Why is accurately labelling simple magnitudes so hard? A past, present and future look at simple perceptual judgment

Chris Donkin[1], Babette Rae[2] Andrew Heathcote[2], & Scott D. Brown[2]
[1] School of Psychology, University of New South Wales, Australia
[2] School of Psychology, University of Newcastle, Australia

Abstract

Absolute identification is a deceptively simple task that has been the focus of empirical investigation and theoretical speculation for more than half a century. Observers are shown a set of N stimuli varying on a single dimension (e.g., length or loudness) and each stimulus is given a label (e.g., $1, .., N$). They then attempt to identify stimuli presented one at at time by producing the associated label. Since Miller's (1956) seminal paper the puzzle of why people are severely limited in their capacity to accurately perform absolute identification has endured. Despite the apparent simplicity of absolute identification, many complicated and robust effects are observed in both response latency and accuracy, including capacity limitations, strong sequential effects and effects of the position of a stimulus within the set. Constructing a comprehensive theoretical account of these benchmark effects has proven difficult, and existing accounts all have shortcomings in one way or another. We review classical empirical findings, as well as some newer findings that challenge existing theories. We then discuss a variety of theories, with a focus on the most recent proposals, make some broad conclusions about general classes of models, and discuss the challenges ahead for each class.

Absolute or perfect pitch, the ability to identify the notes played on a musical instrument, is a very rare ability, and has been a subject of scientific study since the late 19th century (Ellis, 1876). It is surprising that identifying musical notes is so difficult, since humans routinely identify a huge number of things in day-to-day life: such as faces, voices, and places. Throughout the first half of the 20th century, however, psychology researchers uncovered that the difficulty most people experience in naming notes is representative of a general deficit in identifying simple stimuli that vary on just one dimension. This early research culminated in Miller's (1956) seminal "7 ± 2" paper, in which he argued that humans were capable of accurately identifying only 5 to 9 stimuli that varied on a single dimension, regardless of the modality of that stimulus. Miller's work also made prominent the field of study that is the focus of this chapter, absolute identification.

In an absolute identification task, a set of N stimuli that vary on a single physical dimension are assigned a set of labels (usually the numbers 1 through N). On any given trial, participants are presented with one stimulus and asked to produce the corresponding label. For example, absolute pitch is a version of absolute identification in which the stimuli are tones varying in frequency and the labels are the musical note names A#, C, and so on. Other common versions of absolute identification use lines varying in length, or pure tones varying in loudness.

Since Miller's (1956) work, studies of absolute identification using a variety of stimulus dimensions and perceptual modalities have revealed an intricate pattern of phenomena behind this puzzling limitation in human ability. The complexity of the behavior elicited by such a seemingly simple task has ensured the enduring interest of the area, for example as summarized by Shiffrin and Nosofsky (1994). Absolute identification has also been of interest because of its links to other key areas of psychology, such as categorization – absolute identification is just categorization with one item per category (Nosofsky, 1986, 1997), and magnitude production – the reverse of absolute identification: given a label, participants try to produce the stimulus (De Carlo & Cross, 1990; Zotov, Jones, & Mewhort, 2011). More tantalizing is the suggested link between absolute identification and basic short term memory research. As Miller pointed out, both paradigms focus on memory, and share the same severe performance limit (7 ± 2), suggesting a common, and deep-seated, cognitive mechanism (this is also suggested by recently identified links in the sequential effects between short term memory and absolute identification - Malmberg & Annis, 2012).

This extensive study of absolute identification has yielded a wide range of robust benchmark phenomena not only in terms of the accuracy of responses but also in dynamic aspects, such as the time to make responses and the effect of previous responses on subsequent responses. Such a richness of data makes absolute identification a difficult challenge for cognitive modeling. We begin this chapter by giving a overview of the benchmark phenomena. These classical benchmarks focus on response accuracy. We then summarize key theoretical approaches that have been applied to absolute identification and describe how the latest models within these approaches have expanded their explanatory reach to response time (RT) as well as accuracy. We finish by discussing some of the recent, and ongoing, issues in the field.

Benchmark Phenomena

Capacity Limitations

Early in the history of absolute identification researchers were particularly intrigued by the difficulty of the task. A common method of assessing performance limits was to increase the size of the stimulus set, beginning with just two stimuli, and observe classification accuracy. Perfectly accurate performance usually failed after as few as five items. Initially, this result was demonstrated in experiments where the range of the stimulus set is held constant. Figure 1 demonstrates the severe nature of the capacity limitation using data from Pollack (1952) and Garner (1953). One might think that increasing the stimulus range would facilitate performance, but surprisingly the severe limit of 7 ∓ 2 persists even when the range of stimuli is increased dramatically. In fact, once adjacent stimuli are perceptually discriminable (i.e. the spacing between stimuli is well above the just noticeable difference), further increases to the range effectively have no influence on performance

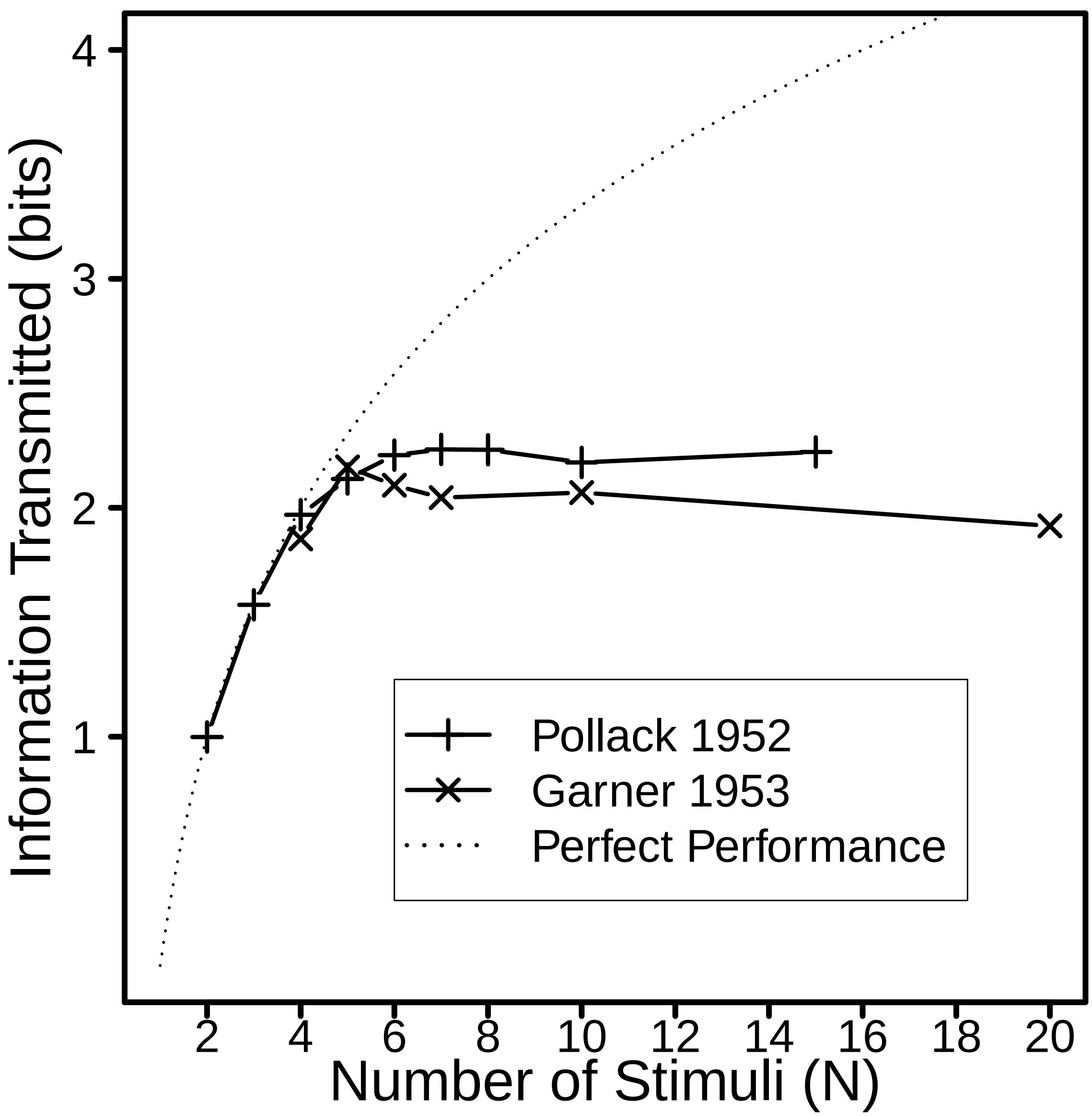

Figure 1. Performance, in terms of information transmitted, as a function of the number of stimuli used in the experiment. The data are taken from two experiments in which severe limits on performance were observed, reported by Pollack (1952) and Garner (1953). Note that k bits transmitted corresponds to the perfect identification of 2^k stimuli. Reproduced from Figure 1, Stewart, Brown and Chater (2005).

(Braida & Durlach, 1972).

A stimulus set can be characterized by the number of bits of information necessary to classify its members. Perfect classification of two stimuli requires 1 bit of information, four stimuli require 2 bits of information, and in general N stimuli require $\log_2 N$ bits of information. In a review of 25 studies using a wide variety of stimuli including frequency and intensity of tones, taste, hue of colours, and magnitude of lines and areas, Stewart, Brown and Chater (2005) found the mean information limit was 2.48 bits. Figure 1 shows how increasing the number of stimuli increases the amount of information required for perfect classification. As illustrated, classification performance falls below the information required for perfect classification as the number of stimuli increases beyond four, and asymptotes not much above two bits. Hence accuracy becomes progressively worse as set size increases above 4.

Historically, it has been found that the capacity limitation is remarkably resistant to practice. For example, Garner's (1953) participants performed around 12,000 identification trials with tones of different loudnesses[1], yet at the end of the experiment they were still limited to identifying the equivalent of just four stimuli correctly. Weber, Green, and Luce (1977) also had participants complete 12,000 trials identifying just six tones varying in loudness and found an improvement in response accuracy of just 6%. Final performance for these participants was well below ceiling, despite there being extensive practice, monetary incentives, and only six tones. Hartman's (1954) participants practiced for 8 weeks, and although they demonstrated substantial improvement, they still could only perfectly identify five stimuli - well within Miller's limit. Such results have established what has subsequently been treated as a truism about absolute identification: there is a severe limitation in human ability to identify unidimensional stimuli, and this limit is unaffected by practice. However, recent results suggest that practice is not always ineffectual, at least for some stimulus dimensions; we will return to this issue later.

