

How much do we orient? A systematic approach to auditory distraction

Article

Accepted Version

Beaman, C. P., Campbell, T. and Marsh, J. E. (2021) How much do we orient? A systematic approach to auditory distraction. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 47 (7). pp. 1054-1066. ISSN 0278-7393 doi: <https://doi.org/10.1037/xlm0000995> Available at <https://centaur.reading.ac.uk/93792/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1037/xlm0000995>

Publisher: American Psychological Association.

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



How much do we orient? A systematic approach to auditory distraction

C. Philip Beaman¹, Tom Campbell^{2,3}, & John E. Marsh^{4,5}

1. School of Psychology & Clinical Sciences, University of Reading, UK
2. Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland
3. Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
4. School of Psychology, University of Central Lancashire, Preston, Lancashire, UK
5. Engineering Psychology, Humans and Technology, Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

Correspondence should be addressed to: C. Philip Beaman, School of Psychology & Clinical Language Sciences, University of Reading, Earley Gate, Whiteknights, Reading RG6 6AL, United Kingdom. Tel: +44 (0) 118 3787637 Email: c.p.beaman@reading.ac.uk

Running Head: HOW MUCH DO WE ORIENT?

Word Count: 8,996

Abstract

Data on orienting and habituation to irrelevant sound can distinguish between task-specific and general accounts of auditory distraction: Distractors either disrupt specific cognitive processes (e.g., Jones, 1993; Salamé & Baddeley, 1982), or remove more general-purpose attentional resources from any attention-demanding task (e.g., Cowan, 1995). Tested here is the prediction that there is no further auditory distraction effect on immediate serial recall with increments in the number of distractors beyond the “changing-state point” of two discrete distractors. A Bayes factor analysis refutes this nil hypothesis: This prediction, a key element of the strong changing-state hypothesis, is shown to be less likely than two competing alternatives. Quantitative predictions for distraction as a function of the number of distractors are derived for an orienting response (OR) and a stimulus mismatch (SMM) hypothesis, representing general and task-specific accounts respectively. The data are shown to be more likely under the SMM hypothesis. Prospects for a parametric account of auditory distraction are considered.

Keywords: auditory distraction, irrelevant sound effect, working memory, orienting, habituation

Cognitive performance upon a range of different tasks has been shown to be susceptible to disruption by auditory distraction. Tasks vulnerable to the disruptive effects of irrelevant sound include such activities as counting (Logie & Baddeley, 1987), arithmetic (Banbury & Berry, 1997, 1998), proof-reading (Jones, Miles, & Page, 1990), and data entry (Beaman, 2005). The effects of distraction are not equivalent across all tasks, being more robust and pronounced when immediate verbal memory is tested (e.g., Bell, Röer, Dentale, & Buchner, 2012; Ellermeier & Hellbrück, 1998; Elliott, 2002; Jones, 1993; Jones, Madden & Miles, 1992; Lange, 2005; Salamé, & Baddeley, 1982) and in particular when the order of the verbal sequence needs to be retained (Beaman & Jones, 1997, 1998; Farley, Neath, Allbritton & Surprenant, 2007; Gisselgård, Uddén, Ingvar & Petersson, 2007; Henson, Hartley, Burgess, Hitch & Flude, 2003; Jones & Macken, 1993; Neath, Guérard, Jalbert & Surprenant, 2009), all of which are features of typical serial short-term memory tasks (Oberauer *et al.*, 2018). In consequence, the auditory distraction known as the “irrelevant sound effect” (Beaman & Jones, 1997) has both featured heavily in the development of short-term memory models and been explained in terms of interference processes in short-term memory (Jones, Beaman & Macken, 1996; Neath, 2000; Norris & Page, 2003; Salamé & Baddeley, 1982).

In a typical short-term memory task involving irrelevant sound, auditory distractors are played alongside the presentation of visual memoranda. Participants are required to ignore anything they may hear having been given the reassurance there will be no test concerning the content of the sound. On presentation of a recall cue, participants attempt to recall the visually-presented items in the original order. The apparent task-specificity of auditory distraction has given rise to a number of hypotheses to account for auditory distraction effects in terms of interference with specific representations, such as the suggestion that speech interferes with the retrieval of phonological representations (Salamé,

& Baddeley, 1982) or other verbal codes (Neath, 2000), or processes such as the maintenance of sequence order (Jones, 1993). Indeed, the original finding that speech is a particularly disruptive distractor resulted in the reformulation of a phonemic output buffer (Baddeley & Hitch, 1974) as a phonological input store within the influential multistore model of working memory (Baddeley, 1986; Salamé, & Baddeley, 1982; for a discussion, see Shallice & Cooper, 2011). To be considered here are the possibilities that a specific working memory account of auditory distraction is unnecessary if it can be assumed by a more general attentional account, and whether a parametric account of auditory distraction as a derivable *function* of the distractors is possible.

Theories of Auditory Distraction by Irrelevant Sound

The phonological store hypothesis, as it originally existed, assumed that unattended and task-irrelevant speech (but not irrelevant sound) disrupted verbal memory. This disruption took place via a confusion between the phonemes in the to-be-ignored auditory distractors and the to-be-recalled visual-verbal memoranda. Later results showed that the phenomenon was not restricted to irrelevant speech as various forms of non-speech also showed the same effect (e.g., Jones & Macken, 1993; Klatte, Kilcher, & Hellbrück, 1995; Schlittmeier, Hellbrück & Klatte, 2008). On top of this, further studies failed to replicate the original Salamé & Baddeley (1982) finding that irrelevant speech disruption was a function of the phonological similarity between irrelevant auditory items and visual memoranda (Jones & Macken, 1995; Larsen, Baddeley, & Andrade, 2000; LeCompte, & Shaibe, 1997), leading to the abandonment of this particular hypothesis. Instead, the changing-state effect (Jones, Madden & Miles, 1992) has come to be seen as the main empirical signature of the “irrelevant sound effect” (Beaman & Jones, 1997) within serial short-term memory. This irrelevant sound effect in serial short-term memory is particularly pronounced when the

distracting sound consists of changing sequences of multiple different task-irrelevant sounds. Those sounds can be speech or non-speech. Sound that changes acoustically from one token to the next (*changing-state* sound, e.g., A-B-C-D) is more disruptive to serial recall than repetitive, *steady-state* sound in which the same token is repeated (e.g., A-A-A; Jones & Macken, 1993; Jones *et al.*, 1992; Meiser & Klauer, 1999). This changing-state effect is one of the benchmark phenomena of working memory (Benchmark 6.2B, Oberauer *et al.*, 2018).

Taking its name from the effect, the changing-state *hypothesis* is the most prominent task-specific hypothesis of auditory distraction. This account assumes that the disruption caused by changing-state irrelevant sound arises from the conflict of processes of seriation within the auditory distractors and the memoranda. Consistent with other data on the perception of order within auditory streams (e.g., Warren & Obusek, 1972), the changing-state hypothesis assumes changes-in-state between auditory events that vary abruptly (e.g., the sequence A-B-C, or a pitch glide interrupted by periods of silence) encode the temporal order of the sequence or stream. Contrastingly, unchanging or repeated (steady-state) auditory events (e.g., the sequence A-A-A, or a single uninterrupted pitch glide) contain little order information (Jones, Alford, Bridges, Tremblay & Macken, 1999; Jones & Macken, 1995). This account is consistent with a number of results on effects of varying the number of auditory streams, that is the number of different sound sources (often voices; Jones *et al.*, 1999; Jones & Macken, 1995b, c; Jones, Macken & Harries, 1997; Klatte *et al.*, 1995) as well as data indicating that tasks which require participants to maintain the serial order of a sequence of memoranda are peculiarly susceptible to the irrelevant sound effect (Beaman & Jones, 1997, 1998; Farley *et al.*, 2007; Gisselgård *et al.*, 2007; Henson *et al.*, 2003; Jones & Macken, 1993; Neath *et al.*, 2009).

More general accounts have also been applied to the phenomenon. Cowan (1995) suggested that the changing-state effect might be explicable in terms of known principles of

orienting and habituation (Sokolov, 1963). According to this account, individuals involuntarily orient towards novel auditory stimuli thereby withdrawing attention from the primary task, resulting in poorer performance. As such, this orienting response becomes habituated over time, restoring performance to something approaching baseline levels. An attraction of this approach is the way in which the changing-state effect arises naturally from already familiar principles.

Numerous studies of habituation to irrelevant sound have been reported as tests of the orienting account (Banbury & Berry, 1997; Bell, Röer, Dentale, & Buchner, 2012; Ellermeier & Zimmer, 1997; Elliott & Cowan, 2001; Hellbrück, Kuwano & Namba, 1996; Jones *et al.*, 1997; Morris & Jones, 1990; Perham & Banbury, 2008; Röer, Bell, Dentale & Buchner, 2011; Röer, Bell & Buchner, 2014; Tremblay & Jones, 1998), but rather fewer direct investigations of orienting yet tests of habituation to irrelevant sound implicitly assume that attentional capture by irrelevant sound is the basis of auditory distraction in the first place. However, this remains an assumption and the term “habituation” has been used rather loosely (e.g., by Morris & Jones, 1990) to refer to both an effect and its cause, where in principle a decrease in distraction over time (an apparent habituation effect) could be accounted for by mechanisms other than reduction in an orienting response (an habituation mechanism). For example, there is substantial evidence that dual-task costs can be reduced with practice (e.g., Oberauer & Kliegel, 2004; Spelke, Hirts & Neisser, 1976; Underwood, 1974) and participants might simply become skilled in avoiding, or resolving, confusions (e.g., of serial order cues) at recall.

