

Habituation Assessment in Infancy

Hoben Thomas and Rick O. Gilmore
The Pennsylvania State University

Infant-control habituation methodology, although serving the research community well, has never been carefully analyzed. A main use is to equate infants in their level of habituation prior to experimental manipulations in a posthabituation phase. When studied analytically and with simulation, it is found to have serious difficulties. It inadvertently recruits infants with large variations in performance while discriminating against those with less variable performance. For nonhabituating infants, its Type I error rates can approach 1. A model-based nonlinear regression framework is proposed, which, because of large individual differences in infants, takes as the unit of analysis the individual infant. It is shown to be more powerful and efficient than existing procedures and can offer practical and theoretical benefits.

Usually, when an infant is repeatedly presented with the same visual stimulus, there is a gradual weakening of attention to that stimulus. This process is called *habituation*, and it continues to be extensively used to study perception, memory, and cognition in infants. For a sense of the breadth of topics and participants, consider these recent examples: auditory–visual recognition memory in special populations of infants (Benasich, 2002), consonant–vowel rule learning in nonhuman primates (Hauser, Weiss, & Marcus, 2002), visual categorization of objects (Arterberry & Bornstein, 2002), infant’s encoding of other’s pointing as indicating actions toward objects (Woodward & Guajardo, 2002), the effects of maternal depression on visual preferences for mother’s versus others’ faces (Hernandez-Reif, Field, Diego, & Largie, 2002), the perception of different patterns of tactile stimulation in pre- versus full-term infants (Fearon, Hains, Muir, & Kisilevsky, 2002), and the

discrimination of odors (Goubet et al., 2002). The focus in this article is on the habituation to visual stimuli, but the theory and methods developed should be relevant to applications in other domains as well. Consequently, the issues considered should be of broad interest.

In most cases of visual habituation, a single stimulus is presented for relatively short periods in a discrete sequence of trials. Attention is often indexed by the amount of time an infant looks at a stimulus before looking away, and the duration of fixation typically increases when a sufficiently novel stimulus replaces the familiar one. This was the initial observation of Fantz (1964), in a study that is generally regarded as marking the onset of infant habituation research.

The infant habituation literature has grown very large (for reviews, see Bornstein, 1985, 1998; Clifton & Nelson, 1976; Cohen, 1976, 2001; Colombo, 1993, 1997; Colombo & Mitchell, 1990; Kaplan, Werner, & Rudy, 1990; Tighe & Leaton, 1976), and habituation has, for a variety of reasons, become the dominant methodology in many studies of infants’ psychological processes. However, except for a polynomial proposal (Ashmead & Davis, 1996) and Dannemiller’s (1984) exponential model, there has been little development of formal theory concerning an issue of central importance: the form of the habituation function for individual infants. Furthermore, the same general procedures for assessing habituation, set down in the early 1970s, are largely followed today. Although there are some variations, the commonly used procedure or criterion, which Dannemiller called the 50%

Hoben Thomas and Rick O. Gilmore, Department of Psychology, The Pennsylvania State University.

This work was supported by National Science Foundation Grant SBR-9617275, by National Institute of Child Health and Human Development Grant HD41476-01A1, by Philipps-Universität Marburg, Germany, and by The Pennsylvania State University. We thank Sandra Smeets, Heather Rettke, and Michael P. Dahlin for their assistance.

Correspondence concerning this article should be addressed to Hoben Thomas, Department of Psychology, The Pennsylvania State University, 513 Moore Building, University Park, PA 16802-3105. E-mail: hxt@psu.edu

decrement criterion, involves taking the infant's first two (or perhaps three) habituation trials and averaging them. Then, when a two- or three-trial window of successive responses has an average that falls below half the initial average, the habituation criterion has been achieved.

Colombo and Mitchell (1990) noted that visual habituation methods have not been subject to rigorous examination but commented that it is necessary to examine them "if we are to understand individual differences in infant cognitive processing" (p. 184). Similarly, Fagen and Ohr (1990) observed that although investigators have assumed there are individual differences in infants' responses, most "have not examined either the reasons for the obtained differences or their reliability, stability, and/or predictiveness" (p. 184). Consequently, there are two main goals of this article: One is to study existing infant habituation techniques, and the second is to provide a uniform model-based framework for characterizing the habituation process at the level of the individual infant. It is important to demonstrate why such a focus is important and useful even if one's interest is only in posthabituation processes.

Because most proponents of habituation methods have focused mainly on an infant's responses in the posthabituation phase, it has been implicitly assumed that what takes place during habituation is largely irrelevant for understanding posthabituation processes. Such beliefs are misplaced, as may be illustrated by considering the power of a simple within-subject posthabituation experiment in which the paired t test would be commonly used to gauge the significance of the response recovery to novel stimuli following habituation. As is demonstrated later in the present article, although all infants may satisfy common habituation criteria, not all infants have necessarily habituated because some infants may achieve criterion simply by chance (cf. Bogartz, 1965). This fact has important practical implications for posthabituation analysis, as we now demonstrate. Let $1 - \lambda$ be the proportion of infants who have habituated, with λ being the proportion of infants who have not habituated. Let η with $\eta \geq 0$ denote the novelty minus familiarity difference for habituating infants, with $\eta = 0$ for the nonhabituating infants. Now, suppose there are 30 infants, with $\lambda = 0$ and $\eta = 4$ yielding power of, say, $p_{t \text{ test}}$. Next suppose $\lambda = .3$ —a realistic proportion based on real data presented later. Figure 1 indicates that n must be 60, doubling the sample size, in order to maintain the same power $p_{t \text{ test}}$ as with $\lambda =$

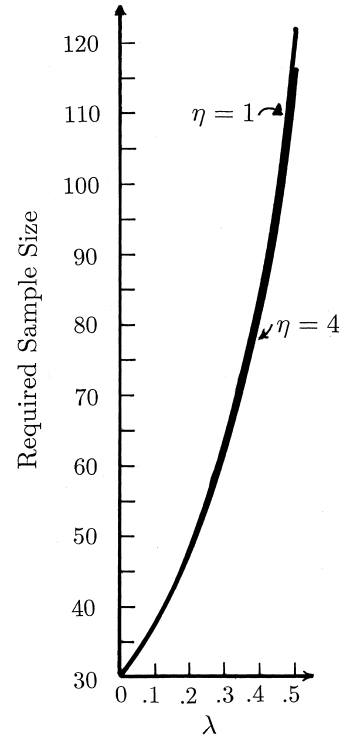


Figure 1. For a λ proportion of nonhabituaors, curves give sample size n required to obtain the same power as when $n = 30$, $\lambda = 0$, for mean differences $\eta = 1$ and 4.

0 and $n = 30$. More generally, Figure 1 shows that the sample size required to maintain the same power $p_{t \text{ test}}$ increases roughly exponentially with increases in λ while being largely independent of the mean difference η .

This result suggests that if even moderate numbers of nonresponsive infants are present in one's sample, sufficient power to detect a novelty response may be absent. Thus, a central question facing researchers is determining which infants do or do not habituate and thereby estimating λ . This can be determined within the framework of a model-based procedure, but not under existing procedures. Estimating the proportion λ is discussed later, and Appendix A (see particularly the last summary paragraph) provides details of the power calculations.

A main reason researchers use habituation methods is to equate in some unspecified way infants' familiarity with a stimulus prior to presenting them with a novel stimulus to discriminate. Unfortunately, precisely what is meant by *equating* has not been made clear. Nonetheless, if existing habituation methods achieve this goal in some satisfactory way, then how the habituation process reaches this state might seem

irrelevant. We argue, to the contrary, that current procedures almost certainly do not equate infants as assumed and that a model-based approach to quantifying the habituation behavior of individual infants is necessary to reasonably equate infants in their levels of habituation. One way to view the model-based approach is that it formalizes what have remained a series of informal ad hoc assumptions about the psychological processes that underlie an infant's habituation behavior. However, more important, the consequences of taking this perspective provide a number of advantages. The power problem is one example.

Although related in spirit to the work of some others (e.g., Ashmead & Davis, 1996; Dannemiller, 1984), the approach taken here is unlike current procedures. It permits the investigator to estimate whether an infant's attention is, in fact, declining, or merely fluctuating randomly. Furthermore, the model-based approach can clarify what it would mean to equate infants and form the basis of more effective techniques for achieving this goal.

The article is organized as follows. First, a brief history of habituation is given. Next, a general model framework is described, and general conditions that seem plausible for any habituation function are specified. This framework provides the background for investigating the infant-controlled 50% decrement habituation criterion discussed next. There are two central issues: (a) whether an infant has habituated or not and the associated decision criteria for deciding the matter and (b) the conditions required for equating infants in their levels of habituation. Then, a general habituation model framework is proposed. Next, the model-based techniques are evaluated in real and simulated data. Then, two examples are provided to illustrate why a model-based approach is relevant even for investigators uninterested in the habituation process. Then, the issue of equating within the model framework is considered, and brief consideration of online implementation is given. A general discussion follows.

History of Habituation Theory and Methodology

Our treatment is necessarily brief (see, e.g., Bornstein, 1985; Colombo, 1993; Colombo & Mitchell, 1990). The original basis for the measurement of declines in infant attention was Sokolov's (1963) observations on the reduction in the autonomic "orienting response" following repeated stimulation. Fantz

(1964) adapted these ideas to the study of infants' patterns of visual attention and presented infants with a series of paired visual displays, each for a fixed time period. One of the displays remained the same across the presentations, whereas the other changed. In measuring the amount of time infants looked at the two displays, Fantz observed that infants looked less often toward the unchanging display, suggesting that a form of stimulus encoding or learning was taking place.

Sokolov's (1963) *comparator theory* was then, and is today, the most widely accepted theoretical account of this sort of habituation behavior. This theory posits that orienting response activity is related to a comparison between an internal representation of a stimulus and the stimulus currently being presented. When the difference is large, as is the case for novel stimuli, the orienting response is also large. With repeated presentations of the same stimulus, the strength of the internal representation grows, and, thereby, the difference between it and the actual stimulus declines. In the context of infant visual habituation, comparator theory suggests that infants form internal representations of stimuli that, over repeated presentations, differ less and less from the actual stimulus. At any given point, the magnitude of the difference between internal representations and external stimuli is roughly proportional to the level of attention as indexed by the length of accumulated fixation time.

In contrast, Groves and Thompson (1970) argued that the observed orienting responses might be composed of two separate processes. Still, the power and simplicity of the comparator theory, coupled with what seem like plausible links to the underlying physiology, have made it the dominant theoretical account. What is surprising, however, is that despite the strong quantitative flavor of comparator theory, there has been relatively little effort made to extend current formulations in quantitative ways.

The Origins of Infant-Control Methodology

Fantz (1964) used what is now called a *fixed trial* procedure. A single stimulus or pair of stimuli were presented for a fixed number of presentations and, within each presentation, for a fixed duration. However, it was soon recognized that infants may vary both in their initial levels of attention and in the rate at which their attention wanes. Consequently, an individual infant may not habituate to a stimulus following the fixed number of repetitions.

Frances Degen Horowitz and her colleagues were

largely responsible for the procedures implemented today. Horowitz, Paden, Bhana, and Self (1972) introduced the notion of an *infant-control procedure* that made the habituation methodology dependent on the infant's own responses. The 50% decrement procedure, or a slight variation of it, was probably first reported in the published literature, although not so named, in Laub and Bhana (1974), an article that borrowed from Laub (1972). However, the general criterion was used in Horowitz's University of Kansas laboratory prior to the appearance of the 1972 article (F. D. Horowitz, personal communication, December 2002).

Under this procedure, a trial is defined as starting when the infant looks at a stimulus and ends when the infant looks away for some fixed predetermined minimum, usually 1 to 2 s. As noted earlier, in what is now often called the 50% decrement criterion, one computes the average response on the first two or three trials and labels this average the baseline. When a subsequent windowed average of the same number of trials falls below 50% of the baseline, the 50% habituation criterion has been achieved. Bornstein (1985) noted the similarity between the 50% decrement and a procedure proposed by McCall and Kagan (1970).¹

This general procedure probably became popular because it is simple, appeared to resolve problems with the fixed trials procedure, and was easily implemented in real time. Indeed, a recent PsycINFO search indicated that about 75% of infancy habituation publications in the last 5 years have used some variant of the 50% habituation criterion. It has also received strong endorsements. Bornstein (1985) said that it enables one "to equate infants for eventual level of habituation" (p. 257). Colombo (1993) stated that the procedure "theoretically equates infants both within and between ages in terms of how well the stimulus is processed before a subsequent test for recovery" (p. 39). Unfortunately, the conceptual basis for these claims has never been made entirely clear.