Bow Effects

The bow effect is one of the most robust results in absolute identification tasks (Kent & Lamberts, 2005; Pollack, 1953; Weber et al., 1977). As shown in Figure 2, response accuracy follows a U-shape when it is plotted as a function of stimulus position. That is, in any set of to-be-identified stimuli, the smaller and larger stimuli are better identified than items in the middle of the stimulus range. This effect is particularly interesting because it is independent of the absolute size of the stimulus. That is, the ability to identify any particular stimulus depends on its relative position within the set of stimuli being identified. For example, a stimulus can go from being very accurately identified, when it is the smallest stimulus in some set, but later be identified at chance-level accuracy, when it is one of the central stimuli in a new set (Lacouture & Marley, 1995; Lacouture, 1997). The addition of new stimuli to a set causes existing stimuli in the set to be identified less accurately, but this effect varies greatly across the stimulus range. Stimuli near the edge of of the range suffer much less from the introduction of additional stimuli than stimuli near the center of the range.

[1] Although the terms loudness and intensity, and the terms pitch and frequency, have different meanings, we use them almost interchangeably for reasons of clarity.

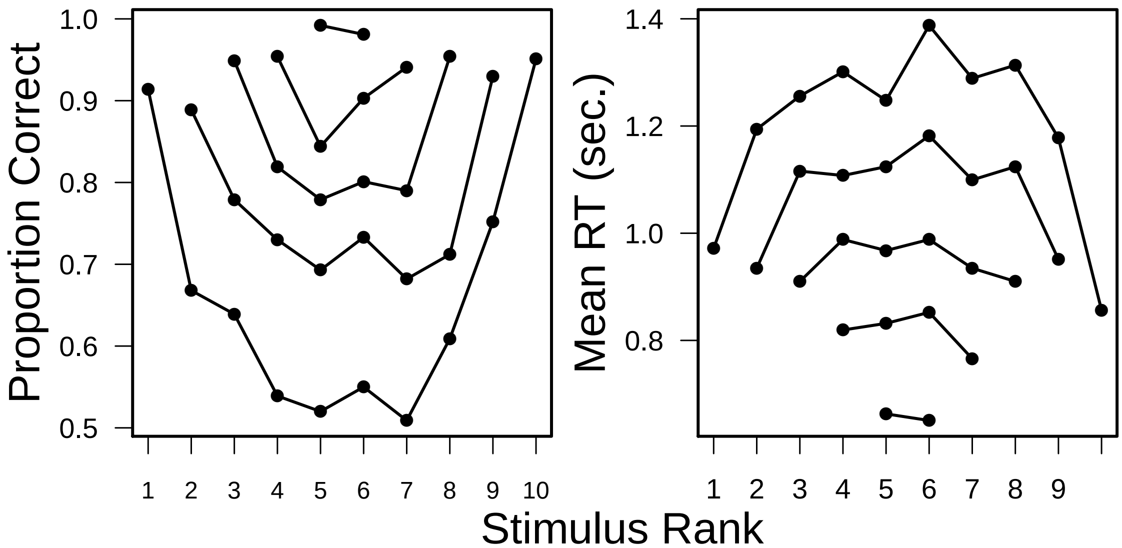

Figure 2. Performance, in terms of proportion of correct responses (left) and mean response time (right) for each stimulus. The different lines correspond to different stimulus set size conditions. The data are taken from Lacouture and Marley (1995), and demonstrate the bow effect, and the consistent performance for stimuli towards the edge of the stimulus range despite the number of stimuli in the set. Reproduced from Figures 5 and 7 of Lacouture and Marley (2005).

Sequential Effects

There has been growing recent interest in sequential effects – the influence of recent stimuli and responses on the current decision (Gilden, Thornton, & Mallon, 1995; Van Orden, Holden, & Turvey, 2003; Wagenmakers, Farrell, & Ratcliff, 2004, 2005). Absolute identification researchers were early pioneers in such analyses (Holland & Lockhead, 1968; Ward & Lockhead, 1970, 1971), showing that the occurrence of identification errors depends in a complicated way on previous stimuli and responses. These sequential effects come in two broad varieties, and turn out to be one of the most challenging aspects of absolute identification for theoretical accounts. Some researchers even believe these sequential effects are the defining feature of absolute identification (e.g., Laming, 1984; Stewart, Brown, & Chater, 2005). In what follows, we refer to the current decision trial as "trial N", which allows us to refer to the preceding trials as $N-1$, $N-2$ and so on.

Assimilation and Contrast. The response made on trial N is "assimilated" towards (i.e., tends to be similar to) the stimulus presented on trial $N-1$. For example, suppose that stimulus #5 is presented on trial N. If the stimulus on trial $N-1$ had been #1 (which is smaller), then an error on trial N is more likely to be an underestimate (e.g., #4) than an overestimate (#6) (Holland & Lockhead, 1968; Luce, Nosofsky, Green, & Smith, 1982; Stewart et al., 2005; Ward & Lockhead, 1970).

Stimuli further back in the sequence of trials have an opposite influence on responses. The response on trial N tends to "contrast" with (i.e., be dissimilar to) the stimuli on trials

$N-2$, $N-3$, and further back in the sequence. For example, incorrect responses tend to be overestimates of current stimuli when the stimuli were small on trial $N-2$, and vice versa. The contrast effect is usually smaller than the assimilation effect, and though it tends to decrease as trials further back are considered, contrast can persist until about trial $N-5$ (Holland & Lockhead, 1968; Lacouture, 1997; Ward & Lockhead, 1971).

One apparent difficulty in thinking about assimilation and contrast is that stimulus labels and responses are correlated, simply because participants typically perform above chance level accuracy. This can make it difficult to establish whether responses on the current trial are assimilated towards the previous *stimulus* or the previous *response*. Nevertheless, progress has been made on this question by using autoregressive measurement models (e.g.: De Carlo, 1992; De Carlo & Cross, 1990). These analyses have generally supported the idea that assimilation and contrast occur mostly as described above: as biases caused by previous stimuli, rather than responses. Intriguingly, recent work by Malmberg and Annis (2012) has shown that assimilation and contrast effects have close analogues in short term memory decisions.

The combination of assimilation and contrast, shown in Figure 3, poses a particularly challenging set of results for theories of absolute identification. A successful theoretical account of absolute identification must predict that responses are biased toward the previous stimulus (assimilation) but that the bias switches direction (contrast) as that stimulus recedes in memory. Several possibilities for how this occurs have been proposed, and we now discuss them.

Theories of Absolute Identification

We now give a brief overview of the different theoretical accounts of the benchmark phenomena just described. Fascination with the counterintuitive and intricate nature of absolute identification has spawned many theories and our coverage does not aim to be encyclopeadic (we recommend Stewart et al., 2005, who give an excellent and comprehensive coverage). Rather, we aim to cover prominent examples of several different conceptual classes of theories, beginning with Thurstonian models and then covering three types of theories that have yielded models still under active investigation.

Thurstonian Models

The basic Thurstonian model assumes that any given stimulus evokes a noisy representation of absolute magnitude on an internal scale. The scale is divided by decision criteria into a set of response categories. The criteria that divide the response categories are allowed to vary between trials because of imperfect memory or shifting response bias. The noisy response categories can account for information limits when the number of stimuli increase for a fixed range, as the bounds become closer and so the noisy stimulus representation and the noisy category bounds lead to more errors in responding.

Durlach and Braida (1969) extended the basic Thurstonian model to account for the invariance of performance with changes in the range of stimuli. They proposed that, because of limits on memory, only recently presented stimuli are used to define the context within which the current stimulus is presented. This additional source of variability is modeled as being proportional to the range of stimuli, and as such can produce the required invariance.

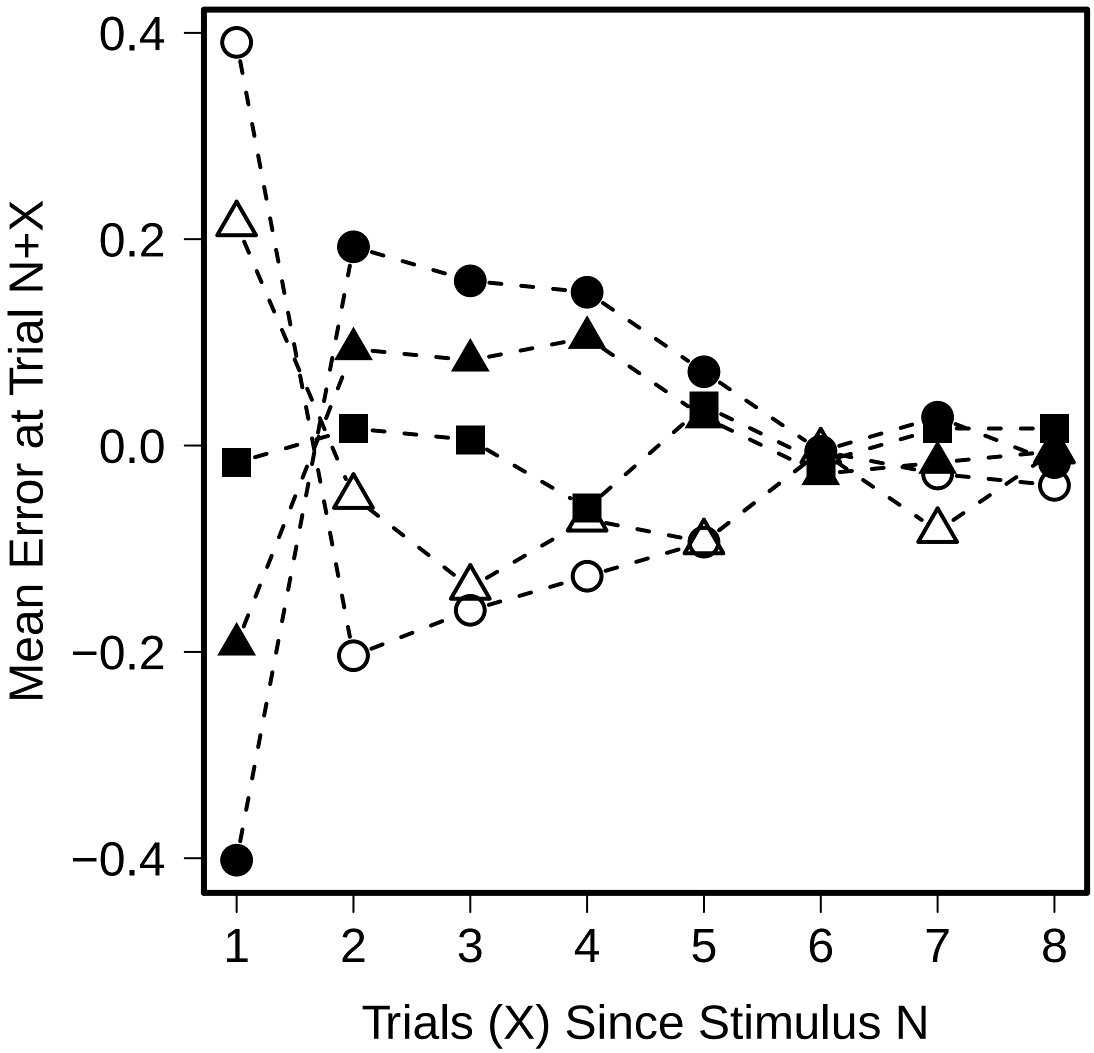

Figure 3. Assimilation (at lag X=1) and contrast effects (at X > 1) in data from Holland & Lockhead (1968). When X=1, the average error is negative for smaller stimuli (filled symbols) and positive for larger stimuli (unfilled symbols). That is, when the stimuli on the previous trial were small, stimulus magnitudes on the current trial were underestimates, while stimulus magnitudes were overestimated when the previous stimuli were large. When X> 1, the pattern reverses, and a contrast effect is observed. Reproduced from Figure 4 of Brown, Marley, Donkin and Heathcote (2008).

