

TouchStat v. 3.00: A new and improved Monte Carlo adjunct for the sequential touching task

WALLACE E. DIXON, JR. AND ROBERT M. PRICE, JR.
East Tennessee State University, Johnson City, Tennessee

MICHAEL WATKINS
Mt. Gilead, Ohio

AND

CHRISTINE BRINK
Medical University of Ohio, Toledo, Ohio

The sequential-touching procedure is employed by researchers studying nonlinguistic categorization in toddlers. TouchStat 3.00 is introduced in this article as an adjunct to the sequential-touching procedure, allowing researchers to compare children's actual touching behavior to what might be expected by chance. Advantages over the Thomas and Dahlin (2000) framework include ease of use, and fewer assumptive limitations. Improvements over TouchStat 1.00 include calculation of chance probabilities for multiple "special cases" and for immediate intercategory alternations. A new feature for calculating mean run length is also included.

An increasingly popular nonlinguistic device for assessing infant and toddler categorical development is the sequential touching task. In the task, an experimenter presents infants and toddlers with arrays of objects from various adult-defined categories, usually four objects from each of two categories, and draws inferences about children's categorical understanding of the objects based on their patterns of touching them. For example, children who touch a lot of objects from a single category, such as "dog," despite the presence of objects from another category such as "horse," may be telling us that they recognize the dog objects as going together, and as being distinct from the horse objects. The problem, of course, lies in determining whether children's dog touching sequences exceed what would be expected by chance, or whether it is better characterized as random touching.

The sequential touching task was popularized in the late 1980s by Jean Mandler and colleagues as a nonlinguistic assessment of conceptual categorization (Mandler & Bauer, 1988; Mandler, Fivush, & Reznick, 1987), but the task ultimately derived from the even earlier work of Ricciuti (1965), and was built upon later by Starkey (1981) and Sugarman (1981). Since Mandler and colleagues' methodological work on the sequential touching elements of the task, popularity of the procedure seems not to have waned, and has been used to explore the extent to which infants and toddlers categorize objects on the basis of their parts (Rakison & Butterworth, 1998; Rakison & Cohen, 1999), their gender (Johnston, Madole, Bittinger,

& Smith, 2001; Levy, 1999), and their significance for semantic understanding (Gopnik & Meltzoff, 1987, 1992; Shore, Dixon, & Bauer, 1995; Smith, Whitmore, Shore, Robinson, & Dixon, 1999). More recently, the task itself has been the subject of study, such as when Oakes and Plumert (2002) examined children's sequential touching strategies as a function of task familiarity.

A popular measure derived from the sequential touching task is *mean run length* (MRL). MRL indexes the average number of sequential intracategorical touches exhibited by a child at a single sitting. MRLs are determined by summing the total number of touches exhibited by a child, and dividing by the number of intracategorical runs of touches exhibited by that child. So, for example, if a child were given a tray containing four toy dogs and four toy horses, and the child touched objects in the pattern dog, dog, dog, horse, dog, dog, dog, horse, horse, horse, horse, dog, dog, the child would have touched 13 total objects, and would have done so in five runs of touches. This child then, would have obtained an overall MRL of $13/5 = 2.60$, incorporating runs to both adult-defined categories. Researchers generally assume that longer mean run lengths indicate greater likelihoods that children are indeed categorizing objects, and not simply touching randomly; however, the strength of this assumption is weakened by the possibility that individual children could be spending a lot of intracategorical time going back and forth between only two of the objects within the category. (Note that in this literature it is conventional to disregard immediately

repeated touches to the same object; therefore calculations of MRLs assume that repeated touches to either the dog or horse category are to different dogs/horses than were the immediately preceding touches.)

To determine whether children are capable of making a particular categorical distinction at a group level (e.g., based on theoretical, methodological, or developmental grounds), individual children are exposed to that categorical distinction, and the group MRL is derived by averaging across individual children's MRL scores. The group MRL is then compared to either another group MRL, depending on the experimental manipulation, or the mean run length that would have been expected had touching been entirely random (i.e., the expected mean run length). According to Mandler et al. (1987), the expected mean run length for children who are touching entirely at random should be 1.75, when they are presented two categories of four objects each. To the extent that the mean run length of the group under consideration significantly exceeds this expected value, children are said to be touching based on something other than chance—which is usually presumed to be information derived from the presented categories.