Modeling Framework

The majority of investigators who use habituation appear to recognize implicitly the need to view an infant's observed response as representing some joint function of the infant's true function (whatever it might be) and error. With only rare exception (e.g., Ashmead & Davis, 1996; Bornstein, 1985) has this idea been made explicit. In this section, the overall structure of a model of habituation is described, and

three conditions that appear to be plausibly required for all habituation functions are discussed. Subsequently, it is shown that the 50% decrement and similar procedures do not reliably distinguish between habituating and nonhabituating infants, do not equate infants as previously assumed, and have other flaws as well.

Let $Y_i(t)$ denote a random response variable measuring looking time for infant i on trial $t = 1, 2, \dots, T_i$. T_i is the last habituation trial for infant i . Observed values of $Y_i(t)$, denoted $y_i(t)$, represent realized looking time values of infant i 's trial t response.² Define

$$Y_i(t) = h_i(t) + E_i(t), \quad t = 1, 2, \dots, T_i \quad (1)$$

where $h_i(t)$ is the infant's unknown habituation function and $E_i(t)$ is a mean zero symmetrical error random variable. Note that the expectation of $Y_i(t)$ is $E[Y_i(t)] = h_i(t)$, which may be viewed as a conditional expectation dependent on t and i . The conceptual idea of expectation and its formal definition are crucial in all that follows. Indeed, our use of the term *expectation* always has this more formal property in mind. Intuitively, *expectation* means average outcome in the long run (e.g., Clarke, 1975). Thus, the focus is not on infant i 's observed responses, $y_i(t)$, but rather on infant i 's expected responses on trial t , however that infant's habituation function, $h_i(t)$, may be defined.

This emphasis on expectation appears to reflect well-established assumptions about the habituation process in addition to having numerous advantages in

¹ Let $Y_i(t)$ denote a random attention response variable for infant i on trial t . The McCall and Kagan (1970, p. 94) criterion is based on

$$\frac{[Y_i(4) + Y_i(5)] - [Y_i(1) + Y_i(2)]}{[Y_i(1) + Y_i(2)]} = \frac{Y_i(4) + Y_i(5)}{Y_i(1) + Y_i(2)} - 1.$$

Whether this ratio was positive or negative, and by how much, determined the infant's habituation status for McCall and Kagan. Suppose the ratio is negative, in which case it is equivalent, with $a = 0$, to

$$\frac{[Y_i(4) + Y_i(5)]/2}{[Y_i(1) + Y_i(2)]/2} < 1 + a,$$

where a measures the magnitude of the negativity. With $a = -1/2$ the inequality is then equivalent to $[Y_i(4) + Y_i(5)]/2 < [Y_i(1) + Y_i(2)]/4$, the 50% decrement criterion.

² Capital letters, for example D_i , T_i , $Y_i(t)$, denote random variables that are always invisible, whereas observed realizations of random variables are denoted in lowercase, d_i , t_i , $y_i(t)$, a distinction important for many expressions.

its own right. Most investigators would probably agree that an infant's expected response on any habituation trial, not the error-perturbed observed response, should be the focus of interest, and yet this has only rarely been done (e.g., Ashmead & Davis, 1996). The change in thrust in Equation 1 toward expected responses forces one to consider the process that underlies the observed response and turns attention away from the individual response observed on any given trial. The fact is that without a model of the habituation process, $h_i(t)$, one does not know whether an observed response is a good portrait of an individual infant's typical behavior on trial t or whether it is a wildly perturbed response that is far from typical. Finally, as is discussed shortly, a model-based perspective allows one to define precisely what it might mean to equate infants in terms of levels of habituation.

Equation 1 may be viewed as a regression model in which $Y_i(t)$ is regressed against t . Suitable candidates for the habituation function $h_i(t)$ will be continuous functions of $t > 0$, with the recognition that observations are typically made only on integer values with $t = 1, 2, \dots, T_i$. Of course, the shape of $h_i(t)$ is an important and perhaps contentious issue. One way of narrowing the family of possible habituation functions is to impose certain conditions on them. These conditions are not required for the general results that follow, but they are sensible properties for $h_i(t)$ to satisfy in most settings.

Condition 1: Ceiling

The weakest condition required is that $h_i(t)$ is bounded from above, or in other words has a "ceiling" value. The practice of averaging the first two or three trials in order to determine the baseline level of attention would seem to imply this condition. The condition also makes explicit the notion that there is some maximum time an infant will inspect a stimulus.

Condition 2: Bounded Monotonicity

This condition requires that $h_i(t)$ be decreasing with increasing t , at least for some $t > t_0$. Assuming that $h_i(t)$ is differentiable, Condition 2 is that $h'_i(t) < 0$, where $h'_i(t)$ denotes the first derivative of $h_i(t)$ with respect to t . Requiring $h'_i(t) < 0$ for all t guarantees that $h_i(t)$ slopes downward with increasing t . Also, $h_i(t)$ must be bounded from below by some nonnegative value as indicated in Condition 3. Some investigators sympathetic to dual-process theory (e.g., Colombo & Mitchell, 1990, Figure 8.1; Kaplan & Werner, 1986) have regarded habituation functions as

being nonmonotonic, with an early peak before a gradual decline reflecting the joint influence of both habituation and sensitization. This issue is considered briefly later. For the time being, note that even these dual-process accounts are compatible with the assumption that there is some value of t , say, $t > t_0$, beyond which $h_i(t)$ is monotonically decreasing.

Condition 3: Soft Landing

This condition ensures, first, that $h_i(t)$ has an asymptotic value or "floor" that is zero or greater and, second, that $h_i(t)$ does not "crash" but guarantees a "soft landing" to whatever the asymptotic value might be. The notion of a nonnegative floor value for $h_i(t)$ seems plausible given that there is likely to be some minimum period of inspection of a stimulus even by an infant who is quite familiar with a display. The soft-landing condition translates into a positive second derivative condition on $h_i(t)$, that is, $h''_i(t) > 0$, for all t greater than some t_0 . It is not required that $h''_i(t) > 0$ for all t —although a candidate $h_i(t)$ might—only for those t after some t_0 . It has been proposed that the habituation function for some infants might require a "sudden drop" steplike function whereas others require a gradually declining function (e.g., Cohen & Gelber, 1975; Cohen & Menten, 1981). Both types of functions can be incorporated within one model as is illustrated below.

These conditions have important implications. The conditions are not satisfied by polynomial models, such as the second-order polynomial approach proposed by Ashmead and Davis (1996). Nor are the conditions fully satisfied by exponential functions used in earlier work to fit functions to group habituation data (e.g., Cohen & Menten, 1981; Dannemiller, 1984). Later, a function that does meet all three conditions is proposed. Next, however, current infant-control procedures for determining whether habituation has actually occurred are discussed. It is shown that these procedures have weaknesses not widely appreciated by most investigators.

The 50% Criterion Method in the Habituation Context

Investigators have apparently made two assumptions about the 50% criterion method: first, that it would detect meaningful declines in infant attention

when they occurred and, second, that infants would be equated in some way once the criterion was achieved. Both of these assumptions are questioned here.

Years ago, Bogartz (1965) showed that in the case of a learning context, individuals could achieve a learning criterion even if no learning took place using a trials-to-criterion procedure. The relevance of this observation for infant-control procedures using the 50% decrement criterion has not been given the attention that it deserves. Investigators have assumed that the 50% decrement method would detect habituation when it appears; more strongly, it apparently has been assumed that “with infant-control procedures . . . every infant tested actually habituates” (Colombo, 1993, p. 42).

However, suppose the stimulus does not influence habituation. Given the acknowledged fussiness of infants and the difficulty of assessing them, it seems plausible to assume that some infants will not be induced to habituate. The consequences of this possibility seem to have gone largely undiscussed in the literature. Here, the concern is not with infants who failed to achieve habituation or whose data were discarded for other reasons such as equipment failure; these cases are routinely reported. The concern is that some infants who reach criterion may do so, in the spirit of Bogartz (1965), solely by chance.

To evaluate the sensitivity of infant-control procedures to this potential problem, we must specify a nonhabituation model. A natural one is a model in which the infant is regarded as showing no change, only variability about some constant value. In a regression context, this is a horizontal line of zero slope. Following the development of Equation 1, the model is denoted $h_{i(1)}^*(t) = \alpha_i$, and the regression model is

$$Y_i(t) = h_{i(1)}^*(t) + E_i(t), = \alpha_i + E_i(t), \quad (2)$$

where α_i is infant i 's constant value at $t = 1, 2, \dots, T_i$. With this model it can be asked how well the 50% decrement criterion performs when there is no habituation occurring. As noted earlier, assessing the proportion of nonhabituating infants, λ , is important for ensuring that tests in the posthabituation phase of a study have sufficient power. This model is evaluated in two ways: (a) in a simulation study reported now and (b) pitted against alternative models as reported in the next section.

In the simulation study, each virtual “infant” received a possible 20 habituation trials. For a given α_i , a normal mean zero, standard deviation σ_{Ei} independent error random deviate was added to each trial

Table 1
Simulation Tests of the 50% Criterion Under the No-Habituation Model of Equation 2

| α_i | σ_{Ei} | <i>Mdn</i> | \bar{y} | s_y | $P_{50\%}$ |
|------------|---------------|------------|-----------|-------|------------|
| 5 | 4 | 7 | 8.28 | 4.31 | .80 |
| 5 | 8 | 6 | 7.48 | 4.01 | .92 |
| 5 | 12 | 6 | 7.08 | 3.71 | .95 |
| 5 | 16 | 6 | 6.93 | 3.64 | .97 |
| 10 | 4 | 9 | 10.29 | 4.89 | .44 |
| 10 | 8 | 7 | 8.23 | 4.29 | .81 |
| 10 | 12 | 6 | 7.66 | 4.08 | .89 |
| 10 | 16 | 6 | 7.39 | 3.94 | .92 |
| 15 | 4 | 11 | 11.40 | 5.05 | .11 |
| 15 | 8 | 8 | 9.36 | 4.70 | .66 |
| 15 | 12 | 7 | 8.30 | 4.37 | .81 |
| 15 | 16 | 6 | 7.83 | 4.18 | .87 |
| 20 | 4 | 11 | 11.79 | 4.70 | .01 |
| 20 | 8 | 9 | 10.30 | 4.87 | .44 |
| 20 | 12 | 8 | 8.96 | 4.53 | .71 |
| 20 | 16 | 7 | 8.32 | 4.36 | .81 |

outcome, yielding $Y_i(t)$ as in Equation 2. Interest focused on when the 50% decrement was achieved and on whether it was achieved before Trial 20. In all simulations involving the 50% decrement procedure, when the average of a rolling 2-trial window dropped below 50% of the “infant’s” first 2 trials, habituation was said to be achieved. The 2-trial criterion window started with trial $t = 3$. Each of 16 conditions was explored. Each condition was defined by a combination of $\alpha_i = 5, 10, 15, 20$ and $\sigma_{Ei} = 4, 8, 12, 16$. These values were chosen because they appear to cover the range of typical infant response patterns in data from our own research. The results, based on 1,000 simulations for each condition, are given in Table 1,³ which specifies the parameters and the corresponding proportions that achieved criterion, denoted by $p_{50\%}$.

First, focus on the third column of Table 1, which shows the median trial, *Mdn*, on which habituation occurred, typically around 6 to 9 or so. These values are in the middle of the range of trials to criterion that Bornstein (1985, p. 264) reported. In addition, \bar{y} in the fourth column gives the corresponding mean number of trials, and s_y gives the standard deviations of trials to criterion. In sum, the simulated values are all within plausible ranges for actual habituation data.

To evaluate whether the 50% decrement procedure

³ Several simulation studies are reported and for the conditions of each, all have 1,000 runs. In all cases the errors

rejects values derived from a constant habituation function, focus on the first two columns and the last column of Table 1. For example, regardless of the error standard deviation, σ_{Ei} for all $\alpha_i = 5$ among the nonhabituating virtual infants, the proportion that were declared to have habituated was from .80 to .97. The remaining proportions failed to reach criterion in 20 trials. The proportions are typically very high for other values of α_i as well. However, note that the proportions decrease as α_i increases, and when α_i is large, and σ_{Ei} is small, the fewest infants reach criterion. This is a consequence of another difficulty discussed below.