Basic Thurstonian models can also account for bow effects, but their account is not entirely satisfactory. The models predict greater accuracy for edge stimuli because Thurstonian models predict erroneous responses are mostly adjacent to the correct response, and edge stimuli have only one adjacent response whereas internal stimuli have two. This account is sometimes called a "response restriction" account, because improved accuracy is predicted for edge stimuli simply because they have fewer neighbouring stimuli to be confused with. Although this restricted response explanation certainly does play some role in bow effects, it fails to predict the observed gradual decrease in accuracy that is observed as stimuli become more internal to the set. Many researchers have used a d'-based measure of performance (Luce et al., 1982), in place of raw accuracy. The d' measure is, under some assumptions, insensitive to response effects such as bias and response restriction. The d' measure still reliably shows deep bow effects, and this is something that response restriction explanations of the bow effect cannot explain.

Braida et al. (1984) elaborated the basic Thurstonian model's account of bow effects to allow for greater variability for stimuli near the center of the range than the ends. They explained this by proposing that stimuli are judged against two "anchors" at the extreme ends of the stimulus range, and the distances between these anchors and the presented stimulus are counted using a noisy measurement unit. The further a stimulus is from those anchors the noisier the distance measurement becomes. This produces a gradual bow effect with increasingly frequent identification errors toward the middle of the measurement range. Luce, Green, and Weber (1976) proposed a related model, suggesting that bow effects may be due to more attention being paid to the edges of the stimulus range, thus reducing the variability of stimulus representations in those regions.

Treisman (1985) modified the basic Thurstonian model in a different way, in order to accommodate sequential effects. Treisman and Williams's (1984) criterion-setting theory proposed that the criteria that set the bounds for response categories change on a trial-by-trial basis due to two factors: a stabilizing mechanism and a tracking mechanism. The tracking mechanism moves the criteria away from the previously observed stimulus. This is based on the assumption that in the real world stimuli do not occur randomly, but that given something is perceived it is likely to reappear again sometime soon. The tracking mechanism moves the criterion away from the previous stimulus (so expanding the stimulus range for the response just given) so that when it is presented again, it is more often correctly identified. The stabilizing mechanism moves the criteria such that they are centered on the mean of the previous stimuli. The stabilizing mechanism acts to counteract the tracking mechanism, moving criteria back in a way that maintains balanced responding in the long run. Treisman assumed that the tracking mechanism is stronger than the stabilizing mechanism, but decays more quickly. In this way, Treisman's model accounts for assimilation to trial $N-1$, where the tracking mechanism dominates, and contrast away from previous trials, where the stabilizing mechanism dominates.

Exemplar Models

Exemplar models have proven very successful in accounting for categorization behavior, and this makes them promising candidates for theories of absolute identification because of the close similarity between absolute identification and categorization. Exemplar models assume absolute identification is accomplished by determining the similarity between a to-be-identified stimulus, and the memory representations for previous stimuli. Each stimulus is assumed to be represented in memory along with its associated label. The probability of response i is proportional to the similarity between the current stimulus j and all exemplars for response i. In general, exemplar models face the problem of not being able to account for the fundamental information limit of absolute identification: when the range of the stimulus set is increased but the number of stimuli remains fixed, these models predict that the memory representations should be more easily discriminable, and so performance should improve. As Braida and Durlach (1972) showed, this does not happen.

Nosofsky (1997) extended exemplar models to also make predictions about the time taken to make decisions, rather than just the decision which is made. His exemplar-based random walk model (EBRW; Nosofsky & Palmeri, 1997) assumes that the representations of stimuli in memory are normally distributed across the stimulus dimension. Upon presentation of a stimulus, the exemplars race to be retrieved from memory with a speed that is an increasing function of the similarity between the current stimulus and each exemplar. Each time an exemplar is retrieved a counter for the associated response is incremented, while the counters associated with other responses are decremented. The race continues until one counter reaches a threshold and the corresponding response is given. Nosofsky's model accounts for bow effects by assuming that stimuli towards the edge of the range have smaller variance in their perceptual representation. The EBRW model makes no attempt to account for sequential effects or the information limit phenomena. It is worth noting, however, that the EBRW model is one of the few absolute identification models to consider response time. Indeed, the EBRW makes detailed and precise predictions for response times, including full response time distributions for both correct and incorrect responses.

Petrov and Anderson (2005) proposed an exemplar model of absolute identification called ANCHOR based on the ACT-R architecture (Anderson, 1990; Anderson & Lebiere, 1998). They assume that unidimensional stimuli are encoded by a perceptual subsystem into an absolute magnitude. The magnitude is then processed by the central subsystem, comparing it with some exemplars or "anchors" stored in long-term memory. The central subsystem is thought to be dynamic and evolves from trial to trial throughout the absolute identification task. The perceptual processing that creates the magnitude and the selection of the anchor exemplar are both stochastic, with selection of the exemplar based on similarity between the current stimulus and memory representations, as well as the previously presented stimulus.

The ANCHOR model accounted for bow effects in accuracy (Petrov & Anderson, 2005) via the restricted response set explanation of bow effects – and therefore suffers the same difficulty as Thurstonian models in accounting for bow effects in d'. Though not explicitly tested, Stewart et al. (2005) suggest that the ANCHOR model would probably be able to account for information limit when range and set size were varied due to noisy parts of the model unrelated to spacing of stimuli. The ANCHOR model accounts for assimilation by making exemplars that have been recently used more likely to be retrieved, but does not attempt to account for contrast effects.

Kent and Lamberts (2005) proposed an exemplar model of absolute identification based on an adaptation of Lamberts's (2000) extended generalized context model for response times (EGCM-RT). EGCM-RT, which in turn is based on Nosofsky's (1986) general context model, assumes that information about the current stimulus is sampled repeatedly until sufficient evidence is accumulated to choose a response. Kent and Lamberts' model successfully accounts not only for the basic bow effect that is observed in accuracy but also for the inverted bow effect that is observed in response time (i.e., responses to stimuli from the center of the range are slower). It does so because stimuli towards the end of the range are relatively more isolated. This adds a different mechanism to the restricted response set explanation of bow effects, because not only do edge stimuli have fewer competing response alternatives, they also have smaller summed similarity than stimuli in the center of the stimulus range. A gradual decrease in accuracy and increase in response time is predicted because summed similarity increases gradually for stimuli that are closer to the center of the range.

Without modification, the EGCM-RT was not able to account for systematic differences in the shape of bow effects as a function of the number of stimuli to be identified (set size; see Figure 2). However, Kent and Lamberts (2005) were able to capture these differences by assuming that the amount of information sampled from a stimulus was a decreasing function of the number of pieces of information already sampled – a plausible assumption of "diminishing returns". By assuming that additional samples from a stimulus were decreasingly useful in discriminating between members of the stimulus set, an information limit was imposed on the basic EGCM-RT model. Kent and Lamberts did not attempt to account for sequential effects.

Relative Judgement Models

The previous Thurstonian and exemplar models tended to focus more heavily on explaining fundamental capacity limitations and bow effects, and paid less attention to explaining sequential effects in absolute identification data. The defining feature of relative judgment models, on the other hand, is that decisions are based on the difference between current and previous stimuli (or responses), rather than being based directly on representations of the absolute magnitudes of stimuli. This leads to a natural focus on sequential effects. Relative vs. absolute judgments is not just a theoretical distinction in absolute identification modeling, but is also important in some applied areas. For example, musicians are frequently trained in the skill of relative judgment (judging musical "intervals") but almost never in absolute judgment, because relative judgment is useful in musical performance while absolute judgment is not (and may even be maladaptive). This difference may help explain the rarity of "perfect pitch" in musicians, which we discuss below.

Holland and Lockhead (1968) proposed that responses are made by combining feedback from the previous trial and the perceived distance between the current and previous stimuli. This basic relative judgment mechanism accounts for assimilation and contrast by simply assuming that the judged difference between the current and previous stimuli is biased towards the previous stimulus and away from earlier stimuli. For example, consider the case in which a small stimulus was presented on the previous trial ($N-1$). This means that, on average, the stimuli presented on earlier trials ($N-2$, $N-3$, ...) were probably larger stimuli. The memories for these larger earlier stimuli interfere with the judgment of the distance between the previous and current stimulus in a way that causes the distance to be underestimated. Hence, a response based on this distance will assimilate to the stimulus presented on trial $N-1$. As Stewart et al. (2005) point out, this approach can explain assimilation and contrast on average, but fails to provide a general account. To recount their example, imagine that a small stimulus on trial $N-1$, say #3 is followed by a smaller stimulus, #2. Now, the interference from more previous stimuli will cause an overestimate of the distance between the stimuli, and produce contrast, whereas assimilation is usually still observed in such cases.

Laming (1984) proposed a strict version of a relative judgment model, assuming that no absolute information is used, and that only the difference between the previous and current stimulus is considered. In particular, decisions are assumed to be made in a relatively coarse manner, such that the current stimulus is judged in terms of only five categories: "much less than", "less than", "equal to", "more than", or "much more than". Such limited categorical information provides a natural account of the fundamental capacity limits, within a relative judgment framework.

Stewart et al. (2005) proposed the most current and successful relative judgment model of absolute identification. In their model, only the series of differences between each stimulus and the next is represented internally. These differences, along with feedback from the previous trial, are used to produce a response. The formula used to generate a response, R_n on trial n is: $R_n = F_{n-1} + \frac{D^C_{n,n-1}}{\lambda} + \rho Z$ Where F_{n-1} is the feedback on trial $n-1$, $D^C_{n,n-1}$ is the difference between the representations of differences on trial n and $n-1$, λ is a scaling parameter, ρ is a parameter that increases with $D^C_{n,n-1}$, and Z is a normally distributed variable with mean zero and standard deviation σ. Stewart et al. (2005) demonstrate that their instantiation of a relative judgment model is capable of producing all of the classical response-choice related benchmark phenomena.