Theoretically, Cowan’s (1995) suggestion of orienting plus habituation seems to provide an existing framework for auditory distraction, in particular the changing-state effect but tests of habituation as a mechanism have tended to look for habituation as an effect over the longer periods of time represented by blocks of trials. A number of studies (Banbury &

Berry, 1997; Bell, Röer, Lang & Buchner, 2019; Bell, Röer, Dentale, & Buchner, 2012; Morris & Jones, 1990) show reduced auditory distraction when participants undertake a serial recall task in the presence of distractors after a passive listening period. In that period, an assumption is that participants listen to and form a neural model of the distractors lending some credence to the prospect of an habituation-based account. It remains uncertain whether an habituation effect over such longer periods necessarily implies an habituation process efficient enough to reduce the changing-state effect. At least one study (Mutschler *et al.*, 2010) has suggested that auditory habituation occurs in parallel by two different means and along two different timescales. A plausible, if less parsimonious, account might therefore allow for very rapid short-term habituation of orienting without attention over the period of a small number of trials – sufficient to account for the changing-state effect – and a smaller, longer-term effect over blocks of trials, requiring pre-exposure and, presumably, attending to the irrelevant sound. If correct, these two processes should be examined independently and not conflated.

Tests of Orienting and Other Hypotheses

Since habituation to a distractor presupposes orienting, one way of avoiding the problems discussed above is to look directly at the evidence for orienting *per se* rather than the evidence for habituation following orienting. Even this direct assessment is not without conceptual problems, however. Näätänen (1986, p. 91) noted that, as a central concept in psychophysiology, orienting was “even too popular” and made the provocative claim that because there were multiple types of orienting, each associated with particular types of surprise or novelty (Näätänen, 1979, p. 61), the scope of the term had grown “too broad” to serve as a useful conceptual tool. In particular, there is no *a priori* basis for deciding what size the effect of orienting should be on the short-term memory tasks most typically used to

investigate auditory distraction unless further analytical work is done to provide such a basis. In the following, a brief conceptual analysis introduces how orienting might operate during auditory distraction before applying Bayesian tools to existing experimental data. This conceptual backdrop then serves to derive specific *quantitative* hypotheses to test and thus differentiate between the theoretical possibilities.

Theoretically Motivated Quantitative Hypotheses

The extent to which orienting might be anticipated over a typical auditory changing-state sequence has remained an open question, which existing evidence, however, has the potential to answer. The presentation rate of such stimuli is typically either at the rate of normal conversation (approximately 2-3 utterances per second if connected speech is played) or somewhere between 1-2 auditory events per second (if speech or non-speech sound is edited and presented according to a regular presentation schedule). It also seems reasonable to presume that the orienting (if orienting occurs) to conversational speech is at, or close to, the maximum that might be expected. That is, speech signals typically prove to be at least as disruptive as any other sound when other factors (e.g., presentation rate, changing-state information) are controlled. These considerations are based only on orienting towards the acoustics of the auditory signal rather than the meaning of the signal. Data suggest that auditory distraction may also be mediated by semantics (Beaman, 2004; Beaman, Hanczakowski, Hodgetts, Marsh & Jones, 2013; Hanczakowski, Beaman & Jones, 2017; Jones, Miles, & Page, 1990; Neely & LeCompte, 1999; Marsh, Hughes & Jones, 2008, 2009; Marsh, Perham, Sorqvist & Jones, 2014; Oswald, Tremblay & Jones, 2000; Röer, Bell, Körner & Buchner, 2018), depending on the processing of meaning within the focal task (Marsh, Hughes & Jones, 2009). They also suggest auditory distraction occurs when there is deviation from an anticipated higher-order structure (Hughes, Vachon & Jones, 2005, 2007;

Vachon, Hughes & Jones, 2012). Both semantics and such deviation characterize everyday, conversational speech but these occurrences speak to the disruptive potential possessed by the informational content of an auditory sequence rather than the purely acoustic or energetic effects of a sound, which are both cognitively and neurally distinct (Beaman, Bridges, & Scott, 2007). These semantic and deviation effects are not the focus of the present investigation.

A repeated, steady-state sound or utterance (e.g., A-A-A-A) can also cause disruption given a sufficiently sensitive design (e.g., Bell, Röer, Lang & Buchner, 2019; LeCompte, 1995; Parmentier & Beaman, 2015) so this, presumably, represents the effects of orienting to a single sound. Similarly, a changing-state sequence that consists of two distinct sounds or utterances (e.g., A-B-A-B) should represent the effects of orienting to two auditory events. A sequence of three sounds (A-B-C-A-B-C) should evoke orienting responses to three events and so on. The changing-state effect is thus an instance of what, in reference to the number of different auditory events, Tremblay and Jones (1998) dubbed the *token set size*. Tremblay and Jones (1998) hypothesized that orienting to each token should be additive. Accordingly, the disruption caused by auditory distraction should be, approximately, a linear function of the token set size. They contrasted this situation with the predictions of a strong version of the changing-state hypothesis: According to the strong version of the hypothesis, the repeated presence of a single change (from “A” to “B” and back again) should be sufficient to set-up order cues and disrupt immediate recall to the extent that further changes would have no greater impact on recall. The results of a direct test by Tremblay and Jones (1998) supported the strong version of the changing-state hypothesis (no significant further disruption was observed after a token set-size of two). That is, despite the predicted significant increase in disruption seen with increments in set size from 1-token (AAA...) to 2-tokens (ABAB...), increments in set size beyond this *changing-state point* did not produce a significant increase

in disruption. However, a consistent although non-significant numerical tendency was seen for higher set sizes to produce an increased disruption.

In a more recent, and statistically more powerful, test of the strong changing-state hypothesis, Bell and colleagues (Bell, Röer, Lang & Buchner, 2019a) re-assessed the strong changing-state hypothesis. Bell *et al.* concluded that, across two experiments, the results were inconsistent with this hypothesis and instead favoured the orienting hypothesis. A problem with this interpretation, however, is that the effect size ($\eta_p^2=.10$ for a difference between two and eight tokens) upon which these researchers based their sample size was not theoretically motivated. Thus, although their results are sufficiently improbable, given the null hypothesis, for them to reject the strong changing-state hypothesis the findings do not *by themselves* provide any indication that these data are any more likely under the orienting account. The reason is that no particular effect size is suggested as a plausible orienting effect (for a more general version of similar arguments see Dienes, 2019; Dienes & Mclatchie, 2018). To take a simplistic example, if orienting predicted an effect size of $\eta_p^2=.3$ or above (i.e., three or more times that observed) then the orienting response prediction is at least twice as far from the observed effect than is the nil effect, yet the null hypothesis is rejected in favour of this less likely alternative. To address this problem, the present investigation advances some possible functions that relate token set size to disruption observed to serial recall. The analysis assesses which function is the most plausible given the data.

In the absence of any more principled basis for estimating the effects of orienting, one possible version of the orienting hypothesis is to assume a reapplication of each percentage decrement every time token set size is incremented by one token to produce the required monotonic, nonlinear function. As an example, if the difference in performance between set size 1 (steady state) and set size 2 (changing state) is 30% then there should be a further 30% decrement from set size 2 to set size 3 and so on. This pattern of data gives a decreasing

function for disruption, as shown in the upper panel of Figure 1(A), consistent with the existing shape of the token set size effect as observed by Tremblay and Jones (1998). It represents a somewhat conservative form of orienting beyond set size 2 as the increase remains constant at every point, resulting in a decelerating increase in disruption as demonstrated in the upper panel of Figure 1(A). This suggestion also has the attractive property of being parameter-free, as the initial disruption from two items relative to one is taken as the baseline from existing data and simply re-applied in an iterative manner.

FIGURE ONE ABOUT HERE

An even simpler alternative to both the strong changing-state hypothesis and the orienting hypothesis is referred to here as the stimulus-mismatch hypothesis (after Campbell, 2000). This stimulus-mismatch hypothesis flows from Beaman's (2000) suggestion that the pattern of disruption observed by Tremblay and Jones (1998) should be taken seriously even if the direct comparisons between memory performance at adjacent set sizes (e.g., set size 2 vs set size 3) are not statistically significant. Beaman (2000) suggested that this pattern could readily be accounted for by assuming that noise is added to memory for each token to the extent that the token contains novel information not already present in the stream of distractors. This resembles the idea of "novelty-gating" espoused more recently by some researchers as a mechanism for limiting encoding in working memory (e.g., Farrell & Lewandowsky, 2003; Oberauer, Lewandowsky, Farrell, Jarrold & Greaves, 2012).