Table 1 reveals that the 50% decrement method does not protect the investigator from declaring nonhabituating responders as habituating infants, particularly when the mean response levels are low and the error variances are high. Accordingly, it appears that the proportion of nonhabituating infants in any given study using conventional procedures could be potentially very large.

The Floor Effect

A second critical weakness in current procedures is susceptibility to the floor effect, which can be understood most easily by considering Figure 2. Figure 2b illustrates a hypothetical habituation function. The values for $y_i(t)$ displayed in Figure 2 were chosen largely for illustration but fall within the values observed for actual infants. Assume for the moment that there is no error in the infant's habituation response, so that $y_i(t) = h_i(t)$. Consider the habituation criterion and trial on which habituation would be reached, and refer to Figure 2b; $y_i(1) = 16$ and $y_i(2) = 13.63$, and so because the two-trial baseline denoted c_i , with $c_i = (16 + 13.63)/2$, the 50% decrement two-trial window criterion would be $c_i/2 = (16 + 13.63)/4 = 7.41$, which is shown in Figure 2b. This criterion would be achieved by trial $t = 4$ because $y_i(3) = 8.44$, $y_i(4) =$

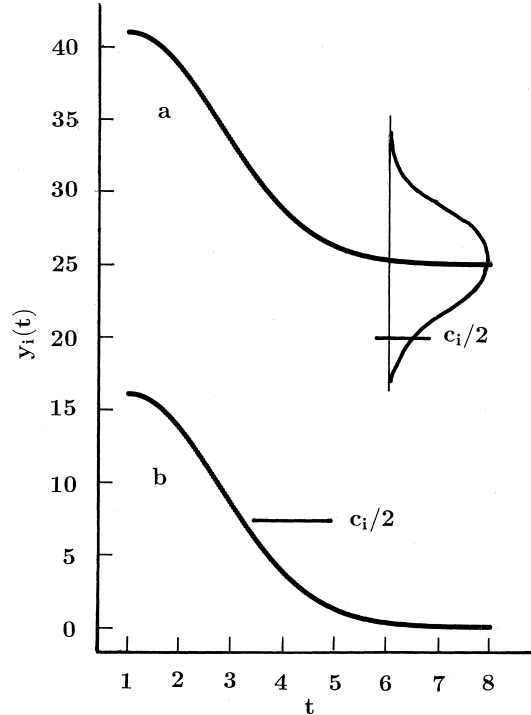


Figure 2. Two habituation functions differing by 25, illustrating the 50% decrement floor effect. $y_i(t)$ represents observed values of a random response variable measuring looking time for infant i on trial t ; c is the habituation baseline; t is trial. a: $y(t)$ in b plus 25; $c/2 = 19.91$. b: $t = 1$ to 8, $y(t)$: 16, 13.63, 8.44, 3.79, 1.24, 0.29, 0.05, 0.01; $c/2 = 7.41$.

3.79, $y_i(5) = 1.24$, so $[y_i(3) + y_i(4)]/2 = 6.12 < c_i/2 = 7.41$.

Now consider another infant, with the same shape of habituation function but simply elevated so the function's floor (i.e., the infant's minimum residual attention value) is not zero as it is in Figure 2b but rather is 25 as in Figure 2a. In this figure $y_i(1) = 41$ and $y_i(2) = 38.63$, so the new 50% decrement criterion $c_i/2$ is 19.91, which is shown in Figure 2a. This criterion could never be achieved by an error-free infant because this value is outside the range of this infant's habituation function, being well below the infant's minimum or floor value of 25. Notice, too, how much greater is the difference between the function's maximum and $c_i/2$ in Figure 2a than is the corresponding distance in Figure 2b. This is a consequence of the 50% decrement criterion's multiplicative effect.

Imagine next a real infant with the underlying habituation function of Figure 2a. This infant will show stochastic variation about his or her true habituation

were normally and independently distributed for each virtual participant and for each trial of each participant. Because the outcomes of each run may be regarded as a Bernoulli trial, the maximum possible standard error associated with any reported simulation proportion is $\sqrt{[(1/2) \times (1/2)]/1,000} = .016$; to the extent the proportions diverge from 1/2, the standard errors are much smaller. Also, each outcome of each run was required to be nonnegative because habituation outcomes are nonnegative; any negative values were set to zero.

function and, depending on the variability of the infant's response, the infant could well achieve the 50% decrement habituation criterion simply as a consequence of this random variation's occasionally plunging the infant's response time on a trial or two below the criterion. This is illustrated in Figure 2a by a normal distribution centered at the $t = 6$ function value and tail extending below the criterion $c_i/2$ value.

This peculiarity of the infant-control habituation criterion, called here the *floor effect*, must be regarded as a true case of pathology in a behavioral decision criterion: Infants are more likely to reach the habituation criterion under conditions of high error variability than when error variability is low. Where in behavioral research—and particularly in infancy research—would one ever value high error variability and deprecate low error variability? And yet, from the perspective of obtaining samples of infants who achieve habituation, this is exactly what one might hope for, although doubtlessly unwittingly so.

The results of a simulation study reported in Table 2 make this point clearly. The table shows the estimated proportion of times, $p_{50\%}$, the 50% decrement criterion was satisfied for the habituation function displayed in Figure 2a for five different σ_{Ei} values. The median trial to criterion is also reported. When $\sigma_{Ei} = 1$, the criterion is never satisfied, but the proportion of virtual infants satisfying the criterion grows systematically as the infant's error standard deviation grows. Another way of thinking about the problem is to say that for any given habituation function with its floor at zero, as in Figure 2b, as the function floor is gradually raised, the probabilities of an infant's satisfying the 50% decrement criterion monotonically decrease. These peculiarities can be summarized as follows.

Consider any habituation function for which the 50% decrement is defined. If this function is simply elevated, there is a point where the 50% decrement will be below the floor of the elevated function. Furthermore, the probabilities of achiev-

ing the habituation criterion monotonically decrease as the function is elevated above a floor of zero.⁴

Returning to Table 1, note that the decreasing probability of achieving the habituation criterion in the nonhabituating case as the value of α_i increases is simply another consequence of the fact that achieving a 50% decrement in the context of a habituation function with a higher floor is less probable than achieving a 50% decrement from a lower floor.

One apparent consequence of the floor effect is that samples of participants achieving habituation will likely be systematically biased toward including those infants with lower floors, and those infants that achieve the habituation criterion are likely to be biased toward including infants with large error variances rather than small.

Further, the present analysis provides a theoretical explanation for a phenomenon long known in the habituation literature—*spontaneous regression*. Suppose an infant has displayed performance that, under the 50% decrement criterion, would be defined as having achieved habituation. Suppose further that this infant is in a subsequent control condition and thus

⁴ Assume $g_i(t) \in (a, b)$ and for some $t = t_b$, $g_i(t_b) = b$. Define $h_i(t) = g_i(t) + \delta$, $\delta \geq 0$, so $h_i(t) \in (a + \delta, b + \delta)$. Then, for $w > 1$, a $(1/w)100\%$ decrement criterion ($w = 2$ for 50%) can be satisfied only when $h_i(t_b)/w = (b + \delta)/w > a + \delta$. This last inequality is equivalent to

$$b > \delta(w - 1) + wa.$$

This inequality may already fail. If it does not, there certainly exists a $\delta > 0$ for a fixed b and fixed $w > 1$ so that the inequality must fail. Thus for any $g_i(t)$ there is a function $g_i(t) + \delta$ for which the $(1/w)100\%$ decrement criterion will fail to hold. This was to be shown. The same result is given when $Y_i(t) = h_i(t) + E_i(t)$ replaces $h_i(t)$ and expectations are considered.

To show the monotonic decrease in probability of achieving the 50% decrement as the function is elevated, for $t \geq 4$, $P\{[Y_i(t) + Y_i(t - 1)]/2 < c_i/2\} = P\{E_i(t) + E_i(t - 1) < c_i - g_i(t) - g_i(t - 1) - 2\delta\}$ after simple manipulation and substitution. Because c_i , $g_i(t)$, $g_i(t - 1)$, δ are all positive non-random quantities, the right side of the argument of the probability inequality becomes smaller as δ grows, and thus the probability becomes smaller. Note that $E_i(t)$ and $E_i(t - 1)$ are symmetric zero centered random variables and so the distribution of their sum is also symmetric about zero. The argument is valid for any two adjacent trials after the first two trials, the outcomes for which define c_i .

Table 2
Proportion, $p_{50\%}$, Satisfying the 50% Criterion for $h_i(t)$ in Figure 2a

| σ_{Ei} | $p_{50\%}$ | <i>Mdn</i> |
|---------------|------------|------------|
| 1 | .000 | — |
| 2 | .004 | 11 |
| 4 | .389 | 11 |
| 8 | .855 | 11 |
| 12 | .938 | 8 |

receives the same stimulus on a subsequent posthabituation trial or two. *Spontaneous regression* refers to the fact that on these posthabituation trials some infants display performance levels above their criterion levels; that is, they rebound from the habituation criterion level observed (Bertenthal, Haith, & Campos, 1983). By now, the reader can probably anticipate our interpretation: Among those infants in the control condition there is a subset of infants who have achieved habituation "by chance," which is to say their floor values were above the required habituation criterion, as in Figure 2a. Those infants would be expected to perform above their habituation criterion level on successive trials simply because of a regression-to-the-mean effect (Davis, 1986).

The Equating Issue

A third flaw with existing techniques concerns the success with which infants are equated prior to the posthabituation test phase. To address this question requires a definition of what it means to equate infants. Neither Bornstein (1985) nor Colombo (1993) stated what is equated by infant-control procedures. Perhaps what is intended is that the relative amount by which infants travel down their habituation functions is the same and that the magnitude of travel is about half the distance from maximum to minimum attention when the 50% criterion has been achieved. This perspective has guided the following discussion.

First, recall that the 50% decrement criterion is satisfied when, for some trial T_p , a rolling average, denoted as y_{iT_i} , falls below half the initial average, denoted as $c_i/2$. Thus $y_{iT_i} = [y_i(T_i - 1) + y_i(T_i)]/2 < c_i/2$, where $[y_i(1) + y_i(2)]/2 = c_i$ is the initial average. Now, if infants were to be equated in terms of their relative proportional decrement of habituation, denoted p_i , then for infant i there would be a proportional decline defined by

$$p_i = \frac{m_i - h_i(t_{pi})}{m_i - f_i}, \quad (3)$$

where m_i denotes the maximum level of attention; f_i the floor or asymptotic value; and t_{pi} the trial associated with infant i 's habituation function, h_i , which corresponds to the proportional decrement p_i . For the moment, assume the floor value $f_i = 0$. Then p_i becomes p'_i where

$$p'_i = 1 - h_i(t_{pi})/m_i.$$

If from the 50% decrement procedure c_i is taken to be the estimate of the infant's maximum attention, and

y_{iT_i} an estimate of the infant's decline, an estimate of p'_i becomes

$$\hat{p}'_i = 1 - \frac{y_{iT_i}}{c_i}.$$

Now, the assumption that the decrement is 50% or greater requires that $p'_i \geq 1/2$, which is equivalent to $y_{iT_i} \leq c_i/2$, or, typically, $y_{iT_i} < c_i/2$.

Thus, the 50% decrement rule is equivalent to specifying that $p'_i \geq 1/2$, which assumes the floor, f_i , is zero. However, assuming a zero floor for all infants is a grossly implausible assumption given even a casual glance at many observed habituation functions (e.g., Colombo, 1993, Figures 2.5 and 2.7). Furthermore, the preceding analysis assumes but does not test whether the habituation baseline, c_i , or rolling average of the current response, y_{iT_i} , are reasonable estimates of the underlying parameters of the infant's habituation function. If rolling averages are not plausible estimates of the relevant habituation function parameters, then the 50% decrement criterion may be further reduced in sensitivity.

