently differ, it remains unclear how exactly the pattern contradicts a unidimensional model. A first attempt to understand Hintzman's reasoning might be to study the correlation between judgments of frequency and duration. If it is low, this would be a good reason to reject the unidimensional model. To derive the correlation between judgments of frequency and duration, the four reported effect sizes are sufficient. In the Appendix we give a formal argument for the following reasoning based on covariance algebra. The general idea can be demonstrated by applying Wright's well-known tracing rules (1934; for an introduction see Kenny, 1979) to the paths in Figure 1. Note that Hintzman used correlations for his argument, but we use standardized regression coefficients, which, in this case, are equivalent.
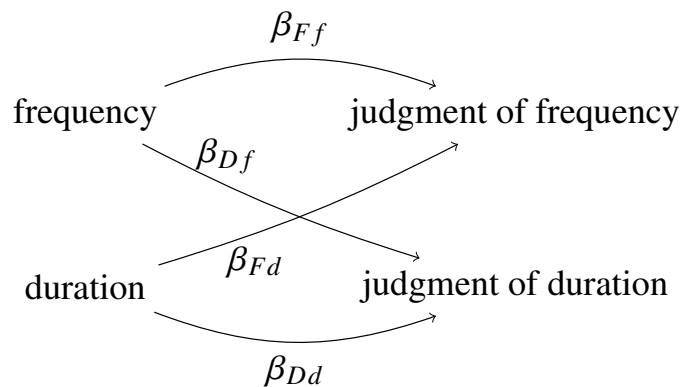
FIGURE 1: Path model for judgments of frequency and duration.

To find the correlation between judgments of frequency and duration one can start at either judgment and trace all paths to the other judgment. For instance, starting from judgment of frequency, there are two paths to judgment of duration[10]: (1) through frequency and (2) through duration. Both paths need to be added to find the total effect:

$$\mathrm{cor}(F, D) = \beta_{Ff}\beta_{Df} + \beta_{Fd}\beta_{Dd}$$

[10]Note that in a controlled experiment the correlation between f and d is 0 and so there is no path between frequency and duration. This is a crucial assumption, demonstrating the value of experimental data.

465

When calculating the correlation between the two judgments with Hintzman's data, one will find $\sqrt{.97 \cdot .85} + \sqrt{.003 \cdot .123} = .93$. With such a high correlation it is difficult to argue against a unidimensional model and for a two-dimensional model.

Another methodological approach to understanding Hintzman's reasoning might be dissociation. Indeed, exposure duration has an effect on judgment of duration (.123), but not on judgment of frequency (.003; compare with the interpretation in Betsch et al., 2010, p. 349). Often this constellation is seen as evidence of two independent dimensions. In general, this is a logical fallacy (Dunn & Kirsner, 1988). In Hintzman's specific case, adding the appropriate amount of noise to judgment of frequency would reduce the effect of duration to zero, without requiring the introduction of a second dimension. But this would only be possible if the effect of frequency would be larger on judgment of duration than on judgment of frequency, which it is not. In other words, the key point is that the effects of frequency and duration are differently ordered in size: $\beta_{Ff} > \beta_{Df}$, but $\beta_{Fd} < \beta_{Dd}$ ($\sqrt{.97} > \sqrt{.85}$, but $\sqrt{.003} < \sqrt{.123}$). This pattern is logically impossible with a single dimension.[11] To illustrate this, imagine a situation where we specify the coefficients between frequency/duration and memory strength, as in Figure 2.
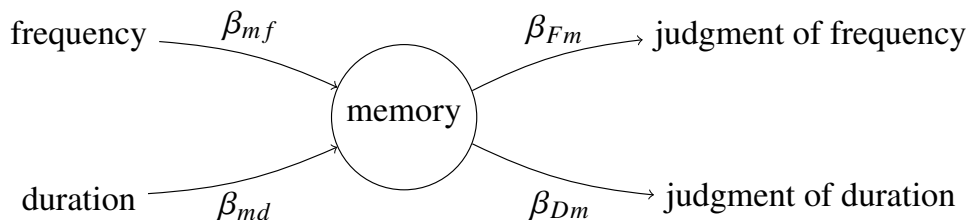
FIGURE 2: Hypothetical unidimensional model for judgments of frequency and duration.

The tracing rules allow us to express the coefficients between the independent and dependent variables via

$$\beta_{Ff} = \beta_{mf}\beta_{Fm} \quad \beta_{Fd} = \beta_{md}\beta_{Fm}$$
$$\beta_{Df} = \beta_{mf}\beta_{Dm} \quad \beta_{Dd} = \beta_{md}\beta_{Dm}$$

Now recall Hintzman's pattern of results:

$$\beta_{Ff} > \beta_{Df} \quad \text{and} \quad \beta_{Fd} < \beta_{Dd}$$

which can also be written as

$$\beta_{mf}\beta_{Fm} > \beta_{mf}\beta_{Dm} \quad \text{and} \quad \beta_{md}\beta_{Fm} < \beta_{md}\beta_{Dm}$$

---

[11]Empirically the pattern is possible since there is always error variance. Thus, a single study will not be sufficient for conclusive evidence. We discuss this problem at the end of the next section.

Assuming that β_{mf} and β_{md} are positive,[12] the equal terms cancel out and one has a contradiction:

$$\beta_{Fm} > \beta_{Dm} \quad \text{and} \quad \beta_{Fm} < \beta_{Dm}$$

Thus, in a unidimensional model, the coefficients have to have the same order: If $\beta_{Ff} > \beta_{Df}$, then $\beta_{Fd} > \beta_{Dd}$. This is violated in Hintzman's data, and assuming the result is not a coincidence, but reliable, there must be two dimensions.