The simplest way of implementing this novelty-gating is to again assume that the difference between steady-state (set size 1) and changing-state (set size 2) conditions represents baseline performance (as in the implementation of the orienting hypothesis). A further assumption necessary is that, because of an overlap in features and content between

tokens, each subsequent token only adds half as much noise again to memory as its immediate predecessor, resulting in a pattern something similar to that shown in the upper panel of Figure 1 (A). Technically this model has one free parameter in that it assumes a 50% overlap between the features of different distractors. This overlap results in each subsequent token only adding half as much noise again. However, this simplifying assumption is itself ultimately derived from the idea within the feature model that items are represented in terms of feature vectors. Each feature takes a binary value. In this way, any randomly generated pair of features would be expected to share, on average, half the values in their respective feature vectors (Beaman, 2000; Nairne, 1990). The assumed 50% overlap is therefore kept as a constant and is not varied in the studies which follow. This hypothesis produces a pattern of results most obviously more similar in form to that reported by Tremblay and Jones (1998, Figure 2, and reproduced in the lower panel of Figure 1(B)) than the strong changing-state and orienting hypotheses. This stimulus-mismatch hypothesis produces neither the polynomial function which is the best-fitting function to Tremblay and Jones (1998), as redrawn in Figure 1, nor the broken power law suggested by Campbell (2000; see also Campbell, Beaman, & Berry, 2002; Campbell, Winkler, Kujala, & Näätänen, 2003; Campbell, Winkler, & Kujala, 2005)¹.

Existing Data on the Token Set-Size Function

The foregoing has demonstrated that token set size is a variable that can readily be varied parametrically, and for which parametric functions can be derived, but a peculiarity of the studies reviewed so far is that only in the first two studies by Tremblay and Jones (1998) was set size varied in this way and in those studies the token set size effect was non-

¹ The pattern is a *broken* power law because an unbroken power law would predict the largest difference to appear between quiet and set size one (steady-state). This conundrum will be addressed in the General Discussion.

significant. Although it is possible to derive token set size predictions in the form of quantitative functions from theories of auditory distraction, as suggested above, this also has not previously been attempted. One reason for this is that Tremblay and Jones (1998) assumed that there was no further disruption beyond the changing-state point because of their non-significant results, and concluded in favor of the strong changing-state hypothesis. All later studies, such as those by Campbell and colleagues (Campbell *et al.*, 2002, 2003, 2005) as well as Bell *et al.* (2019a) took this as a starting point and, as demonstrated in Table 1, constructed experiments specifically to test the hypothesis of no difference between the changing-state point and some larger set size.

The Tremblay and Jones (1998) data are also not the only results in which null effects were observed: Unequivocal further disruption by increasing set size beyond the changing-state point is not universal. Null effects were apparent in Campbell *et al.* (2002, Experiment 3A) for example, and in a later study by Hughes and Jones (2005, Experiment 2) which found no significant increase in disruption with increments in set size from two to eight. Hughes and Jones (2005) provocatively concluded that, “The absence of a basic set size effect casts doubt on the one previous experiment² in which such an effect was found (Campbell *et al.*, 2002: Exp. 3b) and buttresses the claim that the classical changing-state effect is driven by acoustic changes between immediately adjacent items (Tremblay & Jones, 1998).” However, the higher number of trials per condition used by Campbell *et al.* (2002), led to a more sensitive procedure than that of Hughes & Jones (2005) and of Tremblay & Jones (1998). Caution is also warranted in interpreting Experiment 3A of Campbell *et al.* (2002) as unequivocally demonstrating a null significant disruptive advantage of increments in set size beyond two across all serial positions: Increments in set size from one to two to

² This claim was incorrect at the time because Experiment 3 of Tremblay and Jones (1998, Experiment 3) found a significant set size effect, as did Campbell *et al.* (2003).

five produced a linear trend of reduced performance with high set sizes, without a significant nonlinear residual (Campbell, 2000). Thus, whether the numerical trends which appear in Tremblay and Jones (1998) represent reliable parametric functions or not remains debatable. The present investigation resolves that issue.

Following Bell *et al.* (2019b), Table 1 shows the frequency with which statistically significant effects of token set size have been reported as a function of the sample size of the experiment and the differences between the token set sizes considered in all the papers identified using the procedure set out in the Data Selection section below.

TABLE ONE ABOUT HERE

Bayesian Reanalysis

The data so far have indicated that a stimulus-mismatch or orienting hypothesis might be preferred to the strong version of the changing-state hypothesis as potential explanations for the irrelevant sound effect and the changing-state effect. However, as Table 1 shows, although individual papers report data which seem internally consistent, the literature overall is rather less so. Further, the orienting and stimulus-mismatch hypotheses have not been tested against each other: All studies to date have either nullified the strong changing-state hypothesis or have not succeeded in doing so, and they have not directly tested alternate hypotheses. To do this, and to add to the evidence assessing the strong changing-state hypothesis, which remains conflicted despite the significant results reported by Bell *et al.* (2019a), the present enquiry makes use of the Bayes factor analysis method outlined by Dienes (2014, Appendix 1).

Call a Bayes factor comparing theory 1 with the null hypothesis $B_{1/0}$. Call a Bayes factor comparing theory 2 with the null hypothesis $B_{2/0}$. Then, the Bayes factor comparing theory 1 with theory 2,

$$B_{1/2} = \frac{\text{likelihood of obtained data given theory 1}}{\text{likelihood of obtained data given theory 2}} \quad (1)$$

In this notation, a Bayes factor of $B_{1/2} > 1$ represents evidence in favour of the first theory in the pairwise comparison. For example, a Bayes factor of $B_{1/2} > 3$ indicates theory 1 is more than three times as likely as theory 2 given the data. A Bayes factor of $B_{1/2} < 1/3$ represents equivalent evidence in favour of the second theory. A Bayes factor of 1.0 is therefore exactly equivocal between the theories. The Bayes factor gives a continuous measure of support such that any value over 1.0 favours theory 1 (even if barely) and any value under 1.0 favours theory 2. In the current case, for token set sizes above 2, the strong changing-state hypothesis (SCS) is a null hypothesis, the stimulus-mismatch hypothesis (SMM) is theory 1, and the orienting response hypothesis (OR) is theory 2.

$B_{SMM/SCS}$ represents the relative strength of evidence in support of the stimulus-mismatch hypothesis compared to the strong changing-state hypothesis. As already indicated, since the strong changing-state hypothesis gives a null prediction of no further disruption beyond set size 2, this assertion is equivalent to $B_{1/0}$ for these set sizes. Similarly, $B_{OR/SCS}$ is the Bayes factor representing the support for the orienting hypothesis over the strong changing-state hypothesis and again is equivalent to comparing the orienting hypothesis to a null prediction. By extension then, $B_{SMM/SCS}$ represents the relative support for the stimulus-mismatch hypothesis when compared to the orienting hypothesis.

To summarise, this approach is an advance upon previous analyses in two ways. In the first place, the current data – as tested by traditional null hypothesis significance testing

(NHST) – indicate only $p(\text{Data} \mid \text{Null Hypothesis})$. A statistically significant result therefore indicates that the dataset is unlikely to have been generated given the null hypothesis of exactly zero difference between conditions is true. A statistically non-significant result, in contrast, indicates merely that the data are plausible given this null hypothesis. In neither case is the plausibility of the data given a stated alternate hypothesis explicitly considered. Thus, a null and an alternate hypothesis could be equally plausible or implausible but the traditional NHST considers only the plausibility of the null hypothesis. Bayes factors potentially improve upon this situation by their use of likelihood ratios in which the plausibility of two separate hypotheses are considered, allowing for the most plausible of the two to be selected. At present, a small difference between token set sizes two and eight might yield a result which is either non-significant (Tremblay & Jones, 1998) or significant (Bell *et al.*, 2019a). This situation leads to arguments over the balance of type 1 vs type 2 errors and very different theoretical interpretations. In both cases, however, the data might be more plausible under the hypothesis of a small effect than either the null hypothesis of precisely zero effect (Tremblay & Jones, 1998) or an alternate hypothesis that assumes a rather large effect (Bell *et al.*, 2019a).

The second advantage of using this approach is that it is not restricted to testing only a null prediction. As noted by Cohen (1995; see also Gigerenzer, 2018), the null hypothesis was intended by Fisher to indicate the hypothesis intended to be nullified. However, typically with NHST, the null hypothesis has come to mean testing a hypothesis of null difference – an approach referred to by Cohen (1995) as “nil hypothesis testing”. This investigation will test two hypotheses, both of which postulate a non-zero difference between conditions, against each other.

Yet a third advantage of this approach is the opportunity to combine existing data-sets meta-analytically and examine the weight of evidence across published results (see also

Sörqvist, Marsh, & Nössl, 2013). Datasets will be combined in a meta-analysis using the mean and SE from Study 1 as the prior mean and prior SD and the mean and SE from Study 2 as the likelihood's mean and likelihood's SD to calculate the posterior mean and posterior SD. The posterior mean and SD thus obtained are then entered as the new prior mean and SD and the mean and SE from Study 3 used as the likelihood's mean and SD and so on until the data from all available experiments have been combined (Dienes, 2008; see also Parmentier & Beaman [2015] for a previous application of this procedure to irrelevant sound effects). This procedure then gives the likelihood of all the data combined given the strong changing-state hypothesis, the stimulus-mismatch hypothesis, and the orienting response hypothesis.

Data Selection.