Although mean run lengths can be used to test for group differences in categorical ability, analyses of mean run lengths have been unhelpful in determining the categorical abilities of individual children. It is not clear, for example, whether an individual child's mean run length of 2.60 is sufficient to classify the child as a categorizer. Thus, researchers have developed a supplemental strategy of analyzing individual children's observed touching behavior with respect to randomly generated touching patterns using computerized Monte Carlo sampling procedures. Using this strategy, the observed touch sequence of a given child is compared to Monte Carlo generated touch sequences of the same length, and the probability of observing categorization-like Monte Carlo touching sequences that meet or exceed the sequence generated by the child is determined. Through this procedure, researchers can derive an estimate of the probability that an individual target child's "special" category-like touching sequence happened through random touching.

As originally suggested by Mandler et al. (1987), special runs are those in which at least three unique objects are touched within a category. So, in reference to the 13-touch dog/horse sequence reproduced above, three runs are potentially special: the first dog sequence, the second dog sequence, and the second horse sequence. Once any special runs within an individual child's touching sequence have been identified, it is simply a matter of having a computer simulate thousands of random touching sequences, equal in length to the child's total touching sequence, and then counting up the number of simulated touch sequences that contained at least the special case(s) generated by the child. If the special touching patterns exhibited by the child were not likely to occur by chance, as compared with the simulated touching sequences generated by the computer, the individual child can be labeled a categorizer.

Mandler et al. (1987) recommend a 0.10 cutoff value for determining when a special touch sequence was not likely to have occurred by chance. So, for example, if a

child touched a total of 30 objects, and within those touches generated a special touching sequence that was observed in fewer than 10% the computer-generated simulations, that child could be regarded as a categorizer. Incorporation of this procedure adds a tool to the sequential-touching tool chest because it provides a means for researchers to explore individual differences in categorization behavior in addition to any group differences detected through MRL analysis.

Although the analytical logic behind the sequential touching task has been in place for a number of years, the latitude with which researchers can apply its logic has been constrained. For example, run length analyses based on an expected mean run length of 1.75 are predicated on presenting children with two categories of four objects each. Expected mean run lengths for other potential sequential-touching contexts—for example, those involving more than two categories, more or fewer than four objects per category, or unbalanced categories in which one category has more or fewer objects than the other(s)—have not been developed. With one possible exception described below, expected mean run lengths in these contexts cannot easily be described by a general purpose mathematical equation. Moreover, existing look-up tables and Monte Carlo programs, including TouchStat 1.00 (Dixon, Woodard, & Merry, 1998), make it difficult or cumbersome for users to determine *joint probabilities* associated with a child exhibiting two or more special touch sequences to two or more categories, or to a single category multiple times.

Finally, to our knowledge no look-up tables or software codes exist which would allow for the assessment of chance probabilities associated with *immediate intercategory alternations* (IIAs). IIAs occur when objects from contrasting categories are presented to children, who then proceed to touch objects from each of the two categories alternately, such as dog, horse, dog, horse, dog, horse, dog, horse. This pattern of touching behavior can arguably be regarded as a form of categorization because in order to accomplish it, children must recognize when the next object is systematically *different* than the most recently touched object. In terms of measurement, IIAs would be evident in cases where children's mean run lengths fall significantly short of the expected mean run length (e.g., of 1.75 in the two group, four exemplar situation). Indeed, Oakes and Plumert (2002) have suggested that this form of categorization may increase in frequency as a function of task familiarity. In any event, although researchers can conduct mean run length analyses to determine if a group of children are exhibiting mean run lengths that fall significantly below chance expectations, no easily accessible procedures exist for determining when and whether individual children can be classified as intercategory alternators.