Next, consider the random variable $P'_i = 1 - Y_{iT_i}/C_i$ for which p'_i is a realized value. Then, using Equation 1 and the definition of C_i and Y_{iT_i} , we have

$$P'_i = 1 - \frac{h_i(T_i - 1) + h_i(T_i) + E_i(T_i - 1) + E_i(T_i)}{h_i(1) + h_i(2) + E_i(1) + E_i(2)},$$

where recall E_i is a symmetrical zero centered random variable. The expectation to consider, because it is required that $p'_i \geq 1/2$, is

$$E(P'_i | p'_i \geq 1/2) = v_i.$$

Read this expectation as follows: It is the mean of P'_i given that realized values of P'_i , that is, p'_i , are at least 1/2. To equate the expected values of all infants under the 50% decrement criteria requires that all infants have the same proportional decline in their habituation function, or $v_1 = v_2 = \dots = v_n$, which presumably also equals .50. It is not necessary that an explicit value of the expectation be given. What is clear, however, by inspection of P'_i is that the habituation functions could be different for different infants and it could be that $v_1 = v_2 = \dots = v_n$. For example, in the denominator of P'_i there are an infinite number of functions $h_1(t)$ and $h_2(t)$ such that $h_1(1) + h_1(2) = h_2(1) + h_2(2)$. However, in the absence of detail about the form of the habituation function, it is impossible to determine whether the 50% decrement procedure achieves its desired aim. Moreover, it is

difficult to imagine any plausible real-world scenario under which ν_i would be constant for all infants i unless all infants followed the same habituation function and with floor zero, remembering that the definition of P'_i assumes $f_i = 0$ in Equation 3. Consequently, we may conclude the following:

To equate relative habituation levels under the 50% decrement criterion plausibly requires that infants share a common habituation function with floor zero.

In summary, current procedures based on the 50% decrement habituation criterion have a number of apparently unrecognized flaws associated in part with the failure to specify a model of habituation. Table 1 shows that existing procedures do not allow nonhabitators to be discriminated reliably from habituators. Current procedures are subject to a floor effect, which among other deficiencies tends to recruit infants with large error variability but not infants with small error variability. In addition, current procedures can be regarded as equating in relative habituation levels only if a series of implausible assumptions are made. An alternative approach that rectifies many of these difficulties is presented next.

A Strategy and Model

Because infants are sufficiently different in their behavior, we regard group-based analyses in which infants are regarded as being replicates of one another as having flaws and, in some cases, violating critical assumptions. Consequently, four models are proposed; each is applicable at the level of the individual infant. It is assumed that each infant follows one model. To decide which model an infant follows, we propose a model selection strategy. Because infants following the same model are not assumed to share common parameter values, and thus may have unique trajectories, the models may be regarded as random effects models.

A Habituation Model Framework

The function $h_i(t) = h_{i(4)}^\star$ defines a model with four parameters, Model 4:

$$h_{i(4)}^\star(t) = \beta_i \exp[-\delta_i(t - \gamma_i)^2] + \alpha_i$$

This is the general model for infant i with trials $t = 1, 2, \dots, T_i$. It has four nonnegative parameters, α_i , β_i , δ_i , and γ_i . With the parameter γ_i it is able to model nonmonotonic curves, typically thought to be required

for dual-process theory (Groves & Thompson, 1970; Kaplan & Werner, 1986). It fits well peaks in early trial data (e.g., Colombo & Mitchell, 1990, p. 201). Model 4 will not be considered further here, but three special cases of it are important. Setting $\gamma_i = 1$ yields Model 3:

$$h_{i(3)}^\star(t) = \beta_i \exp[-\delta_i(t - 1)^2] + \alpha_i$$

This three-parameter model is generally required for modeling most habituating infants. This model resembles Dannemiller's (1984, Equation 3) and satisfies the three conditions proposed above. Its ceiling occurs at $t = 1$; its floor is α_i , which, for many infants, is often well above the possible floor of zero. β_i is termed the *depth* of the response function because it measures the overall ordinate width of the function from ceiling to floor, whereas δ_i specifies the "slope" and reflects how quickly the infant's habituation response declines over trials; the larger the value of δ_i , the more rapidly the infant habituates. For larger values of δ_i , the function takes the shape of a step function and can accommodate an initial steplike decrease, thus avoiding the need to propose two distinct functions (Cohen & Menten, 1981). Letting the notation $h_{i(3)}^\star(t; \alpha_i, \beta_i, \delta_i)$ denote the values of the function's arguments, Figure 2a is $h_{i(3)}^\star(t; 25, 16, .16)$.

Some infant habituation responses can be modeled with two parameters when their floor values α_i are zero, and so Model 2 is

$$h_{i(2)}^\star(t) = \beta_i \exp[-\delta_i(t - 1)^2].$$

Figure 2b is an example with $h_{i(3)}^\star(t; 0, 16, .16) = h_{i(2)}^\star(t; 16, .16)$. Finally, if $\beta_i = 0$ or depth is zero, and $\alpha_i \geq 0$, the infant's responses do not depend on t . The result is Model 1:

$$h_{i(1)}^\star(t) = \alpha_i$$

This is the nonhabituation model of Equation 2. A nonhabituation model does not necessarily mean that infants are not being familiarized with the stimuli displayed. Infants may respond to novel or changed stimuli quite apart from any habituation process, as Cohen (2001) recently reminded us.

Modeling real data requires an error structure. Define the regression structure as

$$Y_i(t) = h_{i(j)}^\star(t) + E_i(t), \quad j = 1, 2, 3, 4, \quad (4)$$

which is Equation 1 with $h_{i(j)}^\star(t)$ replacing $h_i(t)$. Equation 4 is the basis for all model fitting and estimation.⁵

Model Selection With the Bayesian Information Criterion

Any of the models may be fit to data. The problem is to select the most appropriate model for each infant's response data. The method proposed is use of the Bayesian information criterion (BIC; Ramsey & Schafer, 1997; Schwarz, 1978). The rationale for BIC is straightforward. Better models fit the data by capturing more of the variance in the data. Simple models are preferred over more complex ones, and complexity is measured by the number of model parameters. The more parameters there are in a model, the easier it is to fit the model to data, but adding parameters should have a cost, and this cost is reflected in BIC as a penalty. BIC is defined as

$$\text{BIC}_i(j) = t_i \log(\hat{\sigma}_{Ei}^2) + j \log(t_i),$$

where j denotes the model as well as the number of parameters, $\hat{\sigma}_{Ei}^2$ denotes the estimated error variance under model j , and t_i denotes the number of trials for i . The penalty is $j \log(t_i)$ and linearly increases with the number of parameters, j . BIC is computed for each model, and the model with the smallest BIC is selected as the best model for infant i .

Model Estimation

Estimation under the constant Model 1 is $\hat{\sigma}_{Ei}^2 = \sum_{t=1}^{t_i} [y_i(t) - \bar{y}_i]^2 / (t_i - 1)$, where $\bar{y}_i = \sum_{t=1}^{t_i} y_i(t) / t_i$, which are the ordinary sample variance and sample mean of the observed responses. The height of the constant function for infant i is estimated by $\bar{y}_i = \hat{\alpha}_i$. One could also approach the problem within a linear models framework.

Estimation under Models 2 to 4 requires nonlinear least squares estimation (Seber & Wild, 1989, Equation 13.25), which provides estimates of the parameters of each model and of σ_{Ei}^2 , the residual error variance. Although Model 2 can be log-linearized, and thus approached with conventional linear regression, the error variance estimate under the log-linearization is not the appropriate one for the BIC comparisons, and so nonlinear regression is still required. An estimation algorithm is given in Appendix B.

Model Applications in Simulated and Real Data

In this section, the alternative model-based approach is evaluated both using simulated data and

using real data. Two simulations are reported. First, data are simulated under the nonhabituation model, $h_{i(1)}^\star(t)$, which is then fit to $h_{i(3)}^\star(t)$, the wrong model. This simulation will reveal whether the habituation model can lead to spurious findings, that is, reporting habituation when there is none. In the second simulation, data are simulated under $h_{i(3)}^\star(t)$ and then fit under $h_{i(3)}^\star(t)$. This simulation provides information about the power of the model and when it fails to detect true habituation when it occurs.

Simulations Under $h_{i(1)}^\star(t)$ With Fitting to $h_{i(3)}^\star(t)$, the Wrong Model

Simulations were performed under the nonhabituation model $h_{i(1)}^\star(t)$ using the same combinations of α_i and σ_{Ei} as reported in Table 1. There were eight trials for each simulation. We fit $h_{i(1)}^\star(t)$ and $h_{i(3)}^\star(t)$ to the data for each infant, then $\text{BIC}_i(1)$ and $\text{BIC}_i(3)$ were computed.⁶ The proportions of the 1,000 simulations for which $\text{BIC}_i(3)$ was less than $\text{BIC}_i(1)$, indicating $h_{i(3)}^\star(t)$ (the wrong model) was the best fitting model, were computed. The largest error rate was .031. The median error rate was about .015. Thus, spurious errors are very unlikely. These rates may be compared with the error rates under the 50% decrement procedure reported in Table 1, which sometimes approach 1.

Simulations Under $h_{i(3)}^\star(t)$ With Fitting to $h_{i(3)}^\star(t)$, the Correct Model

Two simulation models were selected. One is the model of Figure 2a, which is $h_{i(3)}^\star(25, 16, .16)$; the other was $h_{i(3)}^\star(25, 36, .16)$. These models differ only in their depth, with $\beta_i = 16$ in the first case and $\beta_i = 36$ in the second. The values of $\sigma_{Ei} = 4, 8, 12, 16$ spanned the ranges of estimated values of σ_{Ei} observed in real infants judged to follow Model 3. There were eight trials for each virtual infant.

The results are given in Table 3. As is to be expected, model selection accuracy is better for small σ_{Ei} than for large, and for $\sigma_{Ei} \leq 4$, it is nearly perfect. However, the role of the depth is critical. The difference in depth between the two functions is 20 ($= 36 - 16$), and this depth influences performance impor-

⁵ Writing $E_{i(j)}(t)$ might be preferable because the error will be different under different models for infant i , but confusion seems unlikely.

⁶ It might be suggested that Models 1, 2, and 3 should all be fit, not just Models 1 and 3. Doing so results in no changes in results.

Table 3
Proportion of Times That $BIC_i(3) < BIC_i(2)$, $p_{3<2}$, With Data From $h_{i(3)}^(t)$*

| σ_{Ei} | $p_{3<2}$ | |
|---------------|------------------------------|------------------------------|
| | $h_{i(3)}^*(t; 25, 16, .16)$ | $h_{i(3)}^*(t; 25, 36, .16)$ |
| 4 | .96 | 1.00 |
| 8 | .51 | .97 |
| 12 | .28 | .80 |
| 16 | .17 | .61 |

tantly. Power drops off substantially as the depth becomes shallower, from .97 to .51 for $\sigma_{Ei} = 8$. Combining a large standard deviation, $\sigma_{Ei} = 16$, and shallow depth results in poor model performance. Fortunately, real data suggest that the depth, at least for 4-month-old habituating infants, is typically large. As a result, the model-based procedure is likely to be superior to existing techniques, particularly when one considers the difficulties associated with conventional procedures such as those displayed in Tables 1 and 2.

Assessing Habituation in Real Data

Experimental method and data. In order to assess the feasibility of the model-based approach, we analyzed data collected from a separate study of infants' responses to different patterns of optic flow (Gilmore & Rettke, 2003). Thirty-five 4- to 6-month-old infants repeatedly viewed an optic flow pattern during the familiarization phase until their looking times, recorded by an observer monitoring gaze duration on a computer-based timer, had reached a habituation criterion. Habituation was determined online using Ashmead and Davis's (1996) proposal; typically, trials were terminated when the infant looked away from the display for 1 s, which was the observer's response time to an infant's looks to and from the display. Following the fifth habituation trial, a second-order polynomial was fitted to the data. When the fitted look time following a given trial dropped below 50% of the fitted (predicted) look time for the first habituation trial, the familiarization phase ended. During the subsequent four-trial test phase, the familiar pattern and a novel pattern were presented twice in alternating order. Details of the method are provided elsewhere (Gilmore & Rettke, 2003).

Description of results. Models 1, 2, and 3 were fitted to the data, and the best fitting model, using BIC, was determined for each infant. In fitting these models, we used all habituation trials; the number of trials on which the estimates were based ranged from 5 to 13.