The current study made use only of existing published data-sets and as such was deemed exempt from the need for ethical review by the institutional ethics board. On 16th April 2018, Google Scholar listed 133 papers which made reference to the Tremblay and Jones (1998) study. Of these, 24 papers were identified as containing comparisons of conditions with different token set sizes within the irrelevant sound. Note that we became aware of the study by Bell *et al.* (2019a) which explicitly addresses the token set-size effect from a statistical power perspective only in the process of writing this paper and so this was also incorporated into our analysis. Of the papers so identified, 6 compared only set sizes of 2 (i.e., a straight comparison of 2-token changing-state conditions with steady-state or quiet controls) but data were available in a form open to analysis from 6 papers either directly or by request to the author.³ Of these papers, the study by Campbell *et al.* (2003) was excluded because the relevant comparison did not find a changing-state effect with set size 2, which is not predicted by the hypotheses under scrutiny, and gives no basis within this system for

³ We thank Emily Elliott for making the data from Elliott and Cowan (2005) available to us.

estimating the effect size of greater token set size effects. The general discussion shall return to this issue, and to the possibility of file-drawer effects (Simonsohn, Nelson, & Simmons, 2014; Rosenthal, 1979).

The final data-set thus consists of 14 comparisons across 10 experiments, as shown in Table 1. As examples of the form taken by the data under consideration, Figure 2 shows data from two of the relevant papers that have not previously been drawn in a form allowing for the direct comparison of set size effects (e.g., in the case of Bell *et al.*, 2019a, all data were presented as a function of serial position so the difference between conditions when collapsed across serial position was indeterminate). Comparable bar charts can already be consulted in the results sections of other relevant papers (e.g., Campbell *et al.*, 2002),

FIGURE TWO ABOUT HERE

Results.

The prediction of the strong changing-state hypothesis was a null prediction beyond the changing-state point. That is, there was predicted to be an effect of precisely zero for comparisons between any changing-state conditions.

The predictions of the orienting and stimulus-mismatch hypotheses were based upon calculations similar to those given in Figure 1. The prediction of the orienting hypothesis for a given experiment was taken to be a constant reapplication of the percentage decrease in performance observed between the control condition and set size 2 every time the set size was incremented by one. For example, if the decrement in performance at set size 2 was 20% lower than control in a given experiment this was used to inform the predictions such that a further 20% decrement was expected on top of this at any set size 3 condition in that experiment, and so on. Thus, if performance at control was 100% and at set size 2 was 80%

then the performance at set size 3 was expected to be 64%, at set size 4 it would be expected to be 51.2% and so on.

The calculated prediction of the stimulus mismatch hypothesis was taken to be the absolute disruption observed at set size 2 and subsequently halved each time the token set size was incremented by one. For example, if the decrement in performance at set size 2 was 20% lower than control, this was used to inform the predictions such that a further 10% decrement was expected on top of this at any set size 3 condition in that experiment, a further 5% at any set size 4 condition and so on. Thus, if performance at control was 100% and at set size 2 it was 80%, then the performance at set size 3 was expected to be 70%, at set size 4 it would be expected to be 65% and so on.

A half-normal distribution was chosen to reflect the plausibility of different population mean values, as described by Dienes (2014). Bayes factors were calculated using the Dienes calculator obtainable at:

http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm. This calculator was coded in Matlab and converted to Flash. Calibration shows that alternative code written independently in R (Baguley & Kaye, 2010) gives identical results.

Once calculated, Bayes factors were combined meta-analytically for those studies (Campbell et al., 2002; Elliott & Cowan, 2005) which made use of the same stimuli and the same set size contrasts. In each case, data from the first study reported in the series was used to inform the prior mean and prior SD, data from the second study reported was used to inform the likelihood's mean and SD, and the two combined to calculate the posterior mean and posterior SD following the first two studies. The posterior mean and SD thus obtained are then entered as the new prior mean and SD and the mean and SE from Study 3 used as the likelihood's mean and SD and so on until the data from all available experiments was combined (Dienes, 2008, 2014; Parmentier & Beaman, 2015). In principle, the same

procedure could be used to combine data across all of the studies examined here but to do so would involve combining data from the different experimental stimuli and procedures (e.g., different list lengths and therefore differences in baseline difficulty) employed by the various labs and therefore a purer approach which avoids these potential confounds was adopted here. We will return to the interpretation of the multiple Bayes factors so obtained in the General Discussion.

TABLE TWO ABOUT HERE

Orienting response versus strong changing-state. Examination of column 3 of Table 2 shows that the Bayes factors in favour of the orienting response hypothesis over the strong changing-state hypothesis are mixed. The combined results of the two experiments by Campbell *et al.* (2002) support the strong changing-state hypothesis over the orienting response account (notwithstanding the statistically significant disruptive advantage for increments in set size beyond two reported for one experiment in the original paper), however the combined results of the three experiments by Elliott and Cowan (2005) are strongly in favour of orienting over strong changing-state. Comparisons originally reported by Tremblay and Jones (1998) are largely inconclusive, however comparisons of 3 versus 5 tokens are strongly supportive of the strong changing-state hypothesis over the orienting hypothesis both for speech and non-speech tokens. In contrast, the Bayes factors calculated for the studies by Bell *et al.* (2019a) provide very strong evidence for the orienting hypothesis over the strong changing-state hypothesis.

Stimulus mismatch versus strong changing-state. By contrast, column 4 of Table 2 shows the Bayes factors in favour of the stimulus-mismatch hypothesis over the strong changing-state

hypothesis. Although, again, a substantial number of the Bayes factors are inconclusive (5 in total are between 1/3 and 3, all of them derived from the Tremblay and Jones (1998) data) only one Bayes factor (a comparison of 3 vs 5 speech tokens by Tremblay and Jones) favours the strong changing-state hypothesis, the remainder support the stimulus-mismatch hypothesis. Juxtaposed with the previous set of comparisons, the combined results from both Elliott and Cowan (2005) and Campbell *et al.* (2002) substantially favour the alternative to the strong changing-state hypothesis. Once again, the results from Bell *et al.* (2019a) are against the strong changing-state hypothesis.

Stimulus-mismatch versus orienting hypotheses. The final set of comparisons (column 5 of Table 2) pits the two experimental hypotheses, which predict a non-zero effect of token set size against each other. Although the majority of Bayes factors are inconclusive in terms of magnitude, it is striking that all of the comparisons favor the stimulus-mismatch hypothesis, even if the evidence in any single comparison would only be classified as “substantial” in three out of the eleven cases. A frequentist sign-test shows that it is unlikely that evidence in favor of one outcome over another would be found in all eleven cases if the SMM and OR hypotheses were equally plausible ($p < .0001$).

General Discussion

The results summarized in Table 2, when taken *in toto*, unequivocally favour the stimulus-mismatch hypothesis over both the strong changing-state hypothesis and the orienting-response hypothesis. The data represent the first meaningful attempt to contrast multiple accounts of the irrelevant sound effect simultaneously and parametrically. As noted, we chose not to combine these results meta-analytically in the manner we combined the results of the individual experiments in the Campbell *et al.* (2002) and Elliott & Cowan

(2005) studies because the analyses we have presented show the relative likelihood of particular quantitative accounts in conditions where the parameters (e.g., the token set size conditions and the difference in set size between conditions) are specific to those particular investigations. Thus, the predicted effect size according to either orienting response or stimulus mismatch hypotheses (which in both cases is a function of such parameters) is similarly experiment-specific. Overall, however, the data do not favour the strong changing-state hypothesis. Five out of the eleven comparisons with the orienting-response account presented here numerically support the changing-state hypothesis but only weakly. One of the comparisons with the stimulus-mismatch hypothesis also supports the strong changing-state hypothesis but again, the evidence in this one comparison is not compelling and is not in line with the results of the other studies.

In contrast, it is notable that in most cases the evidence from individual studies is much stronger in favour of either the orienting-response or the stimulus-mismatch hypotheses over the strong changing-state hypothesis. The verbal labels researchers are willing to place on different quantitative outcomes vary with some (e.g., Raftery, 1995) being more conservative in their use of labels and, technically, Bayes factors provide a continuous measure of confidence in a hypothesis (Dienes & McLatchie, 2018) which should not be reduced to a dichotomous decision. Nonetheless, three of the comparisons between strong changing-state and orienting-response hypotheses produced evidence that would be labelled “substantial” by some authorities (Bayes factors >3) two would be considered “strong” evidence (Bayes factors >10) and one “decisive” (Bayes factor >100 ; all verbal labels taken from Jeffries, 1961).

The pattern of results is even more striking when the strong changing-state hypothesis is contrasted with the stimulus mismatch hypothesis. In these comparisons, two contrasts provide substantial evidence for the mismatch hypothesis, two contrasts provide “very

strong” evidence in the same direction (Bayes factors > 30 ; Jeffries, 1961) and two contrasts come out decisively in favour of the stimulus mismatch hypothesis. Thus, if anything, contrasting both stimulus mismatch and orienting hypotheses with the strong changing state hypothesis favours the stimulus mismatch rather than the orienting response hypothesis, although these are not direct comparisons between these two possibilities.