Stewart et al.'s (2005) model accommodates the information limit and bow effect in accuracy through related mechanisms. The information limit is accounted for via the inclusion of a parameter (Z), which is assumed to represent noise in the mapping of the physical stimulus to its internal representation. Bow effects are accounted for by assuming that Z is scaled by parameter ρ, which depends on the position of the current stimulus within the range of stimuli. For example, if there are eight stimuli to be identified, the previous feedback was #7 and the current stimulus is judged to be larger, then the influence of the noise, σ, is smaller than when the stimulus is closer to the center of the stimulus range. The relative judgment model accounts for assimilation and contrast by assuming that the judgment of the distance between the previous and current stimulus is influenced by previous distance judgments, and that their impact decreases with their position in the trial sequence (i.e., the distance judgment on trial $n-1$ has the most influence, followed by that on trial $n-2$, etc.).

Restricted Capacity

Marley and colleagues have proposed a number of models that attribute poor performance in absolute identification to limited capacity in memory or attention. Marley and Cook (1984, 1986) assume that the full range of stimuli in an experiment are mapped onto an experimental context, which might be thought of as a fixed-capacity attention or memory store. When a stimulus is presented, its relative position within the set of stimuli is located within a context by reference to "anchors" located near the ends of the stimulus range. The relative position of the stimulus is then used to judge its magnitude, and subsequently assign a response label.

Capacity limitations in the model are built into the way the context is maintained. Attention to the stimulus range (context) of the experiment is assumed to be maintained by constant "rehearsal" of that particular segment of the stimulus dimension. Rehearsal is assumed to operate as a Poisson pulse process that directs activity to the relevant context, but this activation is assumed to passively decay. The power of the rehearsal process is assumed to be fixed within an individual, thus yielding a limit on performance when either more stimuli are added, or the range of stimuli is increased.

Marley and Cook's (1984,1986) model naturally predicts the bow effect in response accuracy and d' because the relative magnitude of a stimulus is judged according to the amount of rehearsal activity between the current stimulus location, and the closest anchor. The magnitude of stimuli in the center of the range will, therefore, be estimated with greater variance than edge stimuli, since the noise in the Poisson rehearsal process will have more influence on estimation (i.e., similar to Durlach & Braida, 1969).

Lacouture and Marley (1995, 2004) proposed an alternative means of explaining the information capacity limit and bow effect through their mapping model. The model assumes a basic and very simple absolute magnitude estimate as its input, and transforms it into a set of response-output strengths for each of the possible responses via an internal ("hidden") unit, in a similar way that a set of tuning curves might transform a magnitude estimate into a set of response tendencies. Since the hidden unit normalizes the input into a unit (0-1) range, a small and fixed amount of noise in the input magnitude estimate has greater influence when the number of response alternatives increases, as more stimuli are packed into the same psychological space. In Lacouture and Marley (2004), the output strengths coming from the mapping model were used to drive a leaky, competing accumulator (LCA) model (Usher & McClelland, 2001). This expanded the explanatory scope of their model to include RT, not only in terms of mean RT but also variability in RT (i.e., the full distribution of RT).

The restricted capacity models reviewed above make no attempt to account for any of the benchmark sequential effects in absolute identification. The SAMBA model (Brown, Marley, Donkin, & Heathcote, 2008) extends these earlier models, and incorporates elements of both the Poisson rehearsal and mapping processes from the previous two models, but uses a deterministic version of the LCA, Brown and Heathcote's (2005) ballistic accumulator model. The authors claimed SAMBA to be the most comprehensive theory of absolute identification yet proposed because they showed it was capable of accounting for all of the aforementioned benchmark phenomena in absolute identification not just in terms of response choices, but also in terms of the full distribution of response times.

SAMBA is an acronym for the model's three stages. The **S**elective **A**ttention aspect of Marley and Cook (1984) produces a magnitude estimate for a stimulus, which is then used as an input to the **M**apping model of Lacouture and Marley (1995). The mapping process transforms the single magnitude estimate into N response strengths, which then drive a corresponding number of **B**allistic **A**ccumulators. The ballistic accumulators are a set of evidence accumulators that accrue activation at a rate determined by the output of the mapping stage of the model. Evidence in each accumulator increases ballistically (that is, deterministically, without moment-to-moment noise), suffers from passive decay, and is inhibited by the evidence in other accumulators.

The information limit and bow effects arise out of the first two stages of the SAMBA

Table 1: A simplified summary of the ability of models within each class to account for historical and more recent phenomena. [1]Depending on the definition one allows for a "purely relative" model - see Brown et al. (2009). [2]But see Donkin et al. (2014) for a revised version of SAMBA capable of accounting for this phenomenon

Model Type	Empirical Phenomena						
	Historical Benchmarks			Current Issues			
	Capacity Limitations	Bow Effects	Sequential Effects	RT	Uneven Spacing	Inter-trial Interval	Learning
Thurstonian	Yes	Yes	Yes	No	No	No	No
Exemplar	Yes	Yes	No	Yes	No	No	No
Relative	Yes	Yes	Yes	No	Yes[1]	No	No
Capacity	Yes	Yes	Yes	Yes	Yes	No[2]	No

model. SAMBA provides an advance over earlier restricted capacity models by also accounting for assimilation and contrast. Assimilation is accounted for by assuming that the evidence in accumulators passively decays between trials, beginning from their level at the time at which a response was made. For example, this means that the accumulator that won the race on the previous trial will begin the current trial with the highest level of activation. This benefit for responses close to previous stimuli leads to assimilation toward responses made on the previous trial. Contrast is incorporated into the selective attention stage of the model. Brown et al. (2008) assumed that between trials, rehearsal activity is preferentially directed to the elements of the context corresponding to the stimulus presented on the previous trial. On subsequent trials, this additional activation around previous stimuli leads to contrast away from those stimuli. That is, since magnitude is estimated by summing the amount of activation between an anchor and the stimulus, increased activity will "push" the magnitude estimation away from the previous stimulus.

Table 1 summarizes the ability of each model class to account for existing benchmark phenomena. All of the classes of models discussed so far are basically capable of accounting for what we are calling Historical Benchmarks – capacity limitations, bow effects, and sequential effects. Interested readers should see Stewart et al. (2005) for a more detailed table outlining the various models' ability to account for historical benchmark data. For the remainder of the chapter we will focus on recent issues in absolute identification. As seen in Table 1, these issues pose a challenge for many theories, and there are a number of results that yet remain unresolved. We will not attempt to resolve these issues here, but instead hope to lay out a guide as to what should constrain model development moving forward.

Current Issues in Absolute Identification

Absolute and Relative Judgment

A unresolved tension in the modeling of absolute identification concerns the relative merits of "absolute" vs. "relative" accounts. Recently, the line between absolute and relative models has become blurred, leading to some important tasks for future research. Most immediately, exactly what defines a purely relative vs. purely absolute model needs clarification? Subsequently, the limitations of each class need to be more clearly established, especially if the current "integrated" models, which include both absolute and relative aspects, are to be justified.

Absolute and relative judgment models are defined in a similar spirit to the way the two terms are used in the judgment of musical notes, above. An "absolute" account proposes that each new stimulus is assigned a label by comparison with some stable referents stored in memory – such as the extreme ends of the stimulus range. A "relative" account proposes that each new stimulus is judged instead against the memory of only very recently-observed stimuli (perhaps just the most recent single stimulus). The earliest Thurstonian and exemplar models were purely absolute, in that each stimulus was judged only against long-term referents. Similarly, the earliest relative judgment models (Holland & Lockhead, 1968; Laming, 1984) were purely relative in that stimuli were compared only to the most recently-observed prior stimulus. These "pure" models had difficulty accounting for fundamental phenomena. The purely relative models could not easily account for bow effects in d', or capacity limitations, and the purely absolute models could not easily account for any of the sequential effects.

Purely relative and purely absolute models can be envisaged as the endpoints on a continuum. For example, a purely relative model can be made a little less relative by allowing comparisons against the last two, or three, observed stimuli. The end point of such a process is an absolute model: an exemplar model, where all previous stimuli are used for comparison. Similarly, a purely absolute model can be made a little more relative by allowing a stimulus to be judged by comparison to recent stimuli, as well as to long-term referents.

To accommodate the growing list of complicated benchmark phenomena, recent models have moved away from purely relative or purely absolute accounts. For example, the SAMBA model (Brown et al., 2008) was able to accommodate almost all benchmark phenomena with purely absolute assumptions. However, one phenomenon was only able to be modeled within SAMBA's framework by including a relative process, judgment against the most recently-observed magnitude. The phenomenon in question was the improved accuracy and response times observed for repeated stimuli, and for stimuli which are similar, but not identical, to the previous stimulus. It seems ad hoc to incorporate an otherwise unnecessary model component just to accommodate this one phenomenon, and so future work might investigate whether this phenomenon can be explained more parsimoniously. If it is possible to explain this phenomenon without recourse to a relative judgment process, this could simplify theoretical arguments, by once again having a purely absolute model that fits the data.

On the other hand, the purely relative models of Holland and Lockhead (1968) and Laming (1984) were augmented by Stewart et al. (2005). Stewart et al.'s relative judgment model includes some elements which might or might not be classed as "absolute", depending on taste. For example, in Stewart et al.'s model, the amount of variability in the internal representation of a magnitude depends on the distance between that magnitude and the end of the range. This comparison with the end of the range sounds just like an absolute judgment procedure, but it could avoid that label by instead supposing that magnitudes being represented are *response* magnitudes, rather than stimulus magnitudes. How this argument would hold up in general is unclear: for example, if non-numeric response labels were used, as is customary in categorization tasks.

Brown, Marley, Dodds, and Heathcote (2009) tried to identify the shortcomings of relative judgment accounts without recourse to particular theories or detailed and questionable assumptions about precisely which mechanism are absolute vs. relative. They noted that relative judgment theories proceed by allowing exactly one piece of absolute knowledge about stimulus magnitude: the difference in magnitude between adjacent stimuli in the decision set. This knowledge is necessary to translate observed differences in magnitude between successive stimuli into predicted differences in their response labels. For example, suppose a set of 10 lines are used, with each line differing in length by 2cm from its neighbors. If the previous trial used line #4 (which is known, after the feedback is given) and the current line is 6cm shorter, then knowledge of the 2cm spacing in the set can be used to deduce that the current line must be three responses smaller (line #1). Brown et al. reasoned that this kind of account must break down in the more general case, when the spacing between adjacent stimuli in the set is uneven. In that case, a relative difference of 6cm between the current and previous lines implies a different number of response units depending on which line was shown.