2.2   Relationship to State Trace Analysis

The above deduction seems to come out of nowhere but is in fact just a special case of state trace analysis (for an introduction see Dunn & Kalish, 2018). The goal of state trace analysis is to determine the number of latent variables (here dimensions) required to explain empirical phenomena. The general idea is to theoretically constrain the type of relationship between independent/dependent variables and latent variables. The constraint leads to conditions under which certain empirical observations logically demand a certain number of latent variables. A typical scenario is to assume monotonic functions between a latent variable and two dependent variables (Dunn & Kirsner, 1988).[13] In this scenario, one can infer that the relationship between the dependent variables must also be monotonic, which can be tested empirically. If the relationship between the dependent variables is nonmonotonic, then there must be two latent variables.[14]

Testing the relationship between judgments of frequency and duration for monotonicity is not possible for Hintzman's (1970) study because the needed data is not available. State trace analysis requires means and ideally also standard deviations for all experimental conditions, but Hintzman reported mainly correlations and F values of ANOVAs. But these statistics can still be put to use: When the relationship between the variables is constrained to be linear, state trace analysis can be simplified, resulting in the test condition described above (section Interpreting the Results of Hintzman's Crucial Experiment).

If one assumes that all variables in Figure 2 are already standardized, memory strength (m) can be expressed as

$$m = \beta_{mf}f + \beta_{md}d$$

This is also referred to as the input mapping in state trace analysis, as it maps the independent variables to the latent variable. Mapping the latent variable to the judgments (output

---

[12]This assumption is rather trivial because all coefficients are expected to be positive from a theoretical perspective. Increases in exposure frequency and exposure duration should not lead to a decrease in memory strength. Note that if β_{mf} and β_{md} are both negative, the same argument holds, but the inequalities are reversed.

[13]Note that state trace analysis is a general method. Other relationships and other numbers of variables can be specified.

[14]An alternative is that the monotonicity assumption was wrong, but this assumption is made a priori and is reasonable in most psychological research.

mapping) results in

$$F = \beta_{Fm}m$$
$$D = \beta_{Dm}m$$

On substitution we can also write

$$F = \beta_{Fm}\beta_{mf}f + \beta_{Fm}\beta_{md}d$$
$$D = \beta_{Dm}\beta_{mf}f + \beta_{Dm}\beta_{md}d$$

In matrix notation this is

$$\begin{bmatrix} F \\ D \end{bmatrix} = \begin{bmatrix} \beta_{Fm}\beta_{mf} & \beta_{Fm}\beta_{md} \\ \beta_{Dm}\beta_{mf} & \beta_{Dm}\beta_{md} \end{bmatrix} \begin{bmatrix} f \\ d \end{bmatrix}$$
$$= \mathbf{B} \begin{bmatrix} f \\ d \end{bmatrix}$$

And here is the pivotal part. Note that the matrix $\mathbf{B}$ has rank 1. The rows and columns are multiples of each other with a ratio of $\frac{\beta_{Dm}}{\beta_{Fm}}$ and $\frac{\beta_{md}}{\beta_{mf}}$, respectively. Thus, either judgment can be expressed in terms of the other judgment. This simply means that a single dimension is sufficient to explain both judgments.

To find out whether the rows are independent or dependent, we can study the determinant of matrix $\mathbf{B}$, since a deficient matrix has dependent rows and columns. $\mathbf{B}$ can also be expressed as

$$\mathbf{B} = \begin{bmatrix} \beta_{Ff} & \beta_{Fd} \\ \beta_{Df} & \beta_{Dd} \end{bmatrix}$$

and the determinant is

$$|\mathbf{B}| = \beta_{Ff}\beta_{Dd} - \beta_{Fd}\beta_{Df}$$

If the determinant is zero, the judgments are dependent. If the determinant is nonzero, the judgments are independent. The critical condition described above, $\beta_{Ff} > \beta_{Df}$ and $\beta_{Fd} < \beta_{Dd}$, corresponds to a case where the determinant is positive because the terms of the determinant on the left (β_{Ff}, β_{Dd}) are both larger than the terms on the right (β_{Fd}, β_{Df}). A positive determinant is just a more specific case of a nonzero determinant.

It might be more useful to study the determinant directly because it is the more general condition. But error variance makes it unlikely one will find a determinant that is exactly zero, so one would need a more elaborate statistical framework to conduct a valid test. In contrast, testing the order of correlations is simple and straightforward, while not being less powerful. Furthermore, the boundary conditions are reasonable: All effects are expected to be positive and the effect of frequency should be larger on judgments of frequency than on judgments of duration. Thus, we have found a simple case of falsification. If $\beta_{Fd} < \beta_{Dd}$, the

unidimensional model is incorrect. The test itself does not favor this hypothesis: Technically, the opposite result $\beta_{Fd} > \beta_{Dd}$ is not less likely.

In light of the above derivation, Hintzman's (1970) implicit logical position is strong, but two problems remain. First, Hintzman probably unknowingly used the effect size r_{alerting} for his argument. This statistic is not the correlation between the independent and dependent variable on the individual level, but on the aggregated level (correlation of averages).[15] Since we are interested in individual behavior and not in group behavior, unaggregated data should be used for the analysis.

Second, his conclusion is based on a single study with a sample size of 116. For medium to large effects this might be sufficient to reach high precision, but the effect of duration is quite small and will vary substantially from experiment to experiment. Indeed, in Hintzman's (1970) third experiment, duration actually had a significant effect on judgment of frequency, but he argued that the effect was quite small and a potential artifact (p. 439). This is surprising because the effect had the same size as the effect of duration on judgment of duration (almost the same ANOVA F value with the same degrees of freedom), which Hintzman did not regard as spurious. In the fourth experiment, Hintzman was dealing only with free recall and in the discussion he clearly rejected the unidimensional hypothesis.

In later publications an effect of duration on judgment of frequency was usually found, although its size varied (e.g., Hintzman et al., 1975; Williams & Durso, 1986; Winkler, 2009; Zhao & Turk-Browne, 2011). Betsch et al. (2010) took up the task to explain why this effect varied and along the way resurrected the unidimensional hypothesis.