The data are also illuminating when the stimulus-mismatch hypothesis is directly contrasted with the orienting-response hypothesis. In this case, although the evidence is for many of the studies is weak (e.g., Bell *et al.*, 2019a: Exp. 1; Tremblay & Jones, 1998: Exp. 2) the direction of the Bayes factors favours the stimulus-mismatch hypothesis in all cases. Although at an individual level, only three of the contrasts provide Bayes factors which might be termed “substantial” evidence, it is noteworthy all eleven contrasts favour the same (SMM) hypothesis. Since Bayes factors reflect likelihood ratios for one position over another, the position is not analogous to finding a series of p-values where some statistically significant effects might be called into question by the presence of other more ambiguous results (e.g., $p=.06$) which are insufficient to reject the null hypothesis as only one hypothesis (the null) is under consideration in traditional testing. Consistently finding more – even if weak – evidence in favour of one position over another can only *strengthen* confidence in that position, albeit the extent to which confidence in the conclusions is heightened varies according to the strength of the new evidence. Weaker Bayesian evidence might push less strongly but it still pushes in the same direction. Goodman (1999, Table 1) provides examples of how far particular Bayes factors can move belief in a given hypothesis (the posterior probability) as a joint function of the size of the Bayes factor and the original level of belief (the prior probability) but here, in all cases, the posterior probability in favour of the stimulus mismatch hypothesis over the orienting response has to be revised upwards continuously as a consequence of the additional positive evidence (represented by Bayes

factor >1) each additional study provides. Thus, in this case, the minimum size of the combined Bayes factor is at least as large as the largest individual Bayes factor reported.

A concern might be that the function predicted by the orienting response is squashed by the presence of floor effects as token set size increases. This concern is unwarranted because – as examination of Figures 1 and 2 shows – performance is above floor even at the largest token set sizes. One could, perhaps, argue that “floor” is defined not by very poor percentage performance but, rather, by a minimum absolute number correct that is in some way inviolate to irrelevant sound effects. Thus, for example, it might be possible to be at floor and still score 50% correct if four out of eight items are correctly recalled and four items is the size of a focus of attention, which is immune to interference effects (e.g., Cowan, 1995). Again, however the data show that the same effects are observed with both speech and non-speech stimuli in the same studies. In passing, it should be mentioned that the speech stimuli resulted in greater interference than non-speech stimuli. Thus, even if the participants are performing at floor in irrelevant speech conditions they cannot then be at floor in irrelevant non-speech conditions, and yet the same results are obtained. This investigation therefore concludes that of the three functions considered here, the data favour that predicted by the stimulus-mismatch hypothesis.

Data selection and file-drawer effects.

As with any review, cautious assessment of whether there is a bias in the selection of data is necessary, which might present a misleading picture. Notably, there is the question of the file-drawer problem. This analysis presents the data from Tremblay and Jones (1998), which are the only data representative of null effects of token set size as the data from a similar investigation showing a null result by Hughes and Jones (2005) were not available, so the possibility must be considered that other datasets exist. Such datasets may have not

reached publication either because they were not submitted or rejected by journals as merely showing uninteresting null effects of token set size.

Although there is the known publication bias towards publishing “significant” effects, this tendency in science seems unlikely to have affected the results in this case for a number of reasons. The Tremblay and Jones (1998) data, which established token set size as a variable of interest, are themselves non-significant, with the majority of analyses within their paper showing null effects beyond a token set size of two. Their investigation was more comprehensive than later studies in terms of comparing multiple different set sizes, and contributes a full 50% of the contrasts considered in Table 1. It is therefore not the case that statistically nonsignificant effects are under-represented in these analyses. That is, in all but one of these contrasts the Bayes factor was numerically in favour of the stimulus-mismatch hypothesis rather than the strong changing-state hypothesis. Where the Bayes factors provide only weak or inconclusive evidence, this is also exactly where one would expect given the curvilinear function predicted by the stimulus-mismatch hypothesis and the resulting difference in effect sizes across comparisons: As token set size increases, the rate at which each token further increases disruption decelerates. Thus the difference in the predictions of the stimulus-mismatch and strong changing-state hypotheses also drops and it becomes more-and-more difficult for *any* data set to distinguish between the predictions on a purely pair-wise basis at higher set sizes.

One data-set which was considered and rejected from further analysis was that obtained by Campbell *et al.* (2003), which showed a greater effect of irrelevant sound at token set size five than at set size two. The reason for excluding these data was simply that, contrary to other reports in the literature, there was no changing-state effect at token set size two. That is, in this study two items were no more disruptive than one (and, numerically at least) less so. This dataset therefore gives no basis for comparing the hypotheses considered

here as the difference between set size one and two, the basic changing-state effect, was used to calculate the anticipated further increases in disruption with increasing token set size.

These data are not predicted by any of the hypotheses *considered here* and, in the absence of any further information, can be explained post hoc in a number of ways such as the difference between one and two items was insufficient to invoke a changing-state effect/orienting response/stimulus mismatch that only became apparent subsequently as set size increased. To be borne in mind is that this study used Finnish monosyllables as irrelevant speech items. A peculiarity of Finnish is that it contains only 23 monosyllabic words. In the experiment it is possible that Finnish participants grouped pairs of Finnish-sounding monosyllables into bi-syllables, as they are not used to Finnish monosyllables, thereby conflating the token set size. These data then serve best to reinforce the utility of providing more results in the form of systematic parametric manipulations rather than two or three conditions at a time.

Direction for Further Research

The stimulus-mismatch hypothesis favoured by the data according to the current analyses is novel, despite its similarity to previous ideas (Beaman, 2000; Campbell, 2000; Farrell & Lewandowsky, 2003). The hypothesis is intended at this point simply as a mathematical statement to describe how adding noise to a memory representation (possibly encoded as a vector of different feature values as in the feature model; Beaman, Neath & Surprenant, 2008; Nairne, 1990; Neath, 2000) is subject to a process of increasing redundancy. That process operates such that the noise associated with every additional object is increasingly less potent as a disruptor. A natural interpretation of this interference process is as a form of overwriting in memory, however we are reluctant to advance this as a hypothesis for reasons described in detail elsewhere (Beaman & Jones, 2016). Instead the

interpretation offered is that noisier or more energetic storage, possibly with irrelevant sound items blocking easy access to memoranda, renders retrieval more problematic regardless of the precise nature of the noise. A backward masking effect of irrelevant sound would explain why distraction effects can be seen both when the sound is played at encoding or during a post-encoding maintenance interval (e.g., Miles, Jones & Madden, 1991) but not when the sound immediately precedes the memoranda (Macken, Mosdell & Jones, 1999). This retrospective effect requires further assumptions to be made about rehearsal or other maintenance processes in order to be reconciled with either feature overwriting (or feature adoption; Neath, 2000) or orienting-based explanations. For example, for feature overwriting or feature adoption to work, there must be some form of contiguity between the overwriting and the overwritten material which does not automatically occur if the auditory distractors are presented only within a retention interval. Similarly, for orienting hypotheses one must assume that attention is oriented away from maintenance-rehearsal processes during this same retention period.

A more difficult proposition for the stimulus-mismatch hypothesis is the question why a single repeated (“steady-state”) item is frequently not statistically significant. Beyond set size two, the rapid deceleration of the rate at which disruption increases with token set size would explain why the patterns obtained by Tremblay and Jones (1998) were not statistically significant. This steady-state effect is not significant in designs powerful enough to routinely pick up a changing-state effect (Jones & Macken, 1993; Jones *et al.*, 1992; but see also Bell *et al.*, 2019b; Parmentier & Beaman, 2015), rather this steady-state effect dances either side of significance, suggesting an effect size for which many published studies are underpowered. One possibility is that the cognitive interference caused by steady-state information is at a level from which recovery is usually possible – at least partially – and only beyond this point is recall itself seriously affected (e.g., Beaman, 2000). This

explanation parallels the suggestion that orienting to a single utterance habituates sufficiently to reduce the effect size – an intuitively compelling corollary to the OR hypothesis. However, both suggestions – reasons why recovery might be possible from SMM interference and the possible rate of habituation to an OR remain to be investigated.

In summary, the current data support a version of auditory distraction in which there is a gradual but decelerating increase in distraction as a function of the number of distracters. This pattern of data is less likely under the strong changing-state hypothesis and the orienting-response hypothesis than under the stimulus-mismatch hypothesis as operationalized here. The approach taken, both in terms of the Bayesian analyses used, and their application to a function governing auditory distraction instead of nil hypothesis tests aimed at showing qualitative differences, opens new ways of examining auditory distraction data. As a method, the approach is particularly suitable to examining multiple accounts of effects for which predictions can be quantified – particularly, as here, where the manipulation itself is inherently parametric. Suitable effects amenable to such investigations include the quantifiable effects of word “dose” (the total number of utterances experienced, irrespective of token set size; Bridges & Jones, 1996) and of “babble” – the effect of speech or noise from multiple different sound sources and how these interact with the ability to identify and segregate the different sound streams (Jones et al., 1999; Jones & Macken, 1995). It is also worth noting that a stimulus-mismatch hypothesis is more consistent than a strong changing-state hypothesis with data showing that the extent of changes-in-state within a single stream influences the size of the effect (Beaman & Holt, 2007; Jones & Macken, 1995). Such data are difficult to reconcile with the doctrine that it is the *fact* of change rather than the *extent* of change which is important.