Among the 35 infants, 3 infants followed Model 2, 20 infants followed Model 3, and 12 followed Model 1. Thus, 12 infants of 35, or $12/35 = .34 = \hat{\lambda}$, is the estimated proportion of infants who are nonhabitua-tors, whereas 23 of 35, or $23/35 = .66 = 1 - \hat{\lambda}$, is the estimated proportion of habituators.⁷ Table 4 gives summary information for the parameter estimated of the habituating infants. The last column, r , is the correlation coefficient between the actual values and the fitted values under each infant's fitted model (Model 2 or Model 3). The models generally fit well, and in all cases r is generally high. The estimated error, $\hat{\sigma}_{Ei}$ was about 7 s. There is considerable range in the values of the estimated parameters; for example, the floor estimates of α_i ranged from 0 to 27, whereas the function depth estimate $\hat{\beta}_i$ was typically about 37 and ranged from 6 to 67. Among the 12 infants following Model 1, the mean of the estimated constant, that is, $\hat{\alpha}_i$, was 21.05 with a standard deviation of 11.32. The estimated error, $\hat{\sigma}_{Ei}$ ranged from 2.13 to 24.99.

These results show that even when data have been collected under other habituation criteria, our model-based approach can usefully reveal important information about infants. We illustrate how some of this information can be used next.

Posthabituation Consequences of the Model-Based Approach

Although the focus in this article has been on habituation and not posthabituation procedures, two examples illustrate that modeling the habituation process can provide practical and conceptual advantages even when the focus is on posthabituation settings. The first example focuses on the problem of assessing the influence of novelty responses in the posthabituation phase. The second illustrates the unique possibilities provided by modeling individual responses.

Novelty-familiarity tests. The conventional approach to this problem, in a simple setting at least, is to compare the difference between posthabituation responses to a novel and a familiar stimulus. Let D_i denote the novel minus familiar response difference random variable with realized value d_i . The standard procedure in the simplest setting would be to use a

⁷ Earlier (Gilmore & Thomas, 2002) in a study using the same data, only two models were considered, Models 3 and 1, not three models as here. The three infants following Model 2 were earlier regarded as following Model 1.

Table 4
Parameter Estimates for 23 Infants Fitting $h_{i(2)}^*(t)$ or $h_{i(3)}^*(t)$

| Estimate | $\hat{\alpha}_i$ | $\hat{\beta}_i$ | $\hat{\delta}_i$ | $\hat{\sigma}_{Ei}^2$ | r |
|----------|------------------|-----------------|------------------|-----------------------|---------|
| M | 7.40 | 37.49 | 2.89 | 6.78 | .92 |
| SD | 6.74 | 16.87 | 4.59 | 4.97 | .09 |
| Range | 0–26.64 | 5.64–66.9 | .01–15.84 | .71–15.36 | .61–.99 |

paired t test to test whether the mean difference, $\bar{d} = n^{-1} \sum_{i=1}^n d_i$, is significantly positive. Because in our study 35 infants were presented, in the posthabituation phase, with two trials of the same familiar habituation stimulus and two trials of a novel stimulus, in a counterbalanced design, the most natural comparisons would be to take the average of each infant's two novel responses minus the average of the two familiar responses as d_i . Doing so yields $t(34) = 2.52$, $p < .008$.

There are several reasons for suspecting this procedure would not be optimal, and so one should be leery of accepting this probability level uncritically. First, the assumption of the d_i differences' following a common normal distribution, a requirement of the t test, is surely false ($p < .0001$; Anderson & Darling, 1954). Second, given that the error variances associated with each infant's best-fitting model in the habituation phase ranged from 0.5 to 624—they would be even more widely dispersed under a common model—it is hard to imagine how the variances of the D_i , which are assumed to be constant for all i , under the paired t test assumptions, could be even roughly constant in the posthabituation phase. Consequently, a better approach to evaluating the significance level of t would be to use the bootstrap. A bootstrap probability level called an *achieved significance level* (ASL) by Efron and Tibshirani (1993) is .02.

There is potentially a much more serious assumption violation, however, and that is that the *identically distributed* assumption fails. This assumption requires that the differences in responses, that is, the random variables D_i , for all infants, follow a common distribution (and under the t test, a common normal distribution). This implies that the $E(D_i)$ be all the same and the $\text{var}(D_i)$ be all the same independent of i . Testing and rejecting the hypothesis that the infant difference scores followed a normal distribution means they follow some other distribution. And so, at minimum, this means they cannot follow a common normal distribution and thus are not identically distributed from a normal distribution. In addition, the fact that the observed error variances of the infants, that is, $\hat{\sigma}_{Ei}^2$

under each i 's best fitting model, were hugely different, as just noted above, is further evidence that the D_i do not follow a common distribution with, of course, a common variance. The consequences for the paired t test, if the common normal distribution of the D_i assumption is violated, are troublesome. In the paired t test the sample variance, $s_{d_i}^2$, is based on the observed d_i . However, this estimate, $s_{d_i}^2$ will estimate the wrong quantity and be too large unless the D_i follow a common distribution. This is easily seen by considering the expectation of the sample variance of the difference scores for the paired t test. It contains a quadratic term, $\sum_{i=1}^{n-1} \sum_{j=i+1}^n (\eta_i - \eta_j)^2$, where $E(D_i) = \eta_i$ and $E(D_j) = \eta_j$ for infants i, j . If all infants share the same difference score distribution, this implies that $E(D_i) = E(D_j)$ for all i, j pairs of infants and this quadratic term is zero, as it will be under the paired t test when the assumptions of the paired t test are satisfied. Given the rejection of the hypothesis that the D_i follow a normal distribution—the huge differences among infants in their error variances in the habituation phase—it is difficult to see how this identically distributed assumption could be plausible as an assumptive framework for the posthabituation data.

The quadratic term can be estimated. For the 35 infants in our study, there are $\binom{35}{2} = 595$ squared differences that are summed. This sum could be very large and has the potential to substantially reduce the t test's power because the denominator of the t statistic will be too large. An estimate of this term based on our data is 37,742 (see Appendix C), suggesting the potential for substantial loss of power in the paired t test. There is no fix for the problem in conventional theory, but there is one under our model.

The proposal is to use the estimated error variance $\hat{\sigma}_{Ei}^2$ under each infant's best fitting model obtained in the habituation phase as the estimate of error in the posthabituation or test phase, thus avoiding the use of the sample variance of the observed d_i difference scores. After all, each infant has (in our study) a minimum of five responses to what is the familiar stimulus in the posthabituation phase. Consequently, this should be a very good estimate of error variance for

each infant, and it is assumed the novelty and familiar responses share, for infant i , the same variance. Thus, with estimates $\hat{\sigma}_{Ei}^2$ consider the statistic

$$g = \frac{\bar{d}}{\sqrt{\frac{(1-\hat{\rho}) \sum_{i=1}^n \hat{\sigma}_{Ei}^2}{n^2}}}, \quad (5)$$

where $\hat{\rho}$ is the estimated correlation between novel and familiar responses (see Appendix C). In the present case, $g = 3.16$; reference to a normal distribution gives $p < .0008$. The bootstrap ASL is $p < .007$.

It will rarely be the case that a more efficient procedure will invariably dominate a less efficient one in any given real data set. That is the case here because the decision is the same for both procedures in the present example. In many cases, however, g should make a practical difference. Applied to the habituation group separately, the g statistic mirrored the results just reported. For the 12 nonhabitua-tors, no practical difference between the g and t procedures emerged; actually, the nonhabitua-tors did just show a .05-level difference, revealing that they too attended longer to the novel stimulus. When the assumption of a common normal difference score distribution for all n infants seems unreasonable, g should be more powerful than the paired t test.⁸

Improving individual parameter estimates by "borrowing strength." Consider the problem of using measures in a habituation setting to predict performance of some practical relevance, such as cognitive task performance in later years. It is known that measures of habituation correlate about $r = .41$ with measures of IQ in later childhood (McCall & Carriger, 1993), which is about twice the corresponding correlation with conventional infant test measurements. The habituation predicting variable is usually some function of the response to novelty (McCall, 1994). One likely reason why the correlation is not large enough for useful prediction is because of the uncertainty of the measurements in infancy. This translates into how variable are the estimates of the habituation quantity of focus for each infant. Within the context of the novelty responses, if it were possible to improve the estimate of an individual's η_i , it seems likely that habituation measurements could be much more useful.

Improving the estimate $\hat{\eta}_i$ is certainly possible, and the general strategy for doing so is not new. It is usually termed *empirical Bayes*, and there is both theory and data demonstrating that an individual's

parameter estimate can be improved by using not only the target individual's parameter estimate but also information from other unrelated individuals, a procedure sometimes referred to as *borrowing strength* (e.g., Bryk & Raudenbush, 1992; Casella, 1985; Efron & Morris, 1977; Maritz & Lwin, 1989; Thomas, 1993). Perhaps the most famous example is the improvement in the prediction of a baseball player's end-of-season batting average by using the target player's midseason average along with averages of other players at midseason (Efron & Morris, 1975).

What is required is an estimate of an individual's target parameter under a suitable model. Also required is an estimate of each individual's variability, a function of which typically provides an estimate of the variability of the individual's parameter estimate as well. In the present setting, one has estimates $d_i = \hat{\eta}_i$ and the desire is to obtain an improved estimate, $\tilde{\eta}_i$ which is often of the form $\tilde{\eta}_i = A_i + B_i \hat{\eta}_i$ a unique linear function for each i . The new $\tilde{\eta}_i$ is shrunken in the sense that very large estimates are made smaller, and smaller estimates are made larger; how much change is produced depends on the overall variability of all the $\hat{\eta}_i$ values and the size of $\hat{\sigma}_{Ei}^2$.

Very briefly, $\hat{\eta}_i$ values for the 23 habituating infants were estimated using linear empirical Bayes, a largely distribution-free procedure (Maritz, 1989; Maritz & Lwin, 1989; Thomas, 1993). For instance, for Infant 2, $\hat{\eta}_2 = -17.95$, whereas $\tilde{\eta}_2 = -6.17$; $\hat{\eta}_4 = 33.35$ is shrunk to $\tilde{\eta}_4 = 17.22$, and $\hat{\eta}_{14} = -.165$ is slightly changed to $\tilde{\eta}_{14} = 1.86$. It is interesting that among the 23 habituating infants, 8 had their initial novelty responses less than zero (i.e., $\hat{\eta}_i < 0$), whereas for only 3, adjusted responses were less than zero (i.e., $\tilde{\eta}_i < 0$). This result suggests that substantially

⁸ Based on 20,000 simulations with $n = 30$ and D_i the differences between pairs of bivariate normal random variables with correlation $\rho = .25$, with $E(D_i)$, and σ_{Ei}^2 varying but $E(D) = 0$, G_1 , Appendix C, under the null hypothesis $E(D) = 0$, follows very closely the standard normal distribution. In these studies σ_{Ei}^2 and ρ were estimated from data. In the nonnull setting the power of g_1 depends on σ_{Ei} and η_i but is generally not measurably less and often is much more powerful than the paired t test. For example, sample 29 difference scores from $n(0, \sigma_{Ei} \sqrt{2(1-\rho)} = 1)$, and 1 difference score from $n(6, \sigma_{Ei} \sqrt{2(1-\rho)} = 1)$; $\rho = .25$, $n = 30$, and $E(D) = .2$. The proportion of simulations in which t exceeded the .05 one-tail critical value was .10; for g_1 with estimates of ρ and σ_{Ei} the proportion was .30, so g_1 is 3 times more powerful in this example.

more infants in the habituation group than was originally thought did prefer the novel stimulus, but the variability of their original estimates $\hat{\eta}$ concealed this fact. Whether such procedures are useful in a wider sense remains, of course, an open issue. However, without a model-based framework, there is no possibility of knowing.

A habituation criterion and equating habituation levels. A consideration of Figure 2 suggests that a plausible habituation criterion should be independent of the habituation function's ceiling and floor yet assess how far down the function the infant has traveled. In Equation 3 let $h_{i(3)}^*(t_{pi})$ replace $h_i(t_{pi})$. Under $h_{i(3)}^*(t)$ the maximum is on Trial 1, so $m_i = h_{i(3)}^*(1) = \beta_i + \alpha_i$, and the asymptote is $\lim_{t \rightarrow \infty} h_{i(3)}^*(t) = \alpha_i = f_p$, the floor value. Equation 3 becomes, after substitution,

$$p_i = 1 - \exp[-\delta_i(t_{pi} - 1)^2]. \quad (6)$$

Thus, the proportion of decrement, p_i , is independent of the ceiling and floor parameters as desired. Once δ_i is specified or estimated, the corresponding decrement is easily found. For example, among the 23 infants in our study following Models 2 or 3, 1 had an estimated p_i value of .68 whereas all other estimates exceeded .92. These quantities might be useful as covariates in certain studies. However, interest is likely to be focused on determining the trial on which to terminate habituation for an investigator-specified p ; $p = 1/2$ corresponds to what a 50% decrement criterion would be intended to assess.