2.3   The Symmetry of Biases

Betsch et al. (2010) reframed the unidimensional hypothesis as the common-path hypothesis.[16][17] Their main argument was that biases in such a unidimensional model are symmetrical (or at least very similar) in size: $\beta_{Fd} = \beta_{Df}$. While this sounds intuitive, it is not required nor sufficient for a unidimensional model to be valid. Recall that the general test condition is still $|\mathbf{B}| = 0$ or $\beta_{Ff}\beta_{Dd} - \beta_{Fd}\beta_{Df} = 0$. From this alone it does not follow that $\beta_{Fd} = \beta_{Df}$.

---

[15]This is also why the correlations are much larger than what would be expected in psychological research.

[16]To be consistent we still use the term unidimensional in this section.

[17]Note that Betsch et al. (2010) changed the research design by focusing on total duration instead of single (average) duration. But without additional manipulations, the pattern of results remained the same: Judgments of duration were strongly biased by frequency, while judgments of frequency were not biased by duration (for a replication see Winkler, 2009). This might be one of the reasons why the difference between total and average duration has not yet received much interest in the literature. But it is worth noting that our theoretical derivations are applicable to total and average duration as long as duration is uncorrelated with frequency.

But if one starts from this assumption one will arrive at the required additional constraint:

$$\beta_{Ff}\beta_{Dd} - \beta_{Fd}^2 = 0$$
$$\beta_{Fm}\beta_{mf}\beta_{Dm}\beta_{md} - (\beta_{Fm}\beta_{md})^2 = 0$$
$$\beta_{mf}\beta_{Dm} - \beta_{Fm}\beta_{md} = 0$$
$$\frac{\beta_{mf}}{\beta_{md}} = \frac{\beta_{Fm}}{\beta_{Dm}}$$

Or verbally, the ratio of the effects of the independent variables on the memory strength is the same as the ratio of the effects of memory on the judgments. There appears to be no theoretical or empirical ground to assume that this constraint would be satisfied.

2.4   Conditional Unidimensional Hypothesis

Symmetrical biases might not be a good indicator of unidimensionality, but there is still substantial value in Betsch et al.'s (2010) work. When trying to produce symmetrical biases, they were mainly concerned with increasing the effect of duration on judgments of frequency. The influence of frequency on judgment of duration was well known (Hintzman, 1970; Hintzman et al., 1975; Winkler, 2009), but the influence of duration on judgment of frequency has only sometimes been found and has varied in size (e.g., Hintzman et al., 1975; Winkler, 2009). Betsch et al.'s (2010) main contribution to the field has been to explain this result and to find the conditions under which duration affects judgment of frequency. This attempt can also be seen as pointing out a fallacious dissociation: In some studies, duration had no effect on judgment of frequency only because the manipulation was too weak.

Betsch et al. (2010) postulated that the amount of attention paid to the stimuli is the critical variable. If a participant is required to attend to a stimulus for the whole presentation duration, memory strength will increase continuously and influence a single dimension in memory.[18]

In one experiment they instructed participants to press a key until the stimulus disappeared or to specifically pay attention to either frequency or duration. In another experiment, they used different types of material. Simple stimuli (e.g., words) might take only 2 s for complete processing (Hintzman, 1970), but more complex stimuli (e.g., pictures) can draw attention for a much longer duration. In several experiments, Winkler et al. (2015) studied the effect of arousal as a means of changing processing intensity. In all these studies, duration had an effect on both judgments if attention was high. Thus, biases in both directions were present, but they were not perfectly symmetrical in size. Betsch et al. (2010) concluded that neither the unidimensional nor the two-dimensional hypothesis was supported. They

---

[18]A similar idea was proposed by Cooper and Pantle (1967) to explain deviations from the total time hypothesis. They used the concept of "repeated representational response evocation", which can be defined as the perception of a given unit (Cooper & Pantle, 1967, p. 226), meaning that a stimulus (or whatever the unit might be, e.g., a stimulus pair) was processed once. If this response is evoked repeatedly until the stimulus disappears, the critical variable is total exposure duration.

suggested a conditional unidimensional hypothesis: Only under conditions of high attention will duration and frequency influence a single memory dimension.

One problem with this conceptualization is that it still requires two dimensions under conditions of low attention (when the biases are asymmetrical). Thus, the high-attention condition in the conditional unidimensional model might instead be seen as a case where the empirical data do not speak against a unidimensional model. Still the evaluation of this hypothesis by Betsch et al. (2010) rested on the symmetry criterion, which is not cogent. At the same time, the authors have shown that the relevant effects can vary substantially depending on the material and the attention of participants, so the pattern Hintzman (1970) obtained is literally just one of many.

Overall it remains unclear whether the vast amount of data from Betsch et al.'s (2010) research speaks for a (conditional) unidimensional or a two-dimensional model. Besides, there are other studies that provide enough information to re-evaluate whether judgments of frequency and duration are based on one or two dimensions. Our goal in the remainder of this paper is to gather these studies and test the critical condition, if $\beta_{Ff} > \beta_{Df}$ then $\beta_{Fd} > \beta_{Dd}$.

3 Methods

The critical test can be conducted for every experiment that manipulated frequency and duration and assessed judgments of frequency and duration. Since there are some differences between the relevant studies, it makes sense to perform the test for every single study instead of aggregating the studies first. This also makes it easy to test the conditional unidimensional model: If there is a subset of studies that are consistent with a unidimensional model, we can look for similarities between these studies.

3.1 Selection of Studies

We queried Google Scholar for the search terms frequency AND duration AND ["(judgment OR judgement) of frequency" OR "(judgment OR judgement) of duration"]. We expected to find all studies where both independent variables were manipulated and both of the judgments were assessed. In addition, we did a forward search for the two main studies in this field, the seminal study by Hintzman (1970) and the more recent study including four experiments with many conditions by Betsch et al. (2010). Overall, this resulted in 193 hits. After scanning the abstracts we had to exclude many studies that were unrelated to the topic at hand. A high number of false positives was expected a priori, since the terms frequency and duration can be found in many areas of science. After this first selection, 12 articles remained but we had to exclude some that did not match the independent or dependent variables in question.