A critical precursor to further investigations, however, is to develop mechanistic accounts which can be shown to give rise to particular disruptive functions from first

principles. Presented here is evidence that a particular mathematical function represents a more likely basis for the data examined than two alternative mathematical functions but this represents the first attempt to systematically make such comparisons. The current lack of detail in theorizing forced us to make a number of assumptions – for example the only restriction on orienting response predictions put forward by Bell et al. (2019a) is that it should be monotonic but nonlinear in response to increases in token-set size, which aligns it with the “weak” changing-state hypothesis considered but rejected by Tremblay and Jones (1999) and against their assumption that orienting functions should be strictly linear which Tremblay and Jones also rejected after testing. Thus, the current study is open to the criticism that the functions we examined are not the “real” predictions of these hypotheses but, if so, then absent such more detailed mechanistic accounts there is – as yet – no solid basis on which to make any testable predictions. The current analysis however, indicates what functions any mechanistic account should aim to generate.

Finally, there remains the question of task-specificity. The changing-state hypothesis was initially embedded in the object-oriented episodic record model of short-term memory (Jones et al., 1996) which focussed on serial order retention but assumed amodal, object-based representations. This model has since been superseded by the perceptual-gestural model (Jones, Hughes & Macken, 2006) which retains a reliance upon maintaining serial order but shifts towards a more perception-for-action, embodied idea of representation (e.g., Macken, Taylor, Kozlov, Hughes & Jones, 2016). Note that the changing-state *hypothesis* that a requirement to maintain serial order is necessary to observe a changing-state effect is independent of either of these models. One could assume that the representational assumptions of the perceptual-gestural model to be correct while simultaneously rejecting the idea of a conflict of order cues as the basis of auditory distraction and, as we have seen, the hypothesis could be incorporated into multiple models with different representational

assumptions. As such, it is the strong changing-state hypothesis which is tested here, with the results as discussed. It is worth noting also that the changing-state *effect* – which we have reformulated as a subset (indeed, the starting point) of the token set size effect – has not to our knowledge anywhere been disputed. Similarly, there is no argument that orienting sometimes happens, and it is also possible to build orienting responses into multiple memory models, but critics of the orienting response hypothesis argue simply that orienting (and subsequent habituation) are not the basis for the auditory distraction effects which give rise to changing-state and token set-size effects. A difficulty for all current models is how to account for the extensive data on task-specificity without becoming entangled in circular logic.

For changing-state theorists, it is possible to argue that tasks which require serial order maintenance strategies are those which show changing-state effects, and that tasks which show changing-state effects are those which involve serial order maintenance strategies. For orienting response theorists, it is possible to argue that tasks which require attention show distraction effects when attention is oriented away, and that tasks which do not show distraction effects are insufficiently attention-demanding for orienting away to cause distraction. A possible way of breaking this circularity is to take the memory-blocking suggestion of the stimulus mismatch hypothesis seriously, and argue that tasks which require access to representations of a kind which are either intrinsically similar to the blocking material (i.e., other auditory representations) or can be translated into an auditory-motor output (most typically, vocalized) are those which should show distraction effects. Critically, for all tasks with these characteristics, one would expect to see a token set-size effect of the kind shown in Figure 1A.

References

- Baguley, T., & Kaye, D. (2010). Review of “Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference”. *British Journal of Mathematical & Statistical Psychology*, *63*, 695–698. <http://dx.doi.org/10.1348/000711009X481027>
- Banbury, S., & Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied*, *3*, 181–195. <http://dx.doi.org/10.1037/1076-898X.3.3.181>
- Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology*, *89*, 499–517. <http://dx.doi.org/10.1111/j.2044-8295.1998.tb02699.x>
- Beaman, C. P. (2000). Computational explorations of the irrelevant sound effect in serial short-term memory. *In Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society: August 13-15, 2000, University of Pennsylvania, Philadelphia, PA.*
- Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 1106–1118. <https://doi.org/10.1037/0278-7393.30.5.1106>
- Beaman, C. P. (2005). Auditory distraction from low-intensity noise: a review of the consequences for learning and workplace environments. *Applied Cognitive Psychology*, *19*, 1041–1064. <https://doi.org/10.1002/acp.1134>
- Beaman, C. P., Bridges, A. M., & Scott, S. K. (2007). From dichotic listening to the irrelevant sound effect: A behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex*, *43*, 124–134. [https://doi.org/10.1016/S0010-9452\(08\)70450-7](https://doi.org/10.1016/S0010-9452(08)70450-7)

- Beaman, C. P., Hanczakowski, M., Hodgetts, H. M., Marsh, J. E., & Jones, D. M. (2013). Memory as discrimination: What distraction reveals. *Memory & Cognition*, *41*, 1238–1251. <https://doi.org/10.3758/s13421-013-0327-4>
- Beaman, C. P., & Holt, N. J. (2007). Reverberant auditory environments: the effects of multiple echoes on distraction by ‘irrelevant’ speech. *Applied Cognitive Psychology*, *21*, 1077–1090. <https://doi.org/10.1002/acp.1315>
- Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 459–471. <http://dx.doi.org/10.1037/0278-7393.23.2.459>
- Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology*, *51A*, 615–636. <https://doi.org/10.1080/713755774>
- Beaman, C. P., & Jones, D. M. (2016). The item versus the object in memory: on the implausibility of overwriting as a mechanism for forgetting in short-term memory. *Frontiers in Psychology*, *7*, 341. <https://doi.org/10.3389/fpsyg.2016.00341>
- Beaman, C. P., Neath, I., & Surprenant, A. M. (2008). Modeling distributions of immediate memory effects: No strategies needed? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *34*, 219–229. <https://doi.org/10.1037/0278-7393.34.1.219>
- Bell, R., Mieth, L., Röer, J. P., & Buchner, A. (2019). Effects of auditory distraction on face memory. *Scientific Reports*, *9*, 10185, <https://doi.org/10.1038/s41598-019-46641-7>
- Bell, R., Röer, J. P., Dentale, S., & Buchner, A. (2012). Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of*

Experimental Psychology: Learning, Memory & Cognition, 38, 1542–1557.

<https://doi.org/10.1037/a0028459>

Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019a). Reassessing the token set size effect on serial recall: Implications for theories of auditory distraction. *Journal of Memory & Language*, 45, 1432-1440. <https://doi.org/10.1037/xlm0000658>

Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019b). Distraction by steady-state sounds: Evidence for a graded attentional model of auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 500–512. <https://doi.org/10.1037/xhp0000623>.

Bridges, A. M., & Jones, D. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing state? *Quarterly Journal of Experimental Psychology*, 49A, 919–939. <https://doi.org/10.1080/713755663>

Campbell, T. A. (2000). *The disruptive effects of irrelevant sound in perception and memory* (Doctoral thesis). Retrieved from EThOS. (uk.bl.ethos.270215)

Campbell, T., Beaman, C. P., & Berry, D. C. (2002). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory*, 10, 199–214. <https://doi.org/10.1080/09658210143000335>

Campbell, T., Winkler, I., & Kujala, T. (2005). Disruption of immediate memory and brain processes: an auditory ERP protocol. *Brain Research Protocols*, 14, 77–86. <https://doi.org/10.1016/j.brainresprot.2004.11.001>

Campbell, T., Winkler, I., & Kujala, T. (2007). N1 and the mismatch negativity are spatiotemporally distinct ERP components: Disruption of immediate memory by auditory distraction can be related to N1. *Psychophysiology*, 44, 530–540. <https://doi.org/10.1111/j.1469-8986.2007.00529.x>

- Campbell, T., Winkler, I., Kujala, T., & Näätänen, R. (2003). The N1 hypothesis and irrelevant sound: evidence from token set size effects. *Cognitive Brain Research, 18*, 39–47. <https://doi.org/10.1016/j.cogbrainres.2003.09.001>
- Cohen, J. (1995). The earth is round ($p < .05$): Rejoinder. *American Psychologist, 50*, 1103. <https://doi.org/10.1037/0003-066X.49.12.997>
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York: Palgrave Macmillan,
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (2019). How do I know what my theory predicts? *Advances in Methods and Practices in Psychological Science* <https://doi.org/10.31234/osf.io/yqaj4>
- Dienes, Z., & McLatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25*, 207–218. <https://doi.org/10.3758/s13423-017-1266-z>
- Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in irrelevant speech? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 1406–1414. <https://doi.org/10.1037/0096-1523.24.5.1406>
- Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. *Journal of the Acoustical Society of America, 102*, 2191–2199 <https://doi.org/10.1121/1.419596>

- Elliott, E. M. (2002). The irrelevant speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*, 478–487.
<https://doi.org/10.3758/BF03194948>
- Elliott, E. M., & Cowan, N. (2001). Habituation to auditory distractors in a cross-modal, color–word interference task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 654–667. <http://dx.doi.org/>
- Elliott, E. M. & Cowan, N. (2005). Coherence of the irrelevant-sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory & Cognition*, *33*, 664–675. <https://doi.org/10.3758/BF03195333>
- Farley, L. A., Neath, I., Allbritton, D. W., & Surprenant, A. M. (2007). Irrelevant speech effects and sequence learning. *Memory & Cognition*, *35*, 156–165.
<https://doi.org/10.3758/BF03195951>
- Gisselgård, J., Uddén, J., Ingvar, M., & Petersson, K. M. (2007). Disruption of order information by irrelevant items: A serial recognition paradigm. *Acta Psychologica*, *124*, 356–369. <https://doi.org/10.1016/j.actpsy.2006.04.002>
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 2: the Bayes factor. *Annals of Internal Medicine*. *130*, 1005–1013. doi: 10.7326/0003-4819-130-12-199906150-00019
- Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 838–849. <https://doi.org/10.1037/0278-7393.29.5.838>
- Hanczakowski, M., Beaman, C. P. & Jones, D. M. (2017) When distraction benefits memory through semantic similarity. *Journal of Memory and Language*, *94*, . 61-74.
<https://doi.org/10.1016/j.jml.2016.11.005>

- Hellbrück, J., Kuwano, S., & Namba, S. (1996). Irrelevant background speech and human performance: Is there long-term habituation? *Journal of the Acoustical Society of Japan (E)*, *17*, 239–247. <https://doi.org/10.1250/ast.17.239>
- Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology Section*, *56A*, 1307–1334. <https://doi.org/10.1080/02724980244000747>
- Hughes, R. W. (2014). Auditory distraction: A duplex mechanism account. *Psych J*, *3*, 30–41. <http://dx.doi.org/10.1002/pchj.44>
- Hughes, R. W., & Jones, D. M. (2005). The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human Perception & Performance*, *31*, 316–327. <https://doi.org/10.1037/0096-1523.31.2.316>
- Hughes, R. W., & Marsh, J. E. (2020). When is forewarned forearmed? Predicting auditory distraction in short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* *46*, 427–442. <http://dx.doi.org/10.1037/xlm0000736>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*, 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Jeffries, H. (1961). *Theory of probability*. Oxford: OUP.