Brown et al. (2009) re-ran the classic experiment of Lockhead and Hinson (1986), in which absolute identification was performed on three loudness levels. In one condition, the loudness values were evenly spaced (2dB apart) but in the other conditions, the gap between one pair of stimuli was three times as large (6dB). As shown in Figure 4, a relative judgment account failed to accommodate data from the unevenly-spaced conditions. Brown et al. interpreted this as a failure of purely relative models. However, Stewart and Matthews (2009) explored an augmented version of the relative judgment model, endowed with knowledge of *all* the various stimulus spacings between all pairs of stimuli. Although such a model can easily accommodate the data, what is less clear is the status of that model on the absolute-vs.-relative continuum. Proponents of relative models might construe such knowledge as purely relative, as it involves only knowledge of differences between stimuli. On the other hand, proponents of absolute models might construe such knowledge as purely absolute, as it assumes a long-term memory for the structure of the entire stimulus set.

The most attractive way forward for the debate between absolute and relative accounts is probably by resolution of the exact definition of the basic terms. However, since that is likely a more difficult task than it appears, an alternative way forward is by simply fitting the existing models to data, and comparing their performance on quantitative measures of goodness-of-fit. This approach has worked well in other fields, and avoids many of the problems inherent in the search for qualitative differences between classes of models.

Learning

The limit on people's ability to learn to identify more than around seven unidimensional stimuli was firmly established early in the history of absolute identification research. However, the remarkable levels of achievement that are displayed by experts after extended practice in a broad range of domains – e.g., musicians with absolute pitch – makes this limit puzzling. The following quote, taken from Shiffrin and Nosofsky (1994) sums up this conundrum with capacity limits nicely.

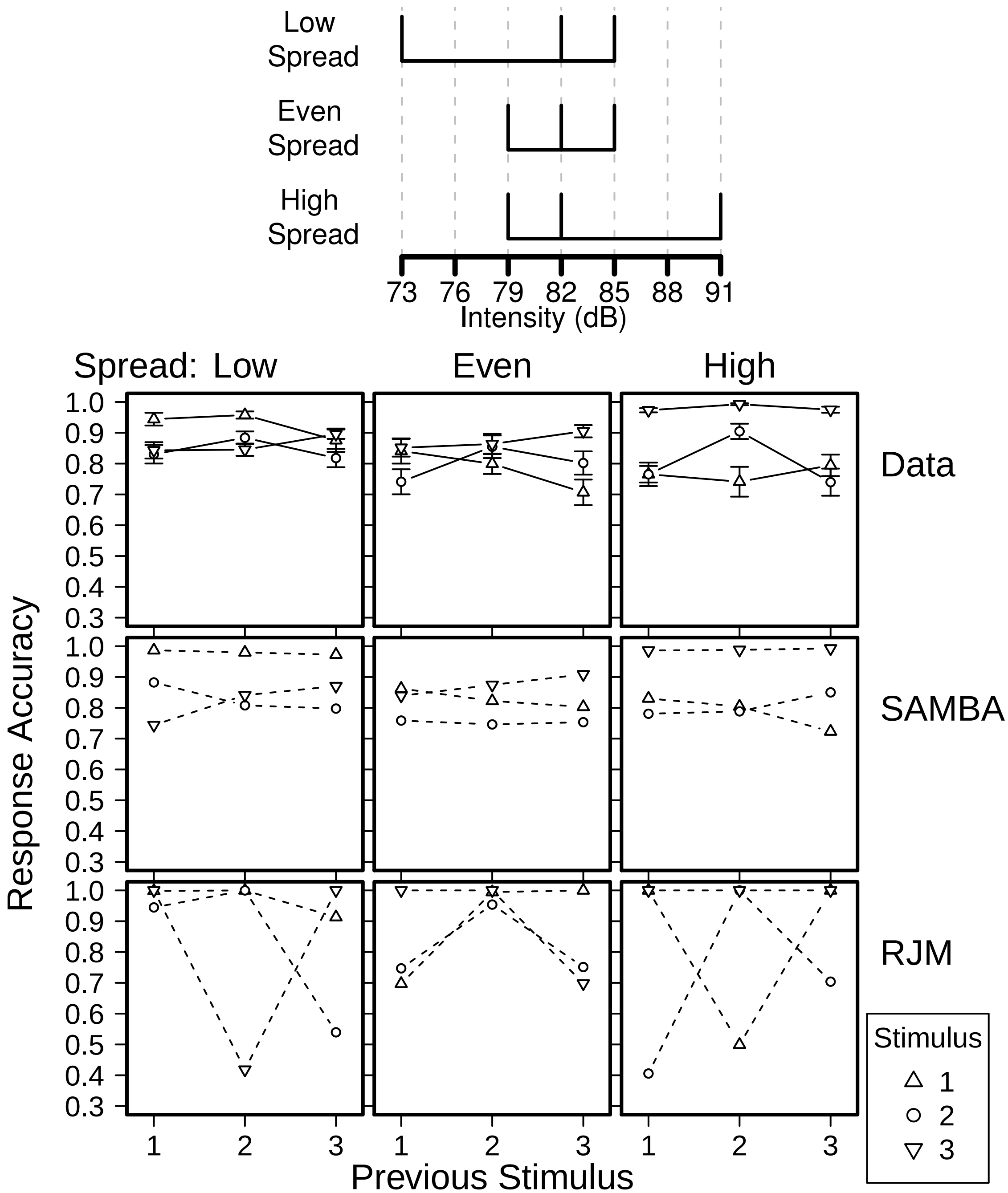

Figure 4. Top panel: Schematic illustration of the stimuli used in Brown et al.'s (2009) experiment. Bottom panel: Response accuracy (y-axis) for each stimulus (symbols) conditioned on the previous stimulus (x-axis). The upper row of panesl show data, the middle row show predictions from the SAMBA model, and the lower row of panels show predictions from the RJM. The three columns correspond to the low-spread, even-spread and high-spread conditions. Reproduced from Figures 1 and 4, Brown, Marley, Dodds and Heathcote (2009).

> "As an anecdotal example, Robert M. Nosofsky started his research career around 1980 in the acoustical laboratory of David Green and R. Duncan Luce, two researchers who happened at the time to be studying absolute identification of loudness ... As a cocky young graduate student, Nosofsky 'knew' that with a bit of practice, he could surely learn to perfectly identify a set of 12 loudnesses. After locking himself in one of Green's sound-proof booths for several weeks, and hearing tone after tone after tone, his absolute identification performance remained unchanged. He did succeed, however, in increasing substantially his need for psychotherapy."

Research in the last 10 years, however, suggests that Professor Nosofsky was just unlucky to choose the particular stimulus continuum (loudness), and that other stimulus dimensions are more amenable to learning, at least for some people. Rouder, Morey, Cowan, and Plaltz (2004) showed that at least one of their participants, the redoubtable RM, was capable of far exceeding a limit of seven items in absolute identification of line length. Indeed, all three participants in their experiments were shown to display some learning. However, after extended practice, RM (one of the authors) displayed a remarkable ability to accurately identify up to almost 20 stimuli. The other two participants, despite not performing at quite the same level, both achieved performance equivalent to the perfect identification of around 10-13 items, both above Miller's limit of 9.

Dodds, Donkin, Brown, and Heathcote (2011) replicated the key finding from Rouder et al.'s (2004) study. Dodds et al. then extended the findings in a series of seven experiments. Many more participants were observed to show substantial learning, though none quite reached the high bar for absolute identification of line lengths set by RM. Dodds et al. (2011) showed that the learning could be observed for stimulus modalities other than line lengths, including the separation between dots (which removed the confound of brightness with line length), and the degree of angular rotation of a line. As shown in Figure 5, auditory stimulus sets, particularly the tone loudnesses that repelled Nosofsky's determined assault, were shown to be much more difficult to learn. Tones of varying frequency presented an interesting case, given the popular belief in the existence of absolute or perfect pitch. Four of the six participants who practiced with tones of varying frequencies showed no substantial improvement. However, one participant showed learning on the order of that displayed by RM with line lengths, and another showed smaller, but still substantial, improvement. By the end of just ten experimental sessions, the best performer could identify about 16 items and was showing of no signs of a slowing in their learning rate, suggesting they could have achieved even better performance with further practice.

Dodds et al.'s (2011) results might be crudely described as showing that absolute identification performance can be improved for all kinds of stimulus sets with extended practice; except for stimulus sets based on tones of varying loudness. This result resolves the apparent contradiction between new and old findings on practice effects. These new results, however, raise a host of new issues. For one, there is the unresolved question of why tone loudness cannot be learned, while other stimulus types can: and further, why the ease of learning might differ between other stimulus sets. In further work, Dodds, Rae, and Brown (in press) found that those continua supporting learning may have a psychological representation of magnitude that is more complex than the simple, one-dimensional physical structure of the stimulus set (see also Dodds, Donkin, Brown, & Heathcote, 2010). For example, although lines vary on a single physical dimension – length – they may, after

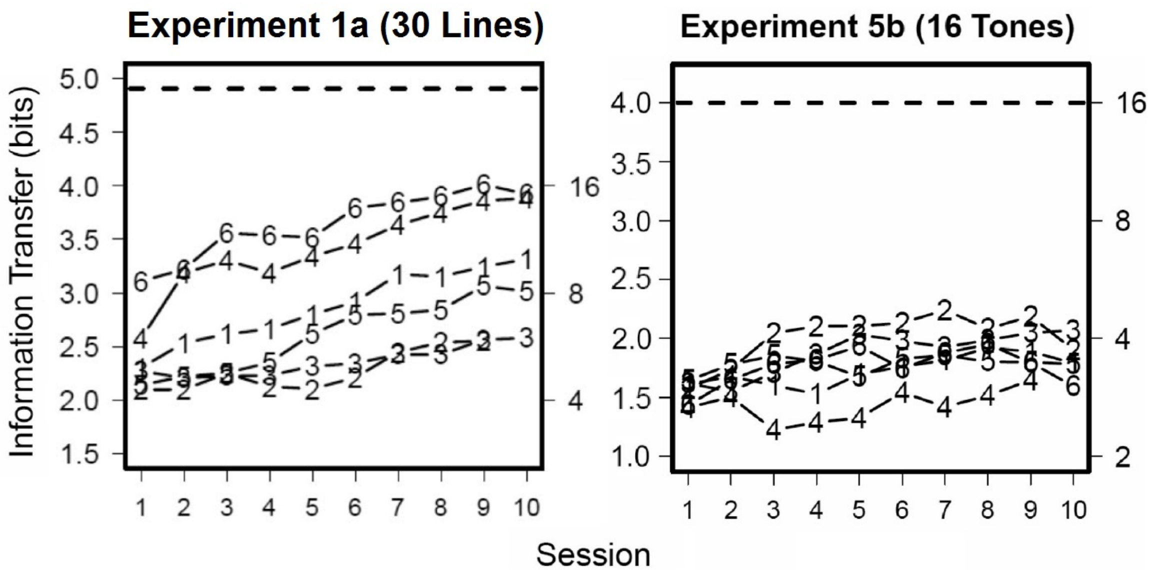

Figure 5. Information transmission as a function of sessions of practice plotted for two experiments from Dodds et al. (2011). The left panel of the figure is a replication of the unusually large amount of learning observed for lines of varying length. In the right panel, however, we see the more standard limited capacity for tones of varying loudness, that was more commonly reported in absolute identification experiments. Reproduced from Figures 1 and 5, Dodds, Donkin, Brown and Heathcote (2011).

extended practice, develop a higher-dimensional psychological representation. This greater dimensionality provides additional information for the participants to learn, and may allow for continued improvement in performance (Rouder, 2001). For lines of various length, it is not clear what these extra dimensions of information might be. However, for some other stimulus sets, more is known. For example, tones that vary in pitch form a physically one-dimensional set, but this dimension is represented psychologically as a two-dimensional helix: the well-known separation of musical notes into chroma and octaves.