To derive an expression for this trial value, replace p_i with p in Equation 6 and solve for t , denoting the solution $t_i(p)$ (to denote its dependence on i and p). The result is

$$t_i(p) = \sqrt{\log[1/(1-p)]/\delta_i} + 1. \quad (7)$$

Because $t_i(p)$ will in general not be an integer, the decision rule becomes

$$\text{stop habituation on trial} = \{\sqrt{\log[1/(1-p)]/\delta_i} + 1\}, \quad (8)$$

where the braces in Equation 8 denote rounding up to the next integer. For example, with $\delta_i = .16$, the value in Figure 2, set $p = 1/2$ for a 50% decrement. This gives $t_i(1/2) = 2.32$ yielding $\{2.32\} = 3$, so habituation would stop on Trial 3.

With this development, the issue immediately surfaces as to whether the current proposal will lead to fewer trials to a habituation criterion than the current 50% decrement procedure. The answer, in general, is

assuredly yes, but to specify how much more efficient it will be is more complicated. First, note that the above calculation assumes δ_i is known, when it will need to be estimated. Second, recall that the present proposal enables one to discriminate which infants habituate and which do not in an extremely efficient way. There is no mechanism, other than an arbitrary one, for deciding the matter under the 50% decrement criterion. Among those 12 nonhabitutors in our illustrative study, only 1 would have achieved the (two-trial window) 50% criterion in the number of trials presented, which ranged from 5 to 13, and in general, because nonhabituating infants typically display responses fluctuating about a horizontal line, they often require many more trials to achieve habituation (again, in the spirit of Bogartz, 1965); how many more trials depends on the magnitudes of their error variances. Turning to the truly habituating infants and those with small error variance, the "ideal infants" may never be identified under the 50% rule, as Table 2 and Figure 2 make clear, whereas the model-based procedure will surely catch them, as Table 3 reveals. Among those 23 in our habituation group, 5 would not have achieved the 50% habituation criterion given the trials presented.

However, perhaps most interesting is the fact that if only the first five trials of each of the 35 infants are used, and the model parameters are estimated with corresponding model selection using BIC, all 12 of the nonhabitutors would retain their status as nonhabitutors; 20 of the 23 habitutors would again be classified as habitutors, and the remaining 3 would have been classified as nonhabitutors. This empirical result of 91% ($100 \times 32/35\%$) classification agreement strongly suggests that, should the current proposal be adopted, then given a return to a fixed trials procedure with perhaps as few as five habituation trials, the efficiencies of the habituation paradigm would markedly improve.

However, rather than focusing on the number of trials to some habituation criterion, which cannot, except in a very crude way, equate infants in their relative levels of habituation, consider a simpler and more elegant procedure. In the equation for $h_{i(3)}^*(t)$, replace t with the right side of Equation 7. Denote the resulting quantity $y_i(p)$:

$$y_i(p) = \beta_i(1-p) + \alpha_i. \quad (9)$$

This expression has the spirit of prediction in simple linear regression. For a given proportional decrement p , there is a linear expression in depth value β_i and the

floor value α_i to predict the corresponding magnitude of the response for any p desired.

This expression has some important conceptual advantages. First, the experimenter can control the proportional decrement by specifying a value for p . Second, different infants are thus relatively matched in their proportional declines by $y_i(p)$. Although $p = 1/2$ is a possibility, there appears to be no particular advantage, other than minimizing the number of habituation trials, in setting $p = 1/2$ when $p = 1$ can be specified, corresponding to a 100% decline. Moreover, doing so means the posthabituation recovery comparison is easier, because the comparison is with the infant's floor value, not some other value above the floor or below the floor, as in Figure 2a.

In practice, β_i and α_i must be estimated, and so Equation 9 becomes

$$Y_i(p) = \hat{\beta}_i(1 - p) + \hat{\alpha}_i, \quad (10)$$

where $Y_i(p)$ is now random because the estimates $\hat{\beta}_i$ and $\hat{\alpha}_i$ are values of random variables. The variance of Equation 10 is

$$\begin{aligned} \text{var}[Y_i(p)] = & (1 - p)^2 \text{var}(\hat{\beta}_i) \\ & + \text{var}(\hat{\alpha}_i) + 2(1 - p) \text{cov}(\hat{\beta}_i, \hat{\alpha}_i), \end{aligned}$$

where cov denotes the covariance. Note that although the variance of $Y_i(p)$ generally depends on the variances of $\hat{\beta}_i$ and $\hat{\alpha}_i$ and their covariance, as p grows the variance of $Y_i(p)$ becomes smaller so that when $p = 1$, $\text{var}[Y_i(p)] = \text{var}(\hat{\alpha}_i)$. Estimates of these variances (and covariance) are readily obtained (see Appendix B). Thus, not only does the investigator have control over the proportion p or an estimate of the percentage decline desired but simultaneous control is acquired over the variance of $Y_i(p)$ as well. In the spirit of the problem, one might call this proposal an *investigator-control* procedure.

Online implementation. As the previous section illustrates, the advantages of the modeling procedure are not dependent on implementing the procedures online. However, code for generating displays and performing online analysis has been implemented within a MATLAB (Version 5.2.1) environment. To determine how well the model can be implemented in real time online, we considered a pilot study with a design similar to that reported above. It involved 6 infants, ages 3 to 6 months. Each infant was presented with the habituating stimulus for six trials. After this trial and on successive trials, online fits were obtained for both $h_{i(1)}^*(t)$ and $h_{i(3)}^*(t)$, along with BIC for each model. Thus, at the end of six trials and at the end of

subsequent trials, parameter estimates and measures of fit are available under each model, and a determination could be made as to whether the infant was a habituator or a nonhabituator based on BIC. In addition, the 50% decrement rolling averages, along with other quantities, are computed. Once the criterion specified by the study's design has been achieved, the software controls the presentation of posthabituation stimuli and records responses according to the design of the study. The habituation intertrial interval required for these calculations was 2 to 4 s, consistent with the intertrial intervals reported by others. No apparent difficulties in implementation were found. The software can be easily modified to reflect the design requirements of different studies. Thus, online implementation is readily achievable.

Discussion

It has been argued that although visual habituation is the dominant method in studies of infant perception and cognition, there has been little progress in understanding the quantitative basis of the phenomenon and only minimal change in methodology for more than 25 years. When this methodology is examined closely, however, some important difficulties are recognized.

Our analysis shows that even if an infant satisfies the 50% decrement criterion, the most commonly used habituation criterion, there is little assurance that the infant has in fact habituated. The infant can achieve the habituation criterion simply by chance as the simulation reported in Table 1 makes clear. Thus, the procedure is highly likely to "report" nonhabituating infants as habituators. It also can fail to recognize ideal infants—those who do actually habituate and have small error variance unless their habituation floor values are at or near zero. Figure 2 illustrates this difficulty, which is reinforced by the corresponding simulated results reported in Table 2. Consequently, it is an inefficient procedure for identifying suitable samples of habituated infants for posthabituation assessment. Among the consequences of these difficulties have likely been the loss of infants—particularly truly habituating infants—to research studies; loss of statistical power in assessing posthabituation processes for those infants remaining, because there are likely to be nonhabitators in the sample (as Figure 1 demonstrates); and difficulties associated with the spontaneous regression phenomenon, difficulties probably spawned by the very use of the 50% decrement criterion.

The proposed model-based procedure overcomes these difficulties and does so with high efficiency, declaring nonhabituating infants as habituators less than 2% of the time (as reported above). It correctly identifies habituating infants as habituators extremely well, as Table 3 reveals, failing only if the infant's error variance is very high, or if the habituation function is quite "flat" and lacks depth (i.e., β is small). However, the procedure goes much further: It does not treat infants as replicates of one another as is required by existing data analysis frameworks but rather takes an individual-differences perspective by seeking the best fitting model, from a broad class of models, for each individual infant. Four models are proposed, along with a model selection criterion, BIC, for deciding which model is most appropriate for a given infant. Models 2 and 3 satisfy the proposed conditions every habituation function should satisfy. Model 4, although not considered here, can accommodate initial rises in attention, and Model 1 models nonhabituating infants. Within these nested models, the parameters for an individual infant are assumed to be random (across infants) and can be estimated for each infant along with the infant's associated error variance under the model.

There has always been, within the wider research community, a recognition that differences among infants in their responses are often very large. In one sense our proposal can be viewed as simply a logical development of this long-held belief. However, coincident with this proposed change of focus from groups to individuals, there has come a concomitant relaxation of data-analytic model assumptions required under current procedures with likely gains in efficiency of alternative procedures. The comparison of the t and g procedures discussed above is one example. Although not resulting in a change in decision for the illustrative data set, g should be more efficient in the longer run (which is the only way its efficiency can be assessed).

To tout the ability of the framework to estimate each infant's error variance is unlikely to spark much initial enthusiasm from investigators. But, in fact, this is an important advantage. Heretofore, there has been no suitable way to assess how "good" or how "bad" a response from an individual infant might be because there lacked an estimate of the individual's variability. However, now, with the estimation capabilities of the framework, individual infant parameter estimates can be reestimated using information from other infants, resulting in what should be improved, more

accurate individual estimates. This is one basic goal of empirical Bayes procedures (e.g., Maritz & Lwin, 1989), a simple example of which was provided above; however, the range of possibilities is much broader. For example, reliability estimates have typically used some simple observed quantities such as lengths of first fixations (Colombo, Mitchell, O'Brien, & Horowitz, 1987) observed at two time points; the resulting $r = .26$ between them is an index of reliability and is obviously low. Consider replacing these quantities with parameter estimates, which are functions of all the infant's data, not just the first trial data, and consequently will be more stable and reliable. For example under Model 3, $\hat{\alpha}_i + \hat{\beta}_i$ estimates the response magnitude on first fixation. Now reestimate this quantity using corresponding information from other infants. The strategy should improve both reliability and, likely, longer term predictions—and, if implemented appropriately for posthabituation tests, might make that setting more sensitive as well.

The model proposed is a simple one, and some assumptions, particularly about the error structures, are certainly questionable. First, is it reasonable to assume the errors are independent over trials for a single infant? Perhaps not; the issue has to be explored, and it can be most easily, perhaps, through standard time series autoregression frameworks (Fox, 1997). If the errors are positively correlated, as seems likely for some infants, the most serious consequence is likely to be that estimates or standard errors are larger than they are reported to be under an independence model. Second, are the variances over trials within an infant constant? They may not be. If a functional dependency of the magnitude of the errors on t could be specified, the structure could likely be accommodated within the existing model framework, requiring no additional parameters. Third, it might be asked if an additive error structure is sensible. Why not a product structure? An easily implemented alternative product model is

$$Y'_i(t) = \exp\{\beta_i \exp[\delta_i(t-1)^2] + \alpha_i\} V_{Ei}$$

Here V_{Ei} is a lognormal random variable with mean 1. Taking logarithms of both sides of the above expression returns Model 3 with $Y_i(t) = \log[Y'_i(t)]$. Thus, by taking logarithms of infant responses and fitting the models proposed above, one is effectively fitting variations of this product model. Models of this form were initially explored but were found to fit the data less well than the additive models proposed.

There remain intriguing questions. For example, it

is clear, at least in some settings, that nonhabituation does not mean that infants are unable to distinguish at least some stimuli in the posthabituation setting. Our data analysis above suggested this. This finding should come as no surprise, because it has long been known that organisms process certain features of the environment that are not always manifested in the specific behavioral measurements taken. The idea goes back at least 70 years to E. C. Tolman's (1932) notions of *latent learning* and, somewhat more recently, to the study of responses to novelty, complexity, and curiosity (e.g., Berlyne, 1960; Thomas, 1973), studies which have typically been outside of a habituation context. It may well be that conceptual insights can be revealed by focusing on nonhabituating infants. Such responses seem less confounded with habituation processes and, consequently, may lead to greater understanding of processes such as the role of sensitization in initial responses.

In conclusion, we think we have shown that our model-based procedure is both more powerful and more efficient than existing techniques. Although there remain important issues to resolve, the current proposal offers both practical and theoretical benefits to researchers. There appears to be very little, if any, downside or risk in considering the proposed alternative methodology except for its initial implementation. But the upside could be a large return.