- Jones, D. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control*. Oxford, England: Clarendon Press.
- Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 464–473. <https://doi.org/10.1037/0278-7393.25.2.464>
- Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. In S.E. Gathercole (Ed.), *Models of short-term memory*. London: Lawrence Erlbaum Associates Ltd.
- Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory & Language*, *54*, 265-281. <https://doi.org/10.1016/j.jml.2005.10.006>
- Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology*, *44A*, 645–669. <https://doi.org/10.1080/14640749208401304>
- Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*, 369–381. <https://doi.org/10.1037/0278-7393.19.2.369>
- Jones, D. M., & Macken, W. J. (1995). Phonological similarity in the irrelevant speech effect: Within-or between-stream similarity? *Journal of Experimental Psychology:*

Learning, Memory, & Cognition, 21, 103–115. <https://doi.org/10.1037/0278-7393.21.1.103>

Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied*, 1, 216–226. <https://doi.org/10.1037/1076-898X.1.3.216>

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition*, 23, 192–200. <https://doi.org/10.3758/BF03197221>

Jones, D. M., Macken, W. J., & Harries, C. (1997). Disruption of short term recognition memory for tones: streaming or interference? *Quarterly Journal of Experimental Psychology Section*, 50A, 337–357. <https://doi.org/10.1080/713755707>

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, 4, 89–108. <https://doi.org/10.1002/acp.2350040203>

Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen. *Zeitschrift für Experimentelle Psychologie*, 42, 517–544.

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory & Language*, 53, 513–531. <https://doi.org/10.1016/j.jml.2005.07.002>

Larsen, J. D., Baddeley, A., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory*, 8, 145–157. <https://doi.org/10.1080/096582100387579>

LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, 2, 391–397. <https://doi.org/10.3758/BF03210978>

- LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 1154–1165. <http://dx.doi.org/>
- LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology*, *50A*, 100-18. DOI: 10.1080/713755679
- Logie, R. H., & Baddeley, A. D. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 310–327. <https://doi.org/10.1037/0278-7393.13.2.310>
- Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 436–448. <https://doi.org/10.1037/0278-7393.21.2.436>
- Macken, W. J., Mosdell, N., & Jones, D. M. (1999). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 810–814. <https://doi.org/10.1037/0278-7393.25.3.810>
- Macken, W. J., Taylor, J. C., Kozlov, M., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, *155*, 113-124. <https://doi.org/10.1016/j.cognition.2016.06.013>
- Marsh, J. E., & Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: The new early filter model. *Frontiers in Neuroscience*, *10*, 136. <https://doi.org/10.3389/fnins.2016.00136>
- Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory & Language*, *58*, 682–700. <https://doi.org/10.1016/j.jml.2007.05.002>

- Marsh, J. E., Hughes, R. W., & Jones, D. M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition, 110*, 23–38.
<https://doi.org/10.1016/j.cognition.2008.08.003>
- Marsh, J. E., Perham, N., Sörqvist, P., & Jones, D. M. (2014). Boundaries of semantic distraction: Dominance and lexicality act at retrieval. *Memory & Cognition, 42*, 1285–1301. <https://doi.org/10.3758/s13421-014-0438-6>
- Marsh, J. E., Sörqvist, P., Hodgetts, H. M., Beaman, C. P., & Jones, D. M. (2015). Distraction control processes in free recall: Benefits and costs to performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 118–133.
<https://doi.org/10.1037/a0037779>
- Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1272–1299.
<https://doi.org/10.1037/0278-7393.25.5.1272>
- Miles, C., Jones, D. M., & Madden, C. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 578–584. <https://doi.org/10.1037/0278-7393.17.3.578>
- Morris, N., & Jones, D. M. (1990). Habituation to irrelevant speech: Effects on a visual short-term memory task. *Perception & Psychophysics, 47*, 291–297.
<https://doi.org/10.3758/BF03205003>
- Mutschler, I., Wieckhorst, B., Speck, O., Schulze-Bonhage, A., Hennig, J., Seifritz, E., & Ball, T. (2010). Time scales of auditory habituation in the amygdala and cerebral cortex. *Cerebral Cortex, 20*, 2531–2539. <https://doi.org/10.1093/cercor/bhq001>
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269. <https://doi.org/10.3758/BF03213879>

- Näätänen, R. (1979). Orienting and evoked potentials. In: H. D. Kimmel, E. H. van Olst, & J. E. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 61–75). Hillsdale, NJ: Erlbaum.
- Näätänen, R. (1986). The orienting response: A combination of informational and energetical aspects of brain function. In: G. R. J. Hockley, A. W. K. Gaillard, & M. G. H. Coles (Ed.s) *Energetics and human information processing*. (pp. 91–111). Dordrecht: Martinus Nijhoff Publishers.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7, 403–423. <https://doi.org/10.3758/BF03214356>
- Neath, I., Guérard, K., Jalbert, A., & Surprenant, A. M. (2009). Irrelevant speech effects and statistical learning. *Quarterly Journal of Experimental Psychology*, 62, 1551–1559. <https://doi.org/10.1080/17470210902795640>
- Oberauer, K., & Kliegl, R. (2004). Simultaneous cognitive operations in working memory after dual-task practice. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 689–707. <https://doi.org/10.1037/0096-1523.30.4.689>
- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., ... & Ma, W. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, 144, 885–958. <https://doi.org/10.1037/bul0000153>
- Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, 19, 779–819. <https://doi.org/10.3758/s13423-012-0272-4>
- Oswald, C. J., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, 8, 345–350. <https://doi.org/10.1080/09658210050117762>

- Öhman, A. (1979). The orienting response, attention, and learning: An information-processing perspective. In H. D. Kimmel, E. H. van Olst, & J. E. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 443–471). Hillsdale, NJ: Erlbaum.
- Page, M. P. A., & Norris, D. (2003). The irrelevant sound effect: What needs modelling and a tentative model. *Quarterly Journal of Experimental Psychology*, *56A*, 1289–1300. <https://doi.org/10.1080/02724980343000233>
- Parmentier, F. B. R., & Beaman, C. P. (2015). Contrasting effects of changing rhythm and content on auditory distraction in immediate memory. *Canadian Journal of Experimental Psychology*, *69*, 28–38. <https://doi.org/10.1037/cep0000036>
- Perham, N., & Banbury, S. P. (2008). You cannot ignore it: Attention to ‘irrelevant’ sound during a habituation period does not produce habituation. In *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society*. New York, NY: HFES.
- Raftery A.E. (1995): Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. <http://dx.doi.org/10.2307/271063>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Röer, J. P., Bell, R., & Buchner, A. (2014). Evidence for habituation of the irrelevant-sound effect on serial recall. *Memory & Cognition*, *42*, 609–621. <https://doi.org/10.3758/s13421-013-0381-y>
- Röer, J. P., Bell, R., Dentale, S., & Buchner, A. (2011). The role of habituation and attentional orienting in the disruption of short-term memory performance. *Memory & Cognition*, *39*, 839–850. <https://doi.org/10.3758/s13421-010-0070-z>
- Röer, J. P., Bell, R., Körner, U., & Buchner, A. (2019). A semantic mismatch effect on serial recall: Evidence for interlexical processing of irrelevant speech. *Journal of*

Experimental Psychology: Learning, Memory & Cognition, 45, 515–525.

<https://doi.org/10.1037/xlm0000596>

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, 21, 150–164. [https://doi.org/10.1016/S0022-5371\(82\)90521-7](https://doi.org/10.1016/S0022-5371(82)90521-7)

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Macmillan.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, 143, 534–547.

<https://doi.org/10.1037/a0033242>

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4, 215–230. [https://doi.org/10.1016/0010-0277\(76\)90018-4](https://doi.org/10.1016/0010-0277(76)90018-4)

Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 24, 659–671.

Underwood, G. (1974). Moray vs. the rest: The effects of extended shadowing practice. *Quarterly Journal of Experimental Psychology*, 26, 368–372.