Dodds et al.'s (2011) study also revealed considerable individuals differences in the ability to learn. The best predictor of such differences seems to be initial performance. That is, participants who performed well in the first few hundred trials also were also those who showed the greatest improvement with subsequent practice. The direction of this correlation can be considered surprising. One might naively expect a negative correlation, since poorer initial performance leaves greater room for improvement. An open question about the nature of learning in absolute identification regards the reason for the observed correlation. One possibility lies in a link between the psychological representation of stimulus sets and performance. Dodds et al. (in press) observed that extended practice was related to a move away from simple (one-dimensional) psychological representations towards more complex representations. Although there is a chicken-and-egg problem that requires attention here, it suggests that improved absolute identification performance occurs through reorganization of the psychological representation of the stimulus set.

An important unanswered question is how theories of absolute identification should incorporate learning. Accounting for learning is no trivial task, especially since most models were developed to account for exactly the opposite – severe and resistant capacity limitations. Dodds et al. (2011) made inroads to understanding how to account for the effects of practice by looking at what practice does to the previously described benchmark phenomena. Bow effects remain across practice with approximately the same magnitude. Similarly, the stimulus presented on trial $N-1$ maintains an assimilative effect at the beginning and end of practice. However, the contrast effect from stimuli further back in the trial sequence ($N-2$, $N-3$, etc.) is greatly reduced by practice. This reduction in contrast effects with practice can be seen as an adaptive behavior. Contrast effects represent incorrect responses, but such errors can be adaptive in situations where the to-be-identified stimuli drift slowly with time. Such drift did not occur in the experiment, which makes learning to reduce contrast effects adaptive. Incorporating learning into absolute identification models by modulating contrast mechanisms is, therefore, a good start to accounting for practice effects. However, changes in contrast are unlikely to be the sole locus of learning – e.g. in the SAMBA model we have found that completely removing the contrast mechanism leads to an improvement in performance that was only about one third of the size observed in some of Dodds et al.'s participants.

Clearly, practice effects provide a challenge to existing models claiming to aspire to a complete account of absolute identification. The set of empirical results we have summarised provide clues as to the locus of the effect of practice, as well as providing concrete targets for model fitting. Further, it seems likely that the increase in performance for some modalities may be, at least partially, driven by the development of higher-order representations of stimuli. Finally, the magnitudes of individual differences in practice effects should be in some way related to initial performance.

Absolute Identification vs. Perfect Pitch

"Perfect pitch", also known as absolute pitch, is the ability to perform almost perfectly on an absolute identification task where the stimuli are a set of musical notes that must be labelled with their chroma and octave (e.g. "C4", also known as "middle C"). Absolute pitch is considered to be rare with only 1 in 10,000 of the general population reported as having the ability (Bachem, 1955; Takeuchi & Hulse, 1993). Rates are said to rise to 1 in 1,500 amongst amateur musicians (Profita & Bidder, 1988), and up to 1 in 7 in highly accomplished musicians (Baharloo, Johnston, Service, Gitschier, & Freimer, 1998). However, those rare people who possess absolute pitch are able to accurately identify dozens of different stimuli, well beyond the limits observed in other stimulus sets. This raises questions about the difference between absolute pitch and absolute identification using other stimulus sets. Miller (1956) noted this apparent dichotomy, but was unable to explain it.

An obvious candidate explanation for the difference in performance between absolute identification tasks in general and absolute pitch tasks is the difference in the physical dimensionality of the stimuli. Live or recorded musical notes are sometimes used in absolute pitch tasks, but these stimuli are physically multi-dimensional, consisting of a fundamental frequency and a series of harmonic overtones. There are also marked variations in timbre, volume, resonance and decay characteristics across the registers of the musical range (Lockhead & Byrd, 1981; Terhardt & Seewann, 1983). Any of these attributes might be used by observers to out-perform the usual limitations of absolute identification. However, even when the stimuli are truly one-dimensional – computer-generated sine tones – some people can still exhibit absolute pitch performance (Athos et al., 2007). This may represent an end-point of the kind of learning observed by Dodds et al. (2011) and Dodds et al. (in press), where extended practice altered the psychological representation of pure tones, supporting improved identification. This explanation is further supported by the prevalence of "octave errors" in identification of pitch (where a note is mistaken for a note with the same chromatic name in a different octave).

Another possible explanation for good performance in absolute pitch tasks is that people, especially musicians, have more exposure to the stimuli generally used in absolute pitch tasks – musical notes. Previous thinking considered absolute pitch to be an innate quality possessed by very few people, and a skill unable to be learned (e.g., Revesz, 1953). However, many recent researchers agree there is a critical period, under the age of approximately eight, where children can learn absolute pitch if exposed to enough musical training, with several researchers noting a correlation between absolute pitch ability and musical training during the critical period (e.g., Levitin & Rogers, 2005). However, exposure to musical training during this critical period is not sufficient to develop absolute pitch, as most people who are musically trained during this period do not develop the ability (Baharloo et al., 1998). There are conflicting views as to whether learning is possible beyond the critical period. Some researchers maintain absolute pitch cannot be learned to any level of fluency after the critical period (e.g., Zatorre, 2003), while others argue that with enough practice, learning is possible (e.g., Lundin, 1963). Dodds et al.'s (2011) results support the latter view.

Evidence that absolute pitch is a learned ability, rather than innate, comes from the non-uniform performance across the range of notes. People with absolute pitch most accurately identify those notes to which they have had more exposure. Middle C and other white-key notes on the piano are often most accurately identified (Lundin, 1963; Takeuchi & Hulse, 1993), perhaps because these are the first notes learned in standard piano training (Athos et al., 2007; Miyazaki, 2004). Similarly, violin players are able to more accurately identify the open A string, as they are very familiar with this tone (Brammer, 1951), and in general musicians more accurately identify pitch on their own instrument (Takeuchi & Hulse, 1993).

Prior work on absolute pitch has often been confounded by the presence of feedback, and by the skill known as "relative pitch". Relative pitch is the ability to identify differences between notes, rather than individual notes in isolation (Miyazaki, 1995). The difference between two notes is known as a "musical interval", and musicians are trained to identify these intervals accurately. Relative pitch plays an important role in musicianship, whereas absolute pitch does not. Confounds arise when absolute pitch is tested in the usual way, by a succession of notes with feedback on the correct answer provided after each. Participants with good relative pitch skills can use the judged intervals, combined with knowledge of the previous note from feedback, to perform accurately – even without possessing absolute pitch skills. Some experimenters have attempted to separate the contribution of relative and absolute pitch by using interference tasks between trials (e.g. Zatorre & Beckett, 1989), or separating successive tones by long time intervals (Baharloo et al., 1998). A simpler technique to control for relative pitch is to remove feedback information, which makes knowledge of the intervals unhelpful. However, feedback must be given at some points, both for motivation and to stop poorly performing participants from becoming wildly inaccurate. A compromise involves alternate blocks with feedback and without feedback (Ward & Lockhead, 1970).

Speakers of tonal languages (e.g., Mandarin, Cantonese, Thai) have been reported to have an advantage in pitch perception due to their use of pitch in conveying the meaning of words (e.g., Deutsch, Henthorn, Marvin, & Xu, 2005). This hypothesis incorporates the ideas that extended practice with identifying pitch (when speaking a tonal language) and exposure to this task during childhood (when learning language) might both support the development of absolute pitch. Various studies have provided indirect support for the assumed link, between tonal language and absolute pitch, such as Pfordresher and Brown (2009), who found that native tonal language speakers were better able to imitate and perceptually discriminate musical pitch. Deutsch et al. (2009) investigated the pitch ranges of female speech in two relatively isolated villages in China, and found that the pitch range of speech is heavily influenced by an absolute pitch template that is developed through long term exposure. However, other studies have reported that tonal language speakers have no advantage in absolute pitch (e.g., Bent, Bradlow, & Wright, 2006; Schellenberg & Trehub, 2007).

In a recent experiment from our lab (not yet published), 60 native Chinese language speakers identified pure sine tones in a standard absolute identification paradigm. All 60 participants were university students at Sichuan University, Chengdu, China, 30 of whom were enrolled in the Music Faculty, and 30 of whom were enrolled in the Faculty of Administration. Although the music students outperformed the administration students when feedback was provided ($p \leq .01$), there was no difference between the two groups when no feedback was given ($p = .57$). Further, both groups' mean performances for feedback and non-feedback blocks were well below Miller's (1956) upper limit of 9 stimuli, suggesting that native Chinese [tonal] language speakers do not have an advantage in identifying uni-dimensional tones.

Response Times

Response times used to be rarely collected in absolute identification experiments, and they have not been nearly as often subjected to the detailed analysis given to response choices. Previous research has identified several effects in mean response times that are analogous to effects in choice. These include the effects already outlined: bow effects, in which responses to extreme stimuli are faster than those to middle stimuli (Lacouture, 1997; Lacouture & Marley, 1995, 2004), and capacity limitations, where RTs slow down as the number of stimuli increases (Kent & Lamberts, 2005) (see Figure 2 for an example of these). Sequential effects on mean response times have also been observed due to response repetition (Petrov & Anderson, 2005) and assimilation (Lacouture, 1997).

It could be argued that response times can be safely ignored because they fail to provide additional constraint on theories of absolute identification beyond the constraints provided by response choice data. This argument rests on the assumption that effects on response time are always just the inverse of effects on response choices (e.g., increased accuracy is always associated with decreased response time). Given this assumption, any theory that can successfully account for the effect of some manipulation on accuracy will need only invert those predictions to also account for response times. Such inversion can be accomplished by feeding the output of the model into evidence accumulators (like those used in the SAMBA model of Brown et al., 2008). Large outputs drive the accumulators more quickly, resulting in the required faster and more accurate responses.