471

Warm and McCray (1969) did not manipulate duration and frequency directly. The authors presented different words only once for 1 s. The words differed in their relative occurrence in the English language (familiarity) and their length (short or long). These manipulations deviate too much from the default procedure and therefore this study was excluded.[19] Huppert and Piercy (1977) assessed not judgment of frequency or duration as the dependent variable but recognition. Mo and Michalski (1972) used pair comparisons and subsecond durations, which deviates too much from other studies. Several researchers (Bonanno, 1996; Hintzman, 2004; Williams & Durso, 1986; Zhao & Turk-Browne, 2011) studied only judgments of frequency and not judgments of duration, so we could not conduct the critical test. In one study (Smith et al., 2017) the independent variables had only two levels, so the effect size is not a normal correlation and our approach is not applicable.

Four articles in peer-reviewed journals remained (Betsch et al., 2010; Hintzman, 1970; Hintzman et al., 1975; Winkler et al., 2015). In addition, we received several unpublished raw data sets from Winkler (2008). One might argue that unpublished data do not meet the quality standards of current scientific practices and thus should be excluded. But in our opinion the better approach is to include these studies and treat their quality as a moderating variable. In our specific case all these studies come from one source (Winkler, 2008), so a subgroup analysis would also resolve the problem. As we show, this is not necessary since the data are conclusive.

Another problem we needed to address is that some authors studied average duration while others studied total duration. In the former the duration of a single presentation is held constant, while in the latter the total duration of all presentations is held constant. This difference has been somewhat neglected in the literature, presumably because the research focus was mainly on the empirical and not on the theoretical side. In our analysis we integrated both types of duration manipulation and can use this information as a moderator variable. Again, this does not play a big role since the data are conclusive.

Table 1 summarizes the selected studies. Since Winkler et al. (2015) is a subset of Winkler (2009), we considered only the latter. Most of the articles report several experiments, which in turn consist of multiple experimental conditions. Overall, our analysis contains 152 effect sizes from 38 conditions out of 19 experiments.[20]

3.2 Attention

According to the attention hypothesis (Betsch et al., 2010), the duration of a stimulus will affect its memory strength only if the stimulus requires prolonged processing. A complex or relevant stimulus attracts attention, and processing can continue for a long time. In this case, duration matters, and according to the conditional unidimensional model (Betsch et al.,

---

[19]But note that judgments of duration were still positively biased by the frequency of word occurrence.

[20]Actually there were 20 experiments, but we treated Experiments 1 and 2 from Hintzman (1970) as one between-subjects experiment because the only difference between the experiments was the dependent variable (in Experiment 1 judgments of frequency and in Experiment 2 judgments of duration).

TABLE 1: Studies Included in the Analysis

Article	No. of experiments	Selected experiments	No. of effects
Hintzman (1970)	3	1–3	8
Hintzman et al. (1975)	3	3	4
Winkler (2008)	10	1–8	48
Winkler (2009)	5	1–5	48
Betsch et al. (2010)	3	1–3	44

Note. In the first two experiments Hintzman et al. (1975) did not assess judgments of duration, so they were excluded. In two studies Winkler (2008) did not assess the type of duration that was manipulated, which makes our approach inapplicable. These two studies were also excluded.

2010) a single dimension is used for judgments of frequency and duration. In the analysis we used stimulus type and other attention manipulations as a moderator.

The material presented to the participants differed between the experiments: words (e.g., first names or foreign language words), images for which the degree of arousal varied, or video scenes. Accordingly, we divided the visual material into three groups (words: 0, pictures: 1, videos: 2) to numerically represent the different degrees of attention the three types of stimuli presumably draw. Furthermore, in some experiments there were special manipulations to increase attention. For instance, Betsch et al. (2010) required participants to press a button during stimulus presentation to increase processing intensity. If there were manipulations of this kind, we coded them with 1, otherwise with 0. We then combined stimulus type and manipulation into an attention score, ranging from 0 to 2 (0+0: words with no special manipulation; 0 + 1: words with special manipulation or 1 + 0: pictures without special manipulation; 2 + 0: videos without special manipulation; or 1 + 1: pictures with special manipulation). Note that there was no experiment that used videos in combination with an additional manipulation, so a score of 3 was never reached.

3.3 Statistics

Most researchers employed analyses of variance (ANOVAs) with contrasts or linear trends as their analysis strategy. To test the critical condition we need the correlation coefficients, which can be calculated from this data. Specifically, in contrast analysis, $r_{effectsize}$ corresponds to the correlation between the contrast weights and the dependent variable. The contrast weights are the values of the independent variables, and thus $r_{effectsize}$ is the correlation we are looking for.

Betsch et al. (2010) reported enough information to calculate the error variance and conduct an ANOVA as well as a contrast analysis without possessing the raw data. The

473

effect size can then be calculated with the following equation (Furr, 2004, p. 10):

$$r_{\text{effectsize}} = \sqrt{\frac{F_{\text{contrast}}}{F_{\text{between}} + \text{df}_{\text{within}}}}$$

where F_{between} and $\text{df}_{\text{within}}$ are from an ANOVA with a single factor (all between conditions).

Hintzman (1970) and Hintzman et al. (1975) correlated the mean values of judgments of frequency for each condition with the corresponding values of duration and frequency. Although the authors did not mention it, this correlation corresponds to r_{alerting} and can be transformed to $r_{\text{effectsize}}$ with the following equation (Rosenthal et al., 1999, p. 46):

$$r_{\text{effectsize}} = \frac{r_{\text{contrast}}}{\sqrt{1 - r^2_{\text{contrast}} + \frac{r^2_{\text{contrast}}}{r^2_{\text{alerting}}}}}$$

where r^2_{contrast} can be calculated from the F value of the reported linear-trend ANOVAs.