<https://doi.org/10.1080/14640747408400426>

Warren, R. M., & Obusek, C. J. (1972). Identification of temporal order within auditory sequences. *Perception & Psychophysics*, 12, 86–90.

<https://doi.org/10.3758/BF03212848>

Author Note

C. Philip Beaman, School of Psychology and Clinical Language Sciences, University of Reading. Tom Campbell, Faculty of Information Technology and Communication Sciences, Tampere University; Department of Psychology and Logopedics, University of Helsinki. John E. Marsh, School of Psychology, University of Central Lancashire; Department of Business Administration, Technology and Social Sciences, Luleå University of Technology.

Correspondence should be addressed to: C. Philip Beaman, School of Psychology and Clinical Language Sciences, Harry Pitt Building, University of Reading, Earley Gate, Reading, Berkshire, United Kingdom, RG6 6ES. Email: c.p.beaman@reading.ac.uk. Tel: +44 (0) 118 3787637

We thank Zoltan Dienes for advice and encouragement, Graham Schafer for the original suggestion that token set size might follow a power function, and Emily Elliott for making data available to us.

Sample Size (N)	Token Difference						Total No. of Comparisons at N
	1	2	3	4	5	6	
21-30	0/1	0/2	2/3	0/1	-	0/1	8
31-40	-	-	-	0/2	-	-	2
41+	0/1	0/2	-	-	0/1	2/2	6
% age significant results	0%	0%	67%	0%	0%	67%	16
Total Sample Size	72	144	70	101	80	300	

Table 1. Frequency (numerator) of statistically significant results reported for comparisons of a particular token difference and the total number of studies (denominator) for which that comparison is reported. Data are given as a function of the sample size for the experiment in question and the difference in token set size being compared (e.g., token difference 2 could be a comparison between 4 and 2 tokens or between 5 and 3 etc).

Source	Contrast	Bayes Factor OR vs SCS	Bayes Factor SMM vs SCS	Bayes Factor SMM vs OR
Tremblay & Jones (1998) Exp 1	Speech, 2 vs 3	1.39*	3.77*	2.85*
	Speech, 3 vs 5	0.24	0.86	5.55*
	Speech, 5 vs 7	0.42	1.03*	2.42*
Tremblay & Jones (1998) Exp 2	Tones, 2 vs 3	2.51*	3.97*	1.51*
	Tones, 3 vs 5	.21	1.11*	4.09*
	Tones, 5 vs 7	.56	1.01*	1.85*
Tremblay & Jones (1998) Exp 3	Speech, 2 vs 7	1.07*	2.44*	2.36*
Campbell, Beaman, & Berry (2002) Exps 3A & 3B combined	Speech, 2 vs 5	.21	170.52*	4.98*
Elliott & Cowan (2005) Exps 1D, 1E, & 1F combined	Tones, 4 vs 8	21.55*	46.74*	2.17*
Bell, Röer, Lang, & Buchner (2019) Exp 1	Speech, 2 vs 8	15981245437.29*	17137777433.56*	1.07*
Bell, Röer, Lang, & Buchner (2019) Exp 2	Notes, 2 vs 8	17.99*	42.14*	2.34*

Table 2. Summary of the evidence in favour of strong changing-state (SCS) orienting (OR) and stimulus mismatch (SMM) hypotheses when direct compared with each other across 14 different contrasts. In this notation, Bayes factors < 1.0 favour the first hypothesis and Bayes factors > 1.0 favour the second hypothesis. The Bayes factor is a continuous measure of relative confidence in an hypothesis but as a basic guide some authors have suggested that only Bayes factors > 3 (or $< 1/3$) give “substantial” or “positive” support for a hypothesis (Jeffreys, 1961; Raftery, 1995). In this table, all the Bayes factors which favour the first of two hypotheses are annotated with an asterisk.

Figure Legends

Figure 1. Panel A: Idealized data on the effects of irrelevant sound according to strong changing state, stimulus mismatch and orienting hypotheses. Baseline (in quiet) was assumed to be 50% error rate and the effects of two token (changing-state) irrelevant sound was assumed to be a 30% increase in error rate for all cases (the average disruptive effect of irrelevant speech according to Ellermeier & Zimmer, 1997). For the strong changing state hypothesis, no further disruption is assumed beyond this. For the stimulus mismatch hypothesis, the additional disruption of recall is given by halving the previous increase in disruption and adding it to the total disruption. For the orienting hypothesis, the additional disruption of recall is given by an additional reduction of recall rate by 30% at every subsequent set size. *Panel B:* Adapted from data from Experiments 1-2 of Tremblay & Jones (1998, Lines of best-fit are 3rd order polynomials ($R^2 = 0.9998$ and 0.9995 for speech and tones respectively)).

Figure 2. Data redrawn from studies by Elliott and Cowan (2005) and Bell et al. (2019a) showing the effects of token set size in these studies when collapsed across serial position.. *A-C* show data from Elliott & Cowan, 2005; *D* shows data from Bell et al. (2019a).

FIGURE ONE

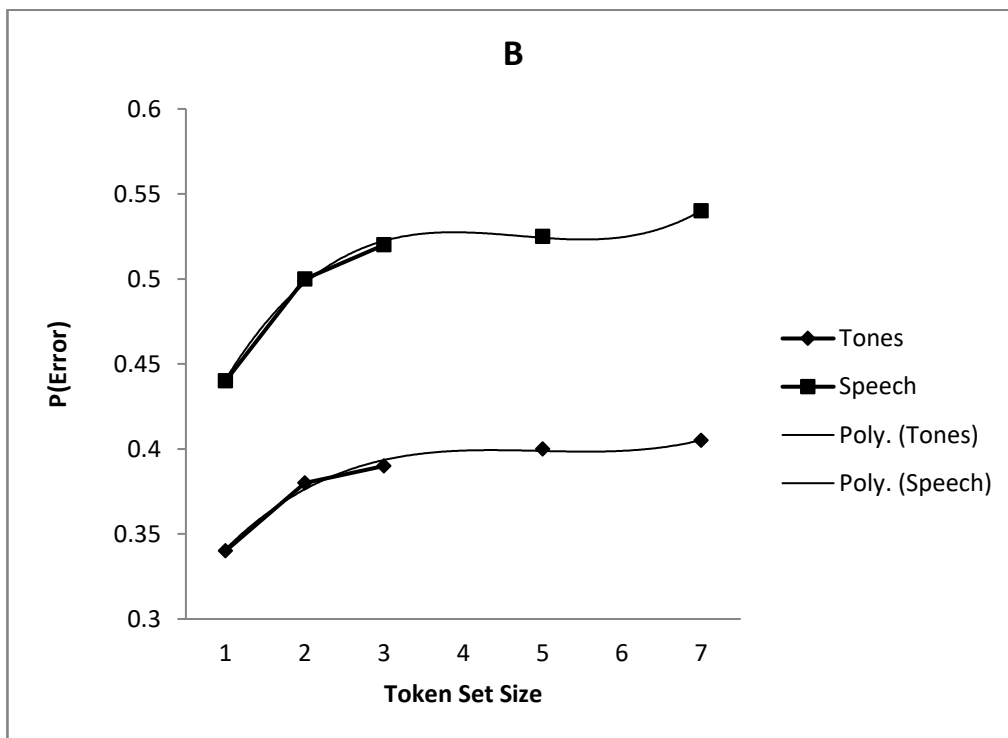
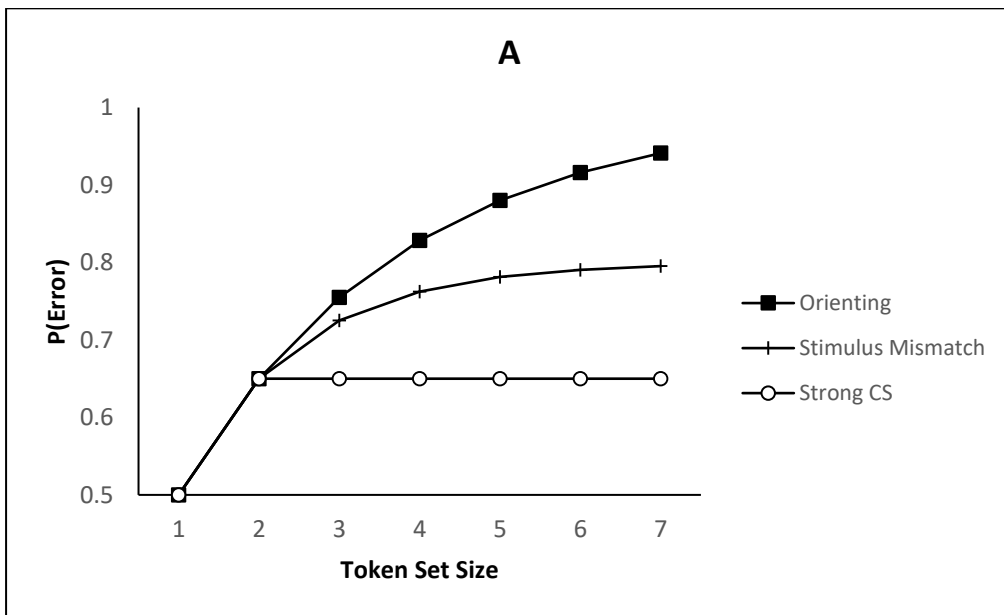


FIGURE TWO