However, Donkin, Brown, Heathcote, and Marley (2009) showed that an inverse relationship between accuracy and response time is not always observed. In a re-analysis of data collected by Lacouture (1997), manipulations of the distance between stimuli sometimes had large effects on choice probability but no effects on response time. For example, in one condition, participants were presented with a stimulus set composed of five smaller lines (#1-5) and five longer lines (#6-10), with a much larger gap between stimuli #5 and #6 than between other adjacent stimuli. As might be expected, participants were more accurate with stimuli that lay on the bounds of the gap between the two subsets of stimuli; they never confused stimuli #5 and #6. Despite this benefit in accuracy, however, responses to the stimuli on the edge of the center gap were just as slow as when standard equally spaced stimuli are used.

This, and similar, results are important because they cannot be accounted for by any model that simply relies on a strict inverse relationship between accuracy and response time. More generally, such a dissociation between accuracy and response time runs counter to the claim that response time is not worthy of investigation in absolute identification.

The result also provides a strong challenge to models of absolute identification. Donkin et al. (2009) showed that the SAMBA model provides a natural account of the dissociation between accuracy and speed due to stimulus spacing (see Figure 6). The credit, however, must go to the mapping model from Lacouture and Marley (1995), for the dissociation arises out of the mapping stage of SAMBA. Recall that the mapping model takes a single magnitude estimate, z, and transforms it into a set of K response strengths, R_i, $i = 1..K$. This transformation is accomplished via the formula $R_i = (2Y_i - 1)z - Y_i^2 + 1$, where Y_i measures the average magnitude estimate for stimuli classified with response i. The transformation is parameter-free, and depends only upon the relative position of each of the K stimuli in the stimulus space (as recorded by Y_n). When these Y_ns represent stimuli with a large, central gap, the mapping produces response strengths that yield high accuracy for stimuli adjacent to the gap, but response times that remain slow.

What made the fits of the SAMBA model to these data particularly compelling was that they were essentially parameter free. The spacing experiments from Lacouture (1997) were part of a larger experiment with other manipulations, including a control condition. In this baseline condition, a standard set of 10 equally-spaced stimuli were used, and Brown et al. (2008) had already shown that SAMBA could fit these control data. Donkin et al. (2009) fixed SAMBA's parameters to those estimated in the controlled condition, and then allowed only the representation of the stimulus spacings (the Y_is) to change across conditions, with the further constraint that those Y_is veridically represented the stimulus spacings used in the experiment. With this setup, SAMBA gave accurate quantitative predictions for the data from all of the spacing conditions.

Inter-trial Interval and Sequential Effects

Different models of absolute identification make different predictions about the influence of manipulating the time between decisions. As noted by Matthews and Stewart

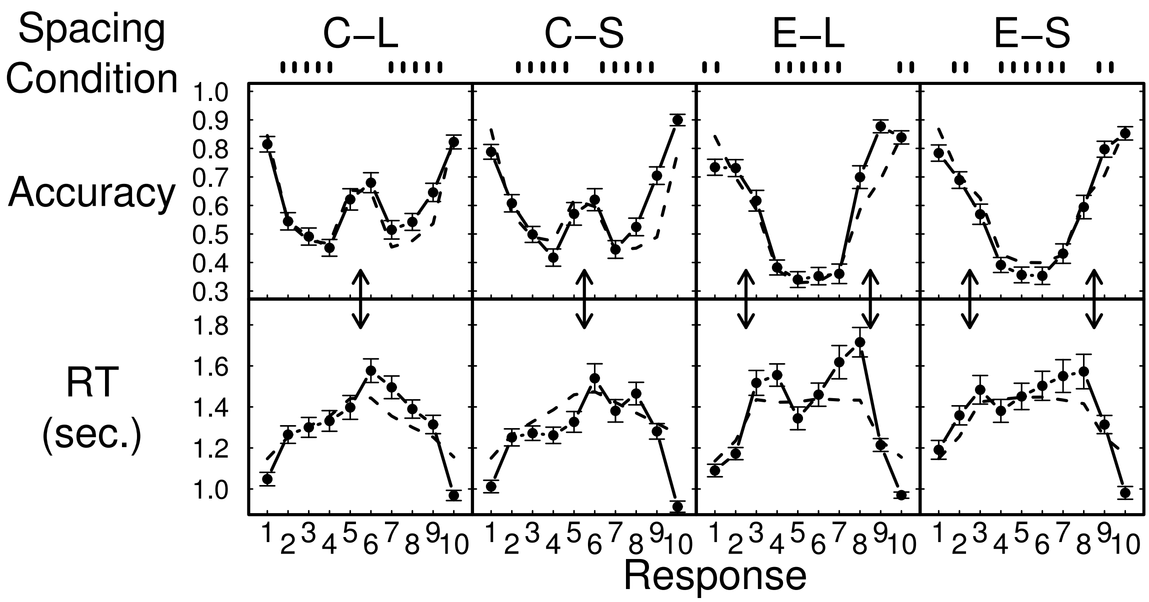

Figure 6. Dashes above the top panels provide a schematic representation of the stimuli used in Lacoutures (1997) experiment. The second row shows response accuracy and the third row shows mean RT for correct responses, both as functions of response. Data are shown as points with standard error bars that are joined by solid lines, and SAMBAs fits are shown with dotted lines, and the arrows between panels show where large gaps separate adjacent stimuli. These gaps lead to different effects on response times than accuracy. Reproduced from Figure 2, Donkin, Brown, Heathcote and Marley (2009).

(2009), there have been three basic accounts for sequential effects: criterion shifts, memory confusions, and selective attention. These three explanations correspond (approximately) to the Thurstonian, relative judgment, and restricted capacity model classes outlined earlier.

The Thurstonian models assume that sequential effects arise out of shifts in criteria; a fast-moving tracking process which produces assimilation, and a relatively slow-moving stabilizing mechanism that produces contrast. Both of these shifting processes decay in strength over time. If the time between trials is increased, then the criteria will have more time to return to their original positions, resulting in a reduction in both assimilation and contrast. The relative judgment models also predict a reduction in sequential effects with increased inter-trial interval due to decay, but the decay is located in a different process. Sequential effects in relative judgment models are caused by the interference from memories of the previous stimuli. Assuming that memory decays with time, then more time between stimulus presentations would reduce the confusion coming from memory.

In agreement with the other classes of models, the selective-attention-based SAMBA model also predicts that assimilation will decrease with inter-trial interval, but it disagrees in predicting that contrast will actually increase. The reduction in assimilation again comes about because of decay, in this case in the activity of response accumulators in the decision stage of the model. More time will allow the starting activity in accumulators to return to baseline levels, thus reducing assimilation to the previous response. The increase in contrast is a by-product of SAMBA's existing explanation for how contrast occurs in general, the re-allocation of attention to the representation of the previous stimulus. Recall that the stimulus context is maintained via a rehearsal process in which attention is randomly allocated across the range of stimuli used in the current experiment. Contrast occurs because the location in which the previous stimulus was presented receives increased attention. So, while rehearsal activity decays across the rest of the stimulus space, the previous stimulus location is given preferential rehearsal. If this selective attention is allowed to continue for a longer time, due to an increased inter-trial interval, the contrast effect will increase in magnitude.

Matthews and Stewart (2009) manipulated the time between trials to be either 0 or 5 seconds in an otherwise standard absolute identification experiment. They found that relative to the small inter-trial interval condition, assimilation was reduced, and contrast increased when the time between trials was longer. This pattern of results appears to align with the predictions from SAMBA and to contradict those from Thurstonian and relative judgment models. However, a more sophisticated analysis of the data that used regression equations to separate out the influence of previous stimuli and previous responses revealed a more challenging pattern of results for theories of absolute identification.

Responses from participants were fit with a regression equation in which both stimuli and responses were included as predictors:

$$R_n = r_0 + \alpha_0 S_n + \alpha_1 S_{n-1} + \beta_1 R_{n-1} + \alpha_2 S_{n-2} + \beta_2 R_{n-2} + e_n$$

Where R_n is the response made on the nth trial, S_n is the stimulus presented on the nth trial, α and β are regression coefficients, and e_n is a normally distributed error term. Based on this equation, Matthews and Stewart (2009) found that assimilation to responses does not decrease as inter-trial interval increases (i.e., that β_1 remains positive and of approximately the same magnitude in both short and long inter-trial interval conditions), but rather that contrast to the previous stimuli increases as inter-trial interval increases (i.e., both α_1 and α_2 become more negative).

The observed increase in contrast with inter-trial interval is predicted by the selective attention explanation for contrast, but is inconsistent with the Thurstonian and relative judgement accounts for sequential effects. Recall, however, that SAMBA predicted a decrease in assimilation to the previous response with inter-trial interval. Though the raw data appear to agree with SAMBA, the regression coefficients reported by Matthews and Stewart (2009) suggest that the observed lack of an assimilation effect in the long inter-trial interval condition is due to an increase in contrast to the previous stimulus, rather than a reduction in assimilation to the previous response. This suggests that either the mechanism by which SAMBA produces assimilation is inappropriate and needs adjustment, or that the decay process is extremely slow.

Conclusion

Absolute identification is an apparently simple task: take a set of lines that vary only in length, or tones that vary only in loudness; label them in a straightforward way; and ask an observer to recall the correct labels, when shown stimuli one at a time. Despite this apparent simplicity, there is a wealth of complicated and robust patterns that have been identified in absolute identification data over the past 60 years.

Early theoretical accounts of absolute identification were inspired by some of the most fundamental data patterns – for example, the relative judgment models were inspired by the ubiquitous sequential effects, and the limited capacity models by the observed performance limits. Since they were inspired by just one of the many fundamental phenomena, these theories all had problems in accounting for the full range of data. More recent theories are more complex, and share structure with more than one type of earlier theory. These new theories aim to provide a comprehensive account of the full range of absolute identification phenomena (Petrov & Anderson, 2005; Stewart et al., 2005; Brown et al., 2008).

This leaves some clear challenges for future empirical and theoretical research in absolute identification:

- What exactly constitutes a "purely relative" and "purely absolute" theoretical account? Can the current models be modified to perfectly fit either category? If so, can a pure model explain the full range of data?
- What mechanism underpins the newly-observed learning effects in absolute identification? Are tones of varying loudness the only stimulus set which is impossible to learn?
- To what extent are the sequential effects in absolute identification malleable? They are apparently altered substantially by practice, and by the timing of the experimental procedure. These findings may provide new constraints on existing theories.

Further into the future, a challenge for absolute identification research will be to build links with other paradigms. Some paradigms clearly are closely related to absolute identification, such as magnitude estimation, magnitude production, and categorization. It should be relatively simple for empirical work to explore the links between these fields, but a more difficult challenge will be to integrate and unify theoretical accounts in the different fields, where possible.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Hillsdale.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Athos, E., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Frimer, N., & J., G. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences*, *104* (37), 14795–14800.

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America*, *27*, 1180-1185.

Baharloo, S., Johnston, P., Service, S., Gitschier, J., & Freimer, N. (1998). Absolute pitch: An approach for identification of genetic and nongentic components. *American Journal of Human Genetics*, *62*, 224–231.