Winkler (2009) calculated statistics for dependent measurements (g), where subject-specific variance is controlled for. These data are not directly comparable to the other studies, but we were able to calculate $r_{\text{effectsize}}$ from the provided raw data.

3.4 Analysis Strategy

The critical test consists of two inequalities and can be performed on each individual data set. Evidence against the unidimensional hypothesis can be acquired by simply counting the number of violations of these inequalities. An intuitive way to do this is by (a) transforming the inequalities such that the right-hand side is 0:

$$\beta_{Ff} > \beta_{Df} \qquad \beta_{Fd} > \beta_{Dd} \quad \text{becomes}$$
$$\beta_{Ff} - \beta_{Df} > 0 \qquad \beta_{Fd} - \beta_{Dd} > 0$$

and (b) plotting the differences on the left-hand side of both inequalities against each other. The resulting plot will have four sectors and one can directly see how many violations of the critical test exist. If the unidimensional hypothesis is correct, we would expect no violations (or only a few violations attributable to chance).

4 Results

As can be seen in Figure 3, in every experiment (or experimental condition) the influence of frequency was larger on judgment of frequency than on judgment of duration ($\beta_{Ff} - \beta_{Df} > 0$; only Quadrant 1 at the top right and Quadrant 4 at the bottom right hold data). For the unidimensional hypothesis to be valid, this pattern should also hold for duration ($\beta_{Fd} - \beta_{Dd} > 0$; values should be found only in Quadrant 1). In only 3 (of 38) cases was

474

this fulfilled, a relative frequency of 7.9% (95% confidence interval [1.7%, 21.4%]). Thus, we found 35 cases that speak against a unidimensional model.
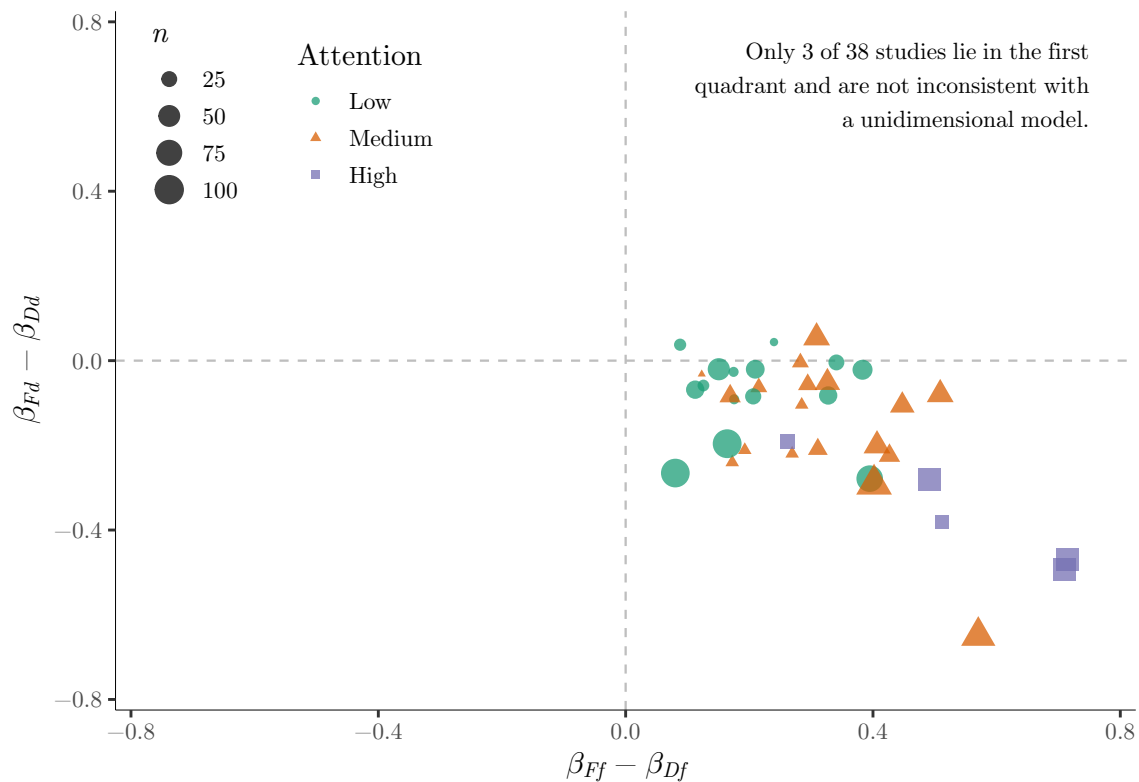

FIGURE 3: Visual analysis of the test condition. R code to reproduce figure: https://github.com/johannes-titz/commonpath.

Since there are only three deviating data points, moderator analyses are not required. Any differences between the studies, such as attention of participants, instructions, or type of duration judgment, would not alter the conclusion that the data are inconsistent with the unidimensional hypothesis. Although the evidence is clear, it might be worth looking at the three deviating cases in more detail (Table 2). Note that in only one condition of Betsch et al.'s (2010) experiments is the unidimensional hypothesis not rejected; in all other conditions (10), it is. The one condition produced effects of -.02/.02 for duration on judgment of duration/frequency, which are essentially null effects. Thus, our analysis contradicts Betsch et al.'s (2010) statement that there is mixed evidence regarding the unidimensional and two-dimensional hypotheses. Independent of the attention condition, the unidimensional hypothesis has to be rejected and thus the conditional unidimensional hypothesis is also unsupported.

Further, note that in the three deviating studies, two effects of duration on judgment of duration are negative, the only ones of the 152. If these effects were larger they might hint at problems in the experimental setup: Stimuli presented for a longer duration should not be judged as shorter. But since these effects are close to 0, they are likely just the result of noise.

475

TABLE 2: Effects consistent with the unidimensional hypothesis.