References

- Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, *49*, 265–269.
- Arterberry, M. E., & Bornstein, M. H. (2002). Infant perceptual and conceptual categorization: The roles of static and dynamic stimulus attributes. *Cognition*, *86*, 1–24.
- Ashmead, D. H., & Davis, D. L. (1996). Measuring habituation in infants: An approach using regression analysis. *Child Development*, *67*, 2677–2690.
- Benasich, A. A. (2002). Impaired processing of brief, rapidly presented auditory cues in infants with a family history of autoimmune disorder. *Developmental Neuropsychology*, *221*, 351–372.
- Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.
- Bertenthal, B. I., Haith, M. M., & Campos, J. J. (1983). The partial-lag design: A method for controlling spontaneous regression in the infant-control habituation paradigm. *Infant Behavior and Development*, *6*, 331–338.
- Bogartz, R. S. (1965). The criterion method: Some analyses and remarks. *Psychological Bulletin*, *64*, 1–14.
- Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of vision and audition in the first year of life* (pp. 253–301). Norwood, NJ: Ablex.
- Bornstein, M. H. (1998). Stability in mental development from early life: Methods, measures, models, meanings and myths. In F. Simon & G. E. Butterworth (Eds.), *The development of sensory, motor and cognitive capacities in early infancy: From sensation to cognition* (pp. 299–331). Hove, England: Psychology Press.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician*, *39*, 83–87.
- Clarke, L. E. (1975). *Random variables*. London: Longman.
- Clifton, R. K., & Nelson, M. N. (1976). Developmental study of habituation in infants: The importance of paradigm, response system, and state. In T. J. Tighe & R. N. Leaton (Eds.), *Habituation: Perspectives from child development, animal behavior, and neurophysiology* (pp. 159–205). Hillsdale, NJ: Erlbaum.
- Cohen, L. B. (1976). Habituation of visual attention. In T. J. Tighe & R. N. Leaton (Eds.), *Habituation: Perspectives from child development, animal behavior, and neurophysiology* (pp. 207–238). Hillsdale, NJ: Erlbaum.
- Cohen, L. B. (2001, April). Uses and misuses of habituation: A theoretical and methodological analysis. In R. O. Gilmore (Chair), *Theoretical advances in habituation*. Symposium conducted at the meeting of the Society for Research in Child Development, Minneapolis, MN.
- Cohen, L. B., & Gelber, E. C. (1975). Infant visual memory. In L. Cohen & P. Salaptek (Eds.), *Infant perception: From sensation to cognition* (Vol. 1, pp. 347–404). New York: Academic Press.
- Cohen, L. B., & Menten, T. (1981). The rise and fall of infant habituation. *Infant Behavior and Development*, *4*, 269–280.
- Colombo, J. (1993). *Infant cognition: Predicting later intellectual functioning*. Newbury Park, CA: Sage.
- Colombo, J. (1997). Individual differences in infant cognition: Methods, measures, and models. In J. Dobbins (Ed.), *Developing brain and behavior: The role of lipids in infant formula* (pp. 339–372). New York: Academic Press.
- Colombo, J., & Mitchell, D. W. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.),

- Individual differences in infancy: Reliability, stability, prediction* (pp. 193–227). Hillsdale, NJ: Erlbaum.
- Colombo, J., Mitchell, D. W., O'Brien, M., & Horowitz, F. D. (1987). Stimulus and motoric influences on visual habituation to facial stimuli at three months of age. *Infant Behavior and Development, 10*, 173–181.
- Dannemiller, J. L. (1984). Infant habituation criteria: I. A Monte Carlo study of the 50% decrement criteria. *Infant Behavior and Development, 7*, 147–166.
- Davis, C. E. (1986). Regression to the mean. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 706–708). New York: Wiley.
- Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association, 70*, 311–319.
- Efron, B., & Morris, C. (1977, May). Stein's paradox in statistics. *Scientific American, 236*, 119–127.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Fagen, J. W., & Ohr, P. S. (1990). Individual differences in infant conditioning and memory. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction* (pp. 157–191). Hillsdale, NJ: Erlbaum.
- Fantz, R. L. (1964, October 30). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668–670.
- Fearon, I., Hains, S. M. J., Muir, D. W., & Kisilevsky, B. S. (2002). Development of tactile responses in human preterm and full-term infants from 30 to 40 weeks postconceptional age. *Infancy, 3*, 31–51.
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.
- Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. New York: Academic Press.
- Gilmore, R. O., & Rettke, H. R. (2003). Four-month-olds' discrimination of optic flow patterns depicting different directions of observer motion. *Infancy, 4*, 177–200.
- Gilmore, R. O., & Thomas, H. (2002). Examining individual differences in infants' habituation patterns using object quantitative techniques. *Infant Behavior and Development, 25*, 399–412.
- Goubet, N., Rattaz, C., Pierrat, V., Allemann, E., Bullinger, A., & Lequien, P. (2002). Olfactory familiarization and discrimination in preterm and full-term newborns. *Infancy, 3*, 53–75.
- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review, 77*, 419–450.
- Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition, 86*, B15–B22.
- Hernandez-Reif, M., Field, T., Diego, M., & Largie, S. (2002). Depressed mothers' newborns show longer habituation and fail to show face/voice preference. *Infant Mental Health Journal, 23*, 643–653.
- Horowitz, F. D., Paden, L., Bhana, K., & Self, P. (1972). An infant-controlled procedure for studying infant visual fixations. *Developmental Psychology, 7*, 90.
- Johnson, N. L., & Kotz, S. (1972). *Distributions in statistics: Continuous multivariate distributions*. New York: Wiley.
- Kaplan, P. S., & Werner, J. S. (1986). Habituation, response to novelty, and dishabituation in human infants: Tests of a dual-process theory of visual attention. *Journal of Experimental Child Psychology, 42*, 199–217.
- Kaplan, P. S., Werner, J. S., & Rudy, J. W. (1990). Habituation, sensitization, and infant visual attention. *Advances in Infancy Research, 6*, 61–109.
- Laub, K. W. (1972). *Habituation of visual fixation in infants 8- and 9-weeks of age*. Unpublished Master's thesis, University of Kansas, Lawrence.
- Laub, K. W., & Bhana, K. (1974). Infant control, response decrement and recovery as an index of visual discrimination in young infants. In F. D. Horowitz (Ed.), *Visual attention, auditory stimulation, and language discrimination in young infants. Monographs of the Society for Research in Child Development, 39*(5–6, Serial No. 158, pp. 52–62).
- Lehman, E. H. (1998). *Elements of large-sample theory*. New York: Springer Publishing Company.
- Maritz, J. S. (1989). Linear empirical Bayes estimates of quantiles. *Statistics & Probability Letters, 8*, 59–65.
- Maritz, J. S., & Lwin, T. (1989). *Empirical Bayes methods*. London: Chapman & Hall.
- McCall, R. B. (1994). What process mediates predictions of childhood IQ from infant habituation and recognition memory? Speculations on the roles of inhibition and rate of information processing. *Intelligence, 18*, 107–125.
- McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64*, 57–79.
- McCall, R. B., & Kagan, J. (1970). Individual differences in the infant's distribution of attention to stimulus discrepancy. *Developmental Psychology, 2*, 90–98.
- Ramsey, F. L., & Schafer, D. W. (1997). *The statistical sleuth*. Belmont, CA: Duxbury.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
- Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
- Sokolov, E. (1963). *Perception and the conditioned reflex*. Oxford, England: Pergamon.

Thomas, H. (1973). Unfolding the baby's mind: The infant's selection of visual stimuli. *Psychological Review*, 80, 468–488.

Thomas, H. (1993). Individual differences in children, studies, and statistics: Applications of empirical Bayes methodology. In M. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development: Foundations* (Vol. 1, pp. 120–153). New York: Springer-Verlag.

Tighe, T. J., & Leaton, R. N. (1976). *Habituation: Perspectives from child development, animal behavior, and neurophysiology*. Hillsdale, NJ: Erlbaum.

Tolman, E. C. (1932). *Purposive behavior in animals and man*. New York: Century.

Woodward, A. L., & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, 17, 1061–1084.

Appendix A

Power Simulation

This Appendix provides details about the power simulation summarized in Figure 1. Let the random variables $W_{2j} - W_{1j} = V$ denote a novelty–familiarity difference in the posttest comparison. To introduce notation, let us assume for the moment that V is continuously distributed with density $g_j(v)$. $E(V) = \eta_j = \mu_{w2j} - \mu_{w1j}$, the means of the corresponding W , and $\text{var}(V) = \sigma_j^2 = 2\sigma_{wj}^2(1 - \rho_{wj})$, where W_{2j} and W_{1j} have the same variance σ_{wj}^2 and have correlation ρ_{wj} . Actually, V has the distribution $\sum_{j=1}^2 \lambda_j g_j(v)$, with nonnegative weights λ_j , where $\lambda_{j=1}$ denotes the proportion of nonhabituated and $\lambda_{j=2}$ denotes the proportion of habituators. $\sum_{j=1}^2 \lambda_j = 1$, and so $E(V) = \sum_{j=1}^2 \lambda_j \eta_j = \sum_{j=1}^2 \lambda_j (\mu_{w2j} - \mu_{w1j})$, $\text{var}(V) = \sum_{j=1}^2 \lambda_j \sigma_j^2 + \lambda_1 \lambda_2 [\eta_1 - \eta_2]^2 = \sum_{j=1}^2 [2\lambda_j \sigma_{wj}^2 (1 - \rho_{wj})] + \lambda_1 \lambda_2 [(\mu_{w21} - \mu_{w11}) - (\mu_{w22} - \mu_{w12})]^2$.

Let V_i , $i = 1, \dots, n$ be a random sample from the same distribution as V , with $\bar{V} = n^{-1} \sum_{i=1}^n V_i$. Then the $E(\bar{V}) = E(V)$, and the $\text{var}(\bar{V}) = \text{var}(V)/n$.

Next, two cases are considered, which are the basis for the power simulations. In Case 2, all infants habituate; in Case 1, a proportion habituate. Subscripts 1 or 2 appear on \bar{V} and n to distinguish each case.

For Case 2, $\lambda_1 = 0$, $\lambda_2 = 1$ (all infants habituate).

$E(\bar{V}_2) = \eta_2 = \mu_{w22} - \mu_{w12}$, and $\text{var}(\bar{V}_2) = 2\sigma_{w2}^2(1 - \rho_{w2})/n_2$. Define

$$U_2 = \frac{\bar{V}_2}{\sqrt{2\sigma_{w2}^2(1 - \rho_{w2})/n_2}}.$$

U_2 is the structure of the paired t test. If S is the standard deviation of the difference scores, V_1, V_2, \dots, V_{n_2} , then $E(S^2) = 2\sigma_{w2}^2(1 - \rho_{w2})$ under t test assumptions.

For Case 1, $\lambda_1 > 0$, $\lambda_2 > 0$ (λ_2 is the proportion of habituators).

Then

$$U'_1 = \frac{\bar{V}_1}{\sqrt{\frac{\sum_{j=1}^2 2\sigma_{\omega j}^2 \lambda_j (1 - \rho_{\omega j}) + \lambda_1 \lambda_2 [(\mu_{\omega 21} - \mu_{\omega 11}) - (\mu_{\omega 22} - \mu_{\omega 12})]^2}{n_1}}}$$

These two cases form the basis for the simulations with simplification: Set $\mu_{\omega 21} = \mu_{\omega 11}$ so $\eta_1 = 0$, $\mu_{\omega 12} = 0$, $\mu_{\omega 22} = \eta_2 \equiv \eta$, $\lambda_1 = \lambda$, $\lambda_2 = (1 - \lambda)$, $\sigma_{\omega}^2 = \sigma_{\omega j}^2$, $\rho_{\omega j} = \rho_{\omega}$, $j = 1, 2$, while letting $c = 2\sigma_{\omega}^2(1 - \rho_{\omega})$. Then U'_1 simplifies to U_1 (proportion $1 - \lambda$ habituate):

$$U_1 = \frac{\bar{V}_1}{\sqrt{[c + \lambda(1 - \lambda)\eta^2]/n_1}},$$

with

$$E(U_1) = \frac{(1 - \lambda)\eta}{\sqrt{[c + \lambda(1 - \lambda)\eta^2]/n_1}},$$

and U_2 (all infants habituate) simplifies to

$$U_2 = \frac{\bar{V}_2}{\sqrt{c/n_2}},$$

with

$$E(U_2) = \frac{\eta}{\sqrt{c/n_2}}.$$

The power problem is to find the value of n_1 required to maintain the same power for a given $\lambda > 0$ as there would be for sample size n_2 if $\lambda = 0$, and for the associated parameters η , ρ_w , and σ_w^2 . This means that $P(U_2 > t_2) = P(U_1 > t_1)$, where t_2 and t_1 are t -distribution .05 one-tail null

critical values corresponding to their associated degrees of freedom. Note, however, that although referenced to a t distribution, as would be done in practice, U_1 has an unknown distribution.