Bent, T., Bradlow, A., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 97–103.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception: Ii. resolution in one-interval paradigms. *Journal of Acoustical Society of America*, *51*, 483–502.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception: Xiii. perceptual anchor model of context-coding. *Journal of the Acoustical Society of America*, *76*, 722–731.

Brammer, L. (1951). Sensory cues in pitch judgment. *Journal of Experimental Psychology*, *41* (5), 336–340.

Brown, S. D., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*, 117–128.

Brown, S. D., Marley, A., Dodds, P., & Heathcote, A. (2009). Purely relative models cannot provide a general account of absolute identification. *Psychonomic Bulletin & Review*, *16*, 583–593.

Brown, S. D., Marley, A., Donkin, C., & Heathcote, A. J. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115* (2), 396–425.

De Carlo, L. (1992). Intertrial interval and sequential effects in magnitude scaling. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1080–1088.

De Carlo, L., & Cross, D. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, *119*, 375–396.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2005). Absolute pitch among american and chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119* (2), 719–722.

Deutsch, D., Le, J., Dooley, K., Henthorn, T., Shen, J., & Head, B. (2009). Absolute pitch and tone language: Two new studies. *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music*.

Dodds, P., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Multidimensional scaling methods for absolute identification data. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society.* Portland, OR.: Cognitive Science Society.

Dodds, P., Donkin, C., Brown, S. D., & Heathcote, A. (2011). Increasing capacity: Practice effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *37*, 477–492.

Dodds, P., Rae, B., & Brown, S. D. (in press). Perhaps unidimensional is not unidimensional. *Cognitive Science*.

Donkin, C., Brown, S. D., Heathcote, A., & Marley, A. A. J. (2009). Dissociating speed and accuracy in absolute identification: The effect of unequal stimulus spacing. *Psychological Research*, *73*, 308–316.

Donkin, C., Chan, V., & Tran, S. (2014). The effect of blocking inter-trial interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology*.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. i. preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372–383.

Ellis, A. J. (1876). On the sensitiveness of the ear to pitch and change of pitch in music. *Journal of the Royal Musical Association*, *3*, 1–32.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, *46*, 373–380.

Gilden, D. L. (2001). Cognitive emissions of $1/f$ noise. *Psychological Review*, *108*, 33–56.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). $1/f$ noise in human cognition. *Science*, *267*, 1837–1839.

Hartman, E. (1954). The influence of practice and pitch-distance between tones on the absolute identification of pitch. *American Journal of Psychology*, *67*, 1–14.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, *3*, 409–414.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*, 628–666.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*, 316–332.

Kent, C., & Lamberts, L. (2005). An exemplar account of the bow and set size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 289–305.

Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Research*, *60*, 121–133.

Lacouture, Y., & Marley, A. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology*, *39*, 383–395.

Lacouture, Y., & Marley, A. A. J. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics*, *66*, 1206–1226.

Lamberts, K. (2000). Information-accumulation theory of speeded classification. *Psychological Review*, *107*, 226–260.

Laming, D. (1984). The relativity of ”absolute” judgements. *British Journal of Mathematical and Statistical Psychology*, *37*, 152–183.

Levitin, D., & Rogers, S. (2005). Absolute pitch: Perception, coding, and controversies. *TRENDS in Cognitive Sciences*, *9*, 26–33.

Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *Journal of the Acoustical Society of America*, *70*(2), 387–389.

Lockhead, G. R., & Hinson, J. M. (1986). Range and sequence effects in judgment. *Perception & Psychophysics*, *40*, 53–61.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics*, *20*, 49–54.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics*, *32*, 397–408.

Lundin, R. (1963). Can perfect pitch be learned? *Music Educators' Journal*, *49*(5), 49–51.

Malmberg, K., & Annis, J. (2012). On the relationship between memory and perception: sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General*, *141*(2), 233–259.

Marley, A. A. J., & Cook, V. T. (1984). A fixed researsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology*, *37*, 136–151.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychological judgments applied to magnitude estimation. *Journal of Mathematical Psychology*, *30*, 339–390.

Matthews, W., & Stewart, N. (2009). The effect of inter-stimulus intervla on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology*, *62*, 2014–2029.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review*, *63*, 81–97.

Miyazaki, K. (1995). Perception of relative pitch with different references: Some absolute-pitch listeners can't tell musical interval names. *Perception & Psychophysics*, *57*(7), 962–970.

Miyazaki, K. (2004). How well do we understand absolute pitch? *Acoustical Science and Technology*, *25*(6), 426–432.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. A. J. Marley (Ed.), *Choice, decision and measurement* (pp. 347–365). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar–based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Petrov, A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*(383–416).

Pfordresher, P., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*, 1385-1398.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America*, *24*, 745–749.

Pollack, I. (1953). The information of elementary auditory displays: Ii. *Journal of the Acoustical Society of America*, *25*, 765–769.

Profita, J., & Bidder, T. (1988). Perfect pitch. *American Journal of Medical Genetics*, *29*, 763-771.

Revesz, G. (1953). *Introduction to the psychology of music.* London: Longmans Green.

Rouder, J. N. (2001). Absolute identification with simple and complex stimuli. *Psychological Science*, *12*, 318–322.

Rouder, J. N., Morey, R. D., Cowan, N., & Plaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review*, *11*, 938–944.

Schellenberg, E., & Trehub, S. (2007). s there an asian advantage for pitch memory? *Music Perception*, *25*, 241–252.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, *101*, 357–361.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881–911.

Stewart, N., & Matthews, W. (2009). Relative judgment and knowledge of the category structure. *Psychonomic Bulletin & Review*, *16*, 594–599.

Takeuchi, A., & Hulse, S. (1993). Absoute pitch. *Psychological Bulletin*, *113*(2), 345–361.

Terhardt, E., & Seewann, M. (1983). Aural key identification and its relationship to absolute pitch. *Music Perception*, *1*, 63–83.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties for a model of absolute judgment. *Journal of Mathematical Psychology*, *29*, 175–230.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68–111.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self–organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*, 331–350.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review*, *11*.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self–organized criticality. *Journal of Experimental Psychology: General*, *134*.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effect and memory in category judgment. *Journal of Experimental Psychology*, *84*, 27–34.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, *9*, 73–78.

Weber, D. L., Green, D. M., & Luce, R. D. (1977). Effect of practice and distribution of auditory signals on absolute identification. *Perception & Psychophysics*, *22*, 223–231.

Zatorre, R. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692–695.

Zatorre, R., & Beckett, C. (1989). Multiple coding strategies in the retention of musical tones by possessors of absolute pitch. *Memory & Cognition*, *17*(5), 582–589.

Zotov, V., Jones, M., & Mewhort, D. J. K. (2011). Contrast and assimilation in categorization and exemplar production. *Attention, Perception, & Psychophysics*, *73*, 621–639.

Textboxes

Textbox 1: Glossary

Absolute Models. These theories assume that identification decisions rely entirely on long-lasting, internal representations. Simple forms of exemplar models, Thurstonian models, and restricted capacity models fall into this category of model.

ACT-R. A production-system based cognitive architecture explanation of human cognitive and perceptual systems that has been mapped to brain areas.

Assimilation. Responses are more likely to be closer to the response (and stimulus) made on the previous trial.

Bow Effect. Improved performance for stimuli towards the edge of the stimulus range, so-called because performance plotted as a function of stimulus position has a bow-shaped curve.

Capacity. Processing resources used to identify stimuli. Performance in absolute identification tasks suggest that humans are severely limited in their capacity when identifying unidimensional stimuli.

Contrast. Responses are more likely to be further away from the stimulus (and response) from 2-5 trials previous.

d'. A signal-detection theory based measure of how well two stimuli can be discriminated.

Stimulus Dimensionality. The number of physical continua on which a set stimuli vary. Absolute identification research focuses on the classification of stimuli varying on just one physical dimension that is assumed to be mapped onto a single psychological dimension.

Exemplar Model. Theory assuming that stimuli are represented and stored as individual memory traces (i.e., exemplars). The similarity between the stimulus presented on a given trial and those stored in memory is critical for decision-making.

Evidence Accumulators. Hypothetical processing units used to account for choice and response time in decision-making tasks. Each potential response is given its own accumulator wherein evidence is collected and races. The first accumulator to reach a threshold level of evidence gives the chosen response, and the time taken to reach threshold gives decision time.

Modality. The perceptual system to which a stimulus is presented. For example, tones varying in frequency are from the auditory modality.

Relative Models. These theories assume that the identification of unidimensional stimuli is based on the judgement of the difference between the current stimulus and temporary representations of recently presented stimuli (and/or the corresponding responses).

Sequential Effects. The influence of previous stimuli and responses on behavior on the current trial.

Textbox 2: Surprising limitations

Perhaps the most surprising aspect of absolute identification is how difficult it is. The task itself sounds almost trivially simple – remember the labels that correspond to, say, ten stimuli. Despite weeks of training, perfect performance remains out of reach; remarkable, when the same person might be able to understand quantum mechanics, identify thousands of species of birds, and almost certainly have a lexicon of many tens of thousands of words.

The key difference between absolute identification and the many tasks which paradoxically are much easier, is that the stimuli being identified vary on only a single dimension. Ten lines that vary only in their length are very difficult to discriminate – even when the spacing between the lines is large enough such that any pair of them can be told apart trivially easily. Of course, if these lines join together to produce shapes that differ on multiple dimensions, accurate identification becomes much easier. As such, the ability to store and retrieve information in long-term memory, as well as our ability to categorize our environment into such a rich structure must depend heavily on the multidimensional nature of our world. Absolute identification is an important paradigm because it represents one of the simplest and clearest cases in which the capacity limits on human information processing can be studied.

Textbox 3: Sequential effects are everywhere

Previously encountered stimuli, and the responses they elicit, have a strong influence on the decision made on any given trial in an absolute identification task. Responses on trial N appear to be more like responses on the immediately preceding trial, and less like the stimuli presented on trials further back in the sequence. Though perhaps most comprehensively studied within the field of absolute identification, sequential effects also appear in many other decision-making tasks. For example, perceptual contrast has been observed in categorization tasks (Jones, Love, & Maddox, 2006; Zotov et al., 2011), assimilation has been observed in recognition memory tasks (though it is argued to be of a different nature than that in absolute identification, Malmberg & Annis, 2012), and the sequence of trials has a systematic influence on performance in simple two-alternative choice tasks (e.g., Gilden, 2001; Jones, Curran, Mozer, & Wilder, 2013).

Absolute identification provides an important window into understanding sequential effects, as the task itself is so simple. There exist a number of potential models for how previous stimuli and responses influence behavior within the realm of absolute identification. In contrast, most models of categorization, recognition memory, and even simple decision-making fail to account for sequential effects, despite their well-documented existence.