The Thomas and Dahlin (2000) Approach

The only published procedures for dealing with sequential touching data besides Mandler et al., (1987) and Dixon, et al. (1998), are those of Thomas and Dahlin (2000). Because of its complexity, we deem it useful to review their approach here; although it is beyond the scope of the present manuscript to provide a comprehensive review. In their approach, Thomas and Dahlin assume that indi-

vidual children are either categorizers or noncategorizers with respect to a given set of objects within a given set of categorical contrasts at a given point in time. Their model for identifying children as categorizers or noncategorizers is best characterized as a mixture of shifted negative binomial distributions. They propose that the sequence of touches for children classified as categorizers will follow one probability distribution, whereas the sequence of touches for children classified as noncategorizers will follow a different probability distribution, and that these distributions will vary as a function of the number and size of the categories, and any number of factors associated with the nature of the sample of children.

Thomas and Dahlin's (2000) model for the noncategorizer assumes that a child's sequence of touches follows a shifted negative binomial distribution with a specified probability for selecting an item from an opposing set—what they call a “switching parameter.” For example, if there are two categories of four objects, then the probability of selecting an item from the opposing category on any given touch can be specified as $4/7$ (because there are four objects outside the current category, and only three other objects within the current category that could be touched next in the sequence). Once a child touches an object outside a given category, the current intracategorical run sequence has ended. Hence, the switching parameter for noncategorizers is known, and is a function of the number and size of the categories.

The touch sequence for the categorizing child is also a shifted negative binomial distribution. However, as noted by Thomas and Dahlin (2000), it is not possible to specify the switching parameter probability for categorizers, because it is not possible to determine a priori when a given intracategorical run sequence has *not* ended. In other words, there are a myriad of ways for how an intracategorical run sequence might *not* end, and so this value must be estimated from the data. This latter, estimable probability is assumed to be constant across all children within the sample, but in all cases is, by definition, less than the probability parameter for the noncategorizers (e.g., $4/7$ in the two category, four exemplar case). Should one have reason to reject this constancy assumption, a probability distribution for the switching parameter must then also be specified.

In order to achieve the final goal of classifying children as categorizers and noncategorizers for specific categorical contrasts, an a priori estimate of the *proportion* of categorizers and noncategorizers for a given sample of children must also be conjectured. Hence, there are two parameters that need to be estimated for each mixed distribution to infer whether a given child within a given sample is a categorizer or noncategorizer for a given category: the switching parameter for categorizers of a given category and the initial estimate of the proportion of children in the sample who are categorizers. The proportion of children in the sample who are not categorizers does not need to be estimated because it is known once its complementary proportion is known [i.e., $p(\text{noncategorizers}) = 1 - p(\text{categorizers})$].

Thomas and Dahlin (2000) propose an iterative procedure for deriving the two estimates. After all unknown pa-

rameters have been iteratively estimated, one next obtains an estimated posterior probability that a target child could be a categorizer or noncategorizer for a given category, and this estimate is determined as a function of both the child's mean run length and the child's number of touches with respect to the given category. This posterior probability, when applied to individual children's sequential touching behaviors, allows the researcher to classify each child as a categorizer or noncategorizer with respect to the specific sequential touching context.

Thomas and Dahlin's (2000) approach generalizes to a wide range of cases in which it is potentially desirable to classify children as either categorizers or noncategorizers. It not only can be applied to the two category, four exemplar case, which is the one illustrated in their article and the most typical in the sequential touching literature, but it can be generalized to any number of atypical sequential touching tasks—including those with more than two categories and those with more or less than four objects per category. It can even be extended to determine whether children can be classified as categorizers by virtue of their immediate intercategory alternations (IIAs). However, as the complexity of the categorization to be modeled increases, the number of parameters to be estimated also increases, and presumably the sample size needed to estimate the various parameters must be increased commensurately.