For Case 2, random samples of $n_2 = 30$ pairs of observations of the form (W_{22}, W_{12}) were obtained from a bivariate normal distribution with correlation $\rho_w = .25$, and common $\sigma_w^2 = 49$ variance. These parameter values were similar to sample values obtained in our empirical study. $E(W_{22} - W_{12}) = \eta$. The difference $V = W_{22} - W_{12}$, with observed values v_1, v_2, \dots, v_{n_2} , formed the data for the paired t test and for varying values of η .

For Case 1, random samples of size n_1 were sampled from two bivariate normal distributions, each with common $\rho_w = .25$ and common marginal variances $\sigma_w^2 = 49$ as in Case 2. With probability λ the observation pair (w_{21}, w_{11}) came from the distribution with difference in marginal means of zero, that is, $\eta = 0$. With probability $1 - \lambda$ the observation pair (w_{22}, w_{12}) came from a distribution with differences in marginal means of η with $\eta > 0$. The corresponding differences v_1, v_2, \dots, v_{n_1} formed the data for the paired t test.

For each fixed set of parameters for each case, there were 5,000 simulations that were used to compute the empirical proportion of the observed u_2 (a value of U_2), which exceeded the t_2 critical value; this estimates $P(U_2 > t_2)$. Similarly, the proportion of the observed u_1 (a value of U_1) that exceed t_1 estimates $P(U_1 > t_1)$. It is a matter of varying n_1

until a match in these empirical proportions is found and then varying λ and η and repeating the process. This is the basis for the curves in Figure 1.

In summary (assuming U_1 and U_2 above), let $c = 2\sigma_w^2(1 - \rho_w)$, where \sqrt{c} is the square root of the expected value of the sample variance in the denominator term in the paired t test. Let η be the expected novel minus familiarity mean difference for habituators, with habituators in the population in proportion $1 - \lambda$. Let n_2 be the number of infants in a study when $1 - \lambda = 1$, so $\lambda = 0$ and all infants in the population habituate. Then n (denoted n_1 above) is the number of infants needed when the proportion of nonhabitulators is $\lambda > 0$ (with mean novel minus familiarity expected difference of zero for nonhabitulators) to maintain the same power, whatever that power might be, relative to the power achieved for a given n_2 with $\lambda = 0$ for $\eta > 0$. An approximation to the number of infants required, n , is given by

$$n \approx n_2 \frac{c + \lambda(1 - \lambda)\eta^2}{c(1 - \lambda)^2}. \quad (\text{A1})$$

With $\eta = 1$, $\lambda = .3$, $c = 73.5 = 2 \times 49(1 - .25)$, $n_2 = 30$, then $n = 62$, whereas if $\eta = 4$, then $n = 64$. In Figure 1 the required sample size to match the power with $\lambda = .3$, $c = 73.5$, $n_2 = 30$ would be about 60 for $\eta = 1$ or $\eta = 4$. Equation A1 also reveals why η does not strongly influence n if c is relatively large. In Figure 1 with $\lambda = 0$, $n_2 = n = 30$, $\eta = 1$, the power is .156; if $\eta = 4$, the power is .802.

Appendix B

Parameter Estimation

This Appendix describes the parameter estimation procedures under Model 3, which can be written with subscript i suppressed and in somewhat simplified notation as follows:

$$E[Y(t)] = h^*(t; \alpha, \beta, \delta) = \beta \exp[-\delta(t - 1)^2] + \alpha,$$

where observations are made on $Y(t)$. The nonlinear optimization problem requires numerical methods with starting values specified for the values of the parameters. Let F be defined as follows:

$$F = \sum_t [y(t) - h^*(t; \alpha, \beta, \delta)]^2; \quad (\text{B1})$$

the desire is to find least squares estimates of the parameters based on F . However, in order to ensure that the estimates are nonnegative, given the uncertainty of good starting values, first estimate parameters under the model

$$E[Y(t)] = h^\circ(t; \alpha_\circ, \beta_\circ, \delta_\circ) = \beta_\circ^2 \exp[\delta_\circ^2(t - 1)^2] + \alpha_\circ^2$$

with

$$F^\circ = \sum_t [y(t) - h^\circ(t; \alpha_\circ, \beta_\circ, \delta_\circ)]^2. \quad (\text{B2})$$

Once estimates $\hat{\alpha}_\circ$, $\hat{\beta}_\circ$, and $\hat{\delta}_\circ$ are obtained, set $\hat{\alpha}_\circ^2 = \hat{\alpha}$, $\hat{\beta}_\circ^2 = \hat{\beta}$, and $\hat{\delta}_\circ^2 = \hat{\delta}$, which are the estimates that satisfy an identical parallel routine defined for F . Thus, if we focus on B2,

$$\mathbf{g} = \left(\frac{\partial F^\circ}{\partial \alpha_\circ}, \frac{\partial F^\circ}{\partial \beta_\circ}, \frac{\partial F^\circ}{\partial \delta_\circ} \right)',$$

the gradient vector. The Hessian matrix is

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 F^\circ}{\partial \alpha_\circ \partial \alpha_\circ} & \frac{\partial^2 F^\circ}{\partial \alpha_\circ \partial \beta_\circ} & \frac{\partial^2 F^\circ}{\partial \alpha_\circ \partial \delta_\circ} \\ \frac{\partial^2 F^\circ}{\partial \alpha_\circ \partial \beta_\circ} & \frac{\partial^2 F^\circ}{\partial \beta_\circ \partial \beta_\circ} & \frac{\partial^2 F^\circ}{\partial \beta_\circ \partial \delta_\circ} \\ \frac{\partial^2 F^\circ}{\partial \alpha_\circ \partial \delta_\circ} & \frac{\partial^2 F^\circ}{\partial \beta_\circ \partial \delta_\circ} & \frac{\partial^2 F^\circ}{\partial \delta_\circ \partial \delta_\circ} \end{pmatrix}.$$

Let $\boldsymbol{\theta} = (\alpha_o, \beta_o, \delta_o)'$, then the basic iterative expression is (Seber & Wild, 1989, Equation 13.25)

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \rho^{(j)} (\mathbf{H}^{(j)} + \nu^{(j)} \mathbf{I})^{-1} \mathbf{g}^{(j)}, \quad (\text{B3})$$

where $0 < \rho \leq 1$, $\nu > \max(0, \text{minus the smallest eigenvalue of } \mathbf{H}^{(j)})$, and \mathbf{I} is a 3×3 identity matrix.

The algorithm begins with starting values of the parameters in $\boldsymbol{\theta}^{(j=1)}$ and where $\mathbf{g}^{(j)}$ and $\mathbf{H}^{(j)}$ denote values of \mathbf{g} and \mathbf{H} evaluated at the value of their derivatives on iteration j , and with $\rho^{(j)}$ and $\nu^{(j)}$ denoting possible dependence on j . Equation B3 is iterated until convergence or until some criterion has been achieved. One can set ρ by line search procedures (Gill, Murray, & Wright, 1981, p. 100; Seber & Wild, 1989, p. 594). The additive $\nu^{(j)} \mathbf{I}$ ensures that the matrix is positive definite. The variance-covariance matrix (Seber & Wild, 1989, pp. 24–26) of the estimates is given by $2\sigma_E^2 \mathbf{H}^{-1}$, where σ_E^2 is the variance of the error associated with the model of Equation B2. Note, however, that a cor-

responding algorithm based on F in Equation B1 must also be given, which differs from the above only in the gradient vector \mathbf{g} and Hessian matrix \mathbf{H} if asymptotic standard errors are to be given under Equation 1 (or Equation B1). With starting values specified by the squares of the estimates under Equation B2, this algorithm can be iterated once to obtain the corresponding \mathbf{H}^{-1} , based on F . The derivative expressions can be obtained using mathematical software.

Starting values for the algorithm are based on a direct search of the three-space defined by values of $\boldsymbol{\theta}$ choosing that triple of values as starting values the values of $\boldsymbol{\theta}$ which minimize Equation B2. The range of values for α_o^2 and β_o^2 is approximately specified by each infant's data. Thus, α_o^2 is in the range from zero to roughly the average of the infant's smallest observed response, whereas the interval for β_o^2 is roughly the interval defined by the range of the infant's responses. The value of δ_o^2 is typically in (0, 16).

Appendix C

The g Statistic

The theory behind g , Equation 5, is developed in this Appendix, except for a slight simplification in which there is one novel and one familiarity trial to compare rather than the average of two novel and two familiar stimuli as in Equation 5.

Let $D_i = N_i - F_i$ denote the difference in responses to a single novel and familiar pair of stimuli for i . $\eta_i = E(D_i)$, $\sigma_{Ei}^2 = \text{var}(N_i) = \text{var}(F_i)$, and $\text{corr}(F_i, N_i) = \rho$.

Case i: D_i Independent but Not Identically Distributed

Then $\text{var}(D_i) = 2\sigma_{Ei}^2(1 - \rho)$. $n^{-1} \sum_{i=1}^n D_i = \bar{D}$, $\text{var}(\bar{D}) = [2(1 - \rho) \sum_{i=1}^n \sigma_{Ei}^2]/n^2$. Define

$$G_1 = \frac{\bar{D}}{\sqrt{[2(1 - \rho) \sum_{i=1}^n \sigma_{Ei}^2]/n^2}}.$$

Under the hypothesis that $E(\bar{D}) = 0$, G_1 will have an asymptotically standard normal distribution, as n increases, by the Liapounov central limit theorem (Lehmann, 1998, p. 97).

Replacing parameters with their estimates in G_1 $\hat{\sigma}_{Ei}^2$ is the estimated residual variance for i 's best fitting model; $\hat{\rho}$ is the correlation among the observed pairs (f_i, n_i) , $i = 1, 2, \dots, n$. Tests and confidence intervals can be constructed using bootstrap procedures (Efron & Tibshirani, 1993) or by using conventional procedures.

The conventional approach is to compute a sample variance of the differences and insert the sample standard error in the denominator: $S_D^2 = \sum (D_i - \bar{D})^2 / (n - 1)$. Define $T_{(i)} = D_i / \sqrt{S_D^2/n}$.

$E(S_D^2/n \text{ Case i}) =$

$$\frac{2(n-1) \sum_i \sigma_{Ei}^2 (1 - \rho) + \sum_{i < j} \sum_{j=2}^n (\eta_i - \eta_j)^2}{n^2(n-1)}. \quad (\text{C1})$$

Let ω_n denote the average of the $\binom{n}{2}$ quadratic terms; then, from Equation C1,

$$E(S_D^2/n \text{ Case i}) = 2(1 - \rho) \sum_{i=1}^n \sigma_{Ei}^2/n + \omega_n/2.$$

Using the sample estimate $d_i = \hat{\eta}_i$ a bias-corrected estimate of the quadratic term based on data from 35 infants in our study is

$$\sum_{i < j} \sum_j [(\hat{\eta}_i - \hat{\eta}_j)^2 - 2(1 - \hat{\rho})(\hat{\sigma}_{Ei}^2 + \hat{\sigma}_{Ej}^2)] = 37,742,$$

with $\hat{\omega}_{35} = 63.41$. Note that

$$[2(1 - \rho) \sum_{i=1}^n \sigma_{Ei}^2]/n^2 < E(S_D^2/n \text{ Case i}).$$

Thus, in general, the denominator of $T_{(i)}$ will tend to be too

large and a null test based on $T_{(i)}$ will tend to have low power.

Case iid: D_i Independent and Identically Distributed (iid)

Let $T_{(iid)} = \bar{D}/\sqrt{S_D^2/n}$. $E(D_i) = E(\bar{D}) = \eta$, independent of i , $E(S_D^2/n | \text{Case iid}) = 2\sigma_E^2(1 - \rho)/n$, and $T_{(iid)}$ is the paired t test. Given the D_i , $i = 1, 2, \dots, n$ are normal, which

requires that F_i and N_i be bivariate normal in distribution (Johnson & Kotz, 1972, p. 59), then $T_{(iid)}$ is T distributed under the null hypothesis.

Received July 10, 2001

Revision received June 17, 2003

Accepted September 16, 2003 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.