Study	β_{Dd}	β_{Fd}	β_{Df}	β_{Ff}	Δf	Δd	n
Betsch et al. (2010) e2 c2	−0.02	0.02	0.33	0.42	0.09	0.04	19
Winkler (2008) e6 c1	0.17	0.22	0.14	0.44	0.31	0.05	48
Winkler (2009) e3 c1	−0.03	0.02	0.30	0.54	0.24	0.04	13

Note. e: experiment, c: condition, f: frequency, d: duration, F: judgment of frequency, D: judgment of duration

5   Discussion

In this paper we have tried to answer whether judgments of frequency and duration are based on one or two dimensions. While this question is old, interest in it has reawakened in the last decade (Betsch et al., 2010; Smith et al., 2017; Winkler et al., 2015; Zhao & Turk-Browne, 2011), probably because it has never been answered conclusively. We first retraced the theoretical roots of this question, starting with Hintzman's (1970) seminal paper. We showed how he transferred the total time hypothesis from memory research to judgments of frequency and duration. Hintzman's (1970) verbal statements relating both judgments to memory were expressed formally to better understand the implications of the unidimensional and two-dimensional models. Although Hintzman's (1970) crucial experiment was not consistent with either of these models, he confidently rejected the unidimensional one. His main argument was that the empirical pattern for both judgments was "too different" for a unidimensional model to be valid. This somewhat vague statement seems to correspond to our more stringent criterion regarding the order of effect sizes. The stringency of this criterion lies in its relation to state trace analysis, a general method for testing the number of latent variables involved in a phenomenon.

While Hintzman's conclusion was implicitly based on strong inferential logic, he ignored reliability issues. It was not clear how robust the effects were or whether a different pattern of results could occur in subsequent experiments. It took some time until Betsch et al. (2010) pointed out this and other problems with Hintzman's (1970) conclusion. They found that the effect of duration on both judgments varied substantially from study to study. As a consequence, Hintzman's (1970) definitive statement could be correct for some studies, but incorrect for others.

To explain why the effect of exposure duration varied so much, Betsch et al. (2010) introduced attention as a moderating variable. They hypothesized that if attention is low, the variation of duration goes unnoticed and one might erroneously think that duration has no effect at all. But if attention is high, duration has an effect on the memory strength and eventually also on both judgments. Thus, only under conditions of high attention is a unidimensional model valid. To test this conditional unidimensional model, Betsch et al. (2010) introduced the idea that symmetrical biases are evidence of unidimensionality. Under

conditions of high attention, they indeed found relatively symmetrical biases and concluded that their data is mostly in line with the conditional unidimensional model. But we showed that this conclusion is invalid in the context of state trace analysis because symmetrical biases is not a sufficient criterion for unidimensionality.

Overall, it has remained unclear whether the data on judgments of frequency and duration speak for a unidimensional or a two-dimensional model. We gathered all available experiments in the literature and tested the critical inequalities. Almost all studies are inconsistent with the unidimensional model. Consequently there are also no moderating variables regarding the number of dimensions, at least not in the data we analyzed. Specifically, under conditions of high attention, there is also no evidence of a unidimensional model.

Hintzman's (1970, p. 442) initial intuition, while somewhat vague, was correct: "Since the JOD data reveal a pattern that differs from that of JOF, the two cannot be based on identical information." In contrast, Betsch et al.'s (2010) conditional unidimensional model turned out to be irrelevant because there is no support for any variant of the unidimensional model. Winkler et al.'s (2015, p. 17) conclusion that their "findings strongly argue for the existence of a common mechanism underlying the processing of frequency and duration" has to be doubted. Furthermore, discussions of the unidimensional model to explain isolated results of a few experiments require an update (Smith et al., 2017; Zhao & Turk-Browne, 2011).

5.1   Two Dimensions

If there are two dimensions that determine judgments of frequency and duration, then the question of what these dimensions are arises. There are two views on this that are both related to memory. The first one follows the idea that judgments of frequency and duration are based on different memory constructs from existing memory theories. The second one states that exposure frequency and duration are different, independent attributes of a remembered event.

The first view was already proposed in the initial publication on judgments of frequency and duration (Hintzman, 1970, p. 441): "The notion that long-term memory involves two mechanisms, one related to F and the other to total time, bears enough similarity to the idea that recognition and recall involve two different processes." Indeed, Hintzman has approached this type of question repeatedly in his research (see Hintzman, 2011, for a summary). For instance, he investigated whether judgments of frequency or judgments of recency might be completely based on the familiarity[21] dimension of memory (Hintzman, 2001), which they were not.

Analogously, one could ask whether judgments of frequency or duration are based on familiarity, while the other is based on recollection. Since recollection and familiarity and judgments of frequency and duration are all affected by exposure frequency and duration,

---

[21]For an overview of the distinction between familiarity and recollection, see Yonelinas (2002).

it appears natural to be interested in their interrelatedness. But Hintzman (2011, p. 259) concluded that judgments of frequency, duration, and recency as well as recognition memory are not based on the same information because of substantial *task dissociations*. This conclusion is problematic because task dissociations can occur with a single dimension. Not dissociation, but state trace analysis is the proper method for studying how many constructs are required to explain judgments of frequency, duration, and recency, as well as recollection and familiarity. One way to find out is to conduct new experiments that are tailored to the method of state trace analysis. Such a research program[22] bears potential for some surprise because its superior methodology could invalidate most of the past studies. Specifically, one might find that judgments of frequency are based on one memory construct and judgments of duration on a different one.

Let us now look at the second view, that judgments of frequency and duration are different attributes of memory that are independent and unrelated to other memory constructs. When remembering an event, different features can be accessed, one of them being the exposure frequency and another the exposure duration. At first glance this idea is appealing since it does not overcomplicate the problem and simply assumes that different things are not the same, but different: "People have no trouble telling repetition, duration and recency apart" (Hintzman, 2011, p. 259). But this statement is not quite correct. Judgments of duration are highly dependent on exposure frequency (e.g., Betsch et al., 2010; Hintzman, 1970; Winkler et al., 2015). The correlation between exposure frequency (the irrelevant variable) and judgments of duration is usually higher than between exposure duration (the relevant variable) and judgments of duration. This is a surprising finding if one assumes that humans have direct access to exposure duration as a distinct dimension. To be consistent with empirical findings, one needs to assume that separate pieces of information about exposure frequency and duration are combined to come up with judgments of frequency and duration. How exactly this happens is a theoretical problem worth approaching, but that it must somehow happen takes away the simplicity initially gained by introducing two independent attributes.

So far, we have mainly focused on theories of memory. When taking a judgment perspective, the theory of magnitude (ATOM; Bueti & Walsh, 2009; Walsh, 2003) is likely the most prominent theory related to judgments of frequency and duration (as already suggested in Betsch et al., 2010; Winkler et al., 2015). According to ATOM's basic assumption, all quantities—speed, size, duration, location, or number—are processed in the same brain region.[23] Judgments related to time, space, and quantity hypothetically all overlap to some degree. This explains why judgments of duration are affected by exposure frequency and judgments of frequency by exposure duration. Although ATOM offers an

---

[22]It cannot be done in a few studies.

[23]The theory does not directly refer to exposure frequency, although frequencies are related to numbers and are prothetic (experiences of "more than" or "less than"), the requirement Walsh postulates for his theory. Because of its focus on motor responses it also does not necessarily extend to episodic memory (Bueti & Walsh, 2009, p. 1836), but some kind of transfer appears plausible.

interesting view, it suffers from not specifying concrete processes. It remains unclear what exactly happens in the specific brain regions. One might naively suggest that some neurons are responsible for processing frequency, some for processing duration, and some for both. In this case, on the macro level, there are three dimensions, but the behavioral predictions do not change much from the memory-oriented views. Biases should be present and the judgments are expected to be correlated to some degree, but state trace analysis will result in at least two dimensions. Thus, the main task in the future is to accurately specify the different views on the two-dimensionality in judgments of frequency and duration to make more elaborate predictions.

5.2   The Assumption of Linear Relationships

Our assumption of linear relationships is mostly a consequence of previous work. Contributors in the field usually did not question it (Betsch et al., 2010; Hintzman, 1970; Hintzman et al., 1975; Williams & Durso, 1986; Winkler et al., 2015; Zhao & Turk-Browne, 2011). All of their analyses, such as ANOVAs with linear trends, contrast analysis with linear contrasts, correlations, regression with untransformed variables, assume linear relationships between independent and dependent variables. Thus, the whole reasoning of these authors is implicitly based on the linearity assumption. To understand their arguments, we had to take their position. Without doing this, we could not have understood why Hintzman (1970) was convinced that his correlation pattern is crucial to rule out a unidimensional model. We might also have missed that symmetry in biases is not a sufficient condition for a unidimensional model and that Betsch et al.'s (2010) data required a re-analysis.

In addition to this pragmatic reason, there is some evidence for the linearity assumption: Hintzman (1970) compared linear with logarithmic models and found very small and inconsistent differences in R^2. The figures in Bonanno (1996) as well as in Winkler (2009) show linear relationships for almost all conditions. The data in Betsch et al. (2010) look fairly linear too. Overall, it appears that, for the particular values of the independent variables chosen in these studies, the relationships in question can be assumed to be linear. If other, more extreme and more finely spaced values are chosen, this assumption might eventually be violated.

If this is the case, a normal state trace analysis could be performed as an alternative. Usually, only the assumption of monotonic relationships is made in state trace analysis, which appears reasonable in judgments of frequency and duration. Most theoreticians would assume that the memory strength is a monotonically increasing function of frequency and duration (i.e., when increasing exposure frequency or duration, the memory strength will not decrease). This assumption is also reasonable for the relationship between the memory strength and the judgments. Note that to conduct state trace analysis, correlations between independent and dependent variables (as used here) are not sufficient, so new experiments need to be run.

5.3   Problems of Comparability

There might be concerns that the data we used to test the critical condition are too diverse. Specifically, the independent and dependent variables regarding duration were not the same in all experiments. Hintzman (1970) was mainly interested in average duration, while Betsch et al. (2010) were interested in total duration.[24] Winkler (2009) and Winkler et al. (2015) studied both. While this concern is definitely problematic when aggregating the data (e.g., in a meta-analysis), it is not a problem in our analysis. Since almost all experiments violate the unidimensional hypothesis, it does not matter whether single or total duration was manipulated and assessed. For both variables, there is no evidence of a unidimensional model. This logic applies to all potential moderators such as range and mean of independent variables, instructions, or attention manipulations.

5.4   Conclusion

On the basis of a methodologically stringent criterion and data from many experiments, we can conclude that judgments of frequency and duration are probably based on two independent dimensions. There seems to be only one convincing way to still argue for a unidimensional model: a violation of the linearity assumption. In this case, novel experiments should be conducted and analyzed with the method of state trace analysis. This is also a good approach to differentiate between different two-dimensional models: State trace analysis can reveal whether judgments of frequency and duration are based on other existing memory constructs. Finally, more theoretical work is required to understand why humans confuse frequency and duration, although the judgments of these attributes are based on two independent dimensions.

References

Betsch, T., Glauer, M., Renkewitz, F., Winkler, I., & Sedlmeier, P. (2010). Encoding, storage and judgment of experienced frequency and duration. *Judgment and Decision Making*, *5*(5), 347–364. http://journal.sjdm.org/10/91221b/jdm91221b.pdf

Bonanno, J. (1996). *The list-strength effect and categorical frequency memory: Tests of availability* (Master's thesis, Wilfrid Laurier University). https://scholars.wlu.ca/etd/568

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Cooper, E. H., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, *68*, 221–234. https://doi.org/10.1037/h0025052

---

[24]Except for one experiment where they assessed both.

Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-73129-2

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101. https://doi.org/10.1037/0033-295X.95.1.91

Furr, R. M. (2004). Interpreting effect sizes in contrast analysis. *Understanding Statistics*, *3*, 1–25. https://doi.org/10.1207/s15328031us0301_1

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, *80*, 139–145. https://doi.org/10.1037/h0027133

Hintzman, D. L. (1970). Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology*, *83*, 435–444. https://doi.org/10.1037/h0028865

Hintzman, D. L. (2001). Judgments of frequency and recency: How they relate to reports of subjective awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1347–1358. https://doi.org/10.1037//0278-7393.27.6.1347

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, *32*, 336–350. https://doi.org/10.3758/BF03196863

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science*, *6*, 253–271. https://doi.org/10.1177/1745691611406924

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). What causes the spacing effect? Some effects of repetition, duration, and spacing on memory for pictures. *Memory & Cognition*, *3*, 287–294. https://doi.org/10.3758/BF03212913

Huppert, F. A., & Piercy, M. (1977). Recognition memory in amnesic patients: A defect of acquisition? *Neuropsychologia*, *15*, 643–652. https://doi.org/10.1016/0028-3932(77)90069-0

Kenny, D. A. (1979). *Correlation and causality*. Wiley.

Mo, S. S., & Michalski, V. A. (1972). Judgment of temporal duration of area as a function of stimulus configuration. *Psychonomic Science*, *27*, 97–98. https://doi.org/10.3758/BF03328900

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*, 10–15. https://doi.org/10.1037/h0027470

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (1999). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804403

Smith, A. R., Rule, S., & Price, P. C. (2017). Sample size bias in retrospective estimates of average duration. *Acta Psychologica*, *176*, 39–46. https://doi.org/10.1016/j.actpsy.2017.03.008

Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *8*(1), 83–94. https://doi.org/10.1016/S0022-5371(69)80015-0

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Warm, J. S., & McCray, R. E. (1969). Influence of word frequency and length on the apparent duration of tachistoscopic presentations. *Journal of Experimental Psychology*, *79*, 56–58. https://doi.org/10.1037/h0026871

Williams, K. W., & Durso, F. T. (1986). Judging category frequency: Automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 387–396. https://doi.org/10.1037/0278-7393.12.3.387

Winkler, I. (2008). *Experiments on JOF and JOD* (Unpublished raw data).

Winkler, I. (2009). *The processing of frequency and duration* (Doctoral dissertation, Chemnitz University of Technology). https://nbn-resolving.org/urn:nbn:de:bsz:ch1-200900917

Winkler, I., Glauer, M., Betsch, T., & Sedlmeier, P. (2015). The impact of attention on judgments of frequency and duration. *PLoS ONE*, *10*, e0126974. https://doi.org/10.1371/journal.pone.0126974

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*, 161–215. https://doi.org/10.1214/aoms/1177732676

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. https://doi.org/10.1006/jmla.2002.2864

Zhao, J., & Turk-Browne, N. B. (2011). Incidental encoding of numerosity in visual long-term memory. *Visual Cognition*, *19*, 928–955. https://doi.org/10.1080/13506285.2011.598482

Appendix

Covariance Algebra for One Dimension

If there is a single dimension on which both judgment of frequency and judgment of duration are based, judgment of frequency and judgment of duration should be correlated. If there are two orthogonal dimensions they should be uncorrelated.

To demonstrate this formally, we need to study the standardized regression coefficients, which will lead us directly to the correlation between the two judgments. In general, standardized regression coefficients play an important role because most arguments in the literature were directly or indirectly based on correlations or equivalent effect sizes (e.g., R^2, η^2) and not on unstandardized effects (e.g., unstandardized regression coefficients).

The correlation between judgment of frequency and judgment of duration reduces to the covariance if the variables are standardized:

$$r(F, D) = \text{COV}(F_z, D_z)$$

Let us assume that both judgments are affected by frequency and duration and thus are linear combinations of these variables:[25]

---

[25]Note that both linear combinations have error terms, but they are uncorrelated with f_z and d_z and thus can be neglected.

$$r(F, D) = \text{COV}(\beta_{Ff} f_z + \beta_{Fd} d_z, \beta_{Df} f_z + \beta_{Dd} d_z)$$

where f_z and d_z are now the standardized variables of frequency and duration and the β's are the standardized regression coefficients. Note that in the case of uncorrelated independent variables, the standardized coefficients are equal to simple correlations and thus we conventionally use β for them.

By applying covariance rules (e.g., Kenny, 1979), in a few steps we get to

$$r(F, D) = \beta_{Ff} \beta_{Df} \cdot \text{VAR}(f_z) + \beta_{Df} \beta_{Fd} \cdot \text{COV}(f_z, d_z) +$$
$$\beta_{Fd} \beta_{Dd} \cdot \text{VAR}(d_z) + \beta_{Ff} \beta_{Dd} \cdot \text{COV}(d_z, f_z)$$

In a controlled experiment the covariance between f_z and d_z is 0 and because the variables are standardized, the variance of both variables is 1, which reduces the formula to

$$r(F, D) = \beta_{Ff} \beta_{Df} + \beta_{Fd} \beta_{Dd}$$

The presented logic can also be followed purely functionally (cf. Dunn & Kirsner, 1988). If judgment of duration and judgment of frequency are based on the same memory strength, then

$$D = g_D(m)$$

and

$$F = g_F(m)$$

Applying the inverse function to express m in terms of F results in

$$g_F^{-1} F = g_F^{-1} g_F(m) = m$$

and now substituting m results in

$$D = g_D(g_F^{-1}(F))$$

Or in words, judgment of duration will be a function of judgment of frequency. Under the assumption of a linear relationship, a positive correlation will result between the two variables. The analysis of Dunn and Kirsner (1988) shows that this functional dependence does not logically exclude a two-process model. For instance, if duration and frequency are functionally dependent or two memory dimensions are functionally dependent, we cannot exclude a two-process model that just looks like a one-process model. But, duration and frequency are uncorrelated in an experiment and the simplest model with the fewest degrees of freedom will be a unidimensional model.