Unfortunately, there are significant drawbacks for those wishing to incorporate the Thomas and Dahlin (2000) approach. Most obvious is its computational complexity. For example, applying the model to the typical two category sequential touching task requires the user to identify probability parameters associated with identifying children as categorizers of Category A, noncategorizers of Category A, categorizers of Category B, and noncategorizers of Category B. Moreover, the user has to use previously obtained run length data in order to calculate a priori estimates of the proportion of each sample that corresponds to each of these four groups present in the sample. Then, the proportion of the sample estimated to be comprised by each of the four groups, as well as the switching parameter probabilities for categorizers, has to be determined through the application of iterative formulae published both in the main body of the manuscript as well as in their appendices. Even the most cursory glance at the mathematical complexity of their formulae is sure to startle the faint of heart. In any case, the level of mathematical expertise required to implement the Thomas and Dahlin procedure probably exceeds the mathematical sophistication of the typical end user. In the absence of user-friendly interfaces which incorporate their negative binomial mixture model, alternative approaches will need to continue to be developed.

But there are other limitations of their approach as well. First, because the entire approach is based on mean run length data, it suffers from the same limitations as other mean-run-length based approaches. In particular, it fails to distinguish between children who “rack up” long mean run length scores by touching only two intracategory objects over and over, from children who earn long mean run length scores by touching all or almost all of the objects within the category. Surprisingly, Thomas and

Dahlin (2000) disregard the importance of this distinction, despite themselves citing evidence attesting to the possibility that the latter pattern of sequential touching is indicative of more mature categorical behavior. Monte Carlo procedures of the sort advocated by Mandler et al. (1987), at least allow the user to specify requirements in which a minimum number of intracategory objects be touched within the intracategorical sequential touching run. When users are able to require that all or nearly all the objects within a category be touched, they can reduce the likelihood of overattributing categorization competence to children who may have simply detected two objects within a given category that they find interesting.

Second, Thomas and Dahlin's (2000) approach requires users to have knowledge of the behaviors of an entire sample of children in order to classify individual children. Specifically, the user must estimate the proportion of categorizers and noncategorizers in a given sample for a given categorical distinction before being able to determine the likelihood that an individual child is a categorizer or noncategorizer. It strikes us that such an approach might have limited utility when circumstances are such that the categorical status of an individual child needs to be determined. Relatedly, the fact that children get classified as categorizers or noncategorizers based on their performance in the context of the entire sample, leaves open the possibility that children could be classified differently if evaluated in the context of different samples. That is, because their classification scheme is based on probability distributions, a child who in one sample is classified as a categorizer of Category A, could, with some probability, be classified as a noncategorizer of Category A in a different sample.

Finally, Thomas and Dahlin's (2000) approach presently lacks published evidence of its convergent validity. Designers of the sequential touching task intended the procedure to be a nonlinguistic tool that would nevertheless converge with linguistic measures of children's categorization. Although researchers have produced convergent validity data using the standard procedures Thomas and Dahlin oppose (e.g., Shore et al., 1995; Smith et al., 1999), Thomas and Dahlin themselves fail to present convergent data in support of their approach. In addition, no published studies have to date incorporated the Thomas and Dahlin approach.

TOUCHSTAT VERSION 3.00

The purpose of the present article is to introduce TouchStat 3.00, a freeware program for calculating expected mean run lengths and for determining chance probabilities associated with categorically based touching sequences exhibited by young children in the sequential touching task. Version 3.00 improves significantly over version 1.00 (Dixon et al., 1998) because (1) it exists as an executable file which functions within a windows environment (TouchStat 1.00, written in C, required compiling prior to running), (2) it allows users to determine chance probabilities associated with both single and multiple intracategory touch sequences, (3) it allows users to determine chance probabilities associated

with immediate intercategory alternations, and (4) it adds functionality by incorporating a feature for calculating mean run lengths directly from observed data (which can be cut and pasted from a text file or manually entered directly into the program). TouchStat 3.00 is available for download from the International Society on Infant Studies website, isisweb.org/touch_stat_pgm.htm, and runs on any Windows-based machine updated with the .Net framework. Implementation of TouchStat 3.00 is based on the premise that sequential touching is congruent with a permutation testing model of statistical inference which can be modeled by sampled permutations derived from Monte Carlo procedures (see Dixon et al., 1998, for a fuller explanation).

Running the Program

The user interface for TouchStat 3.00 can be seen in Figures 1–3. Figure 1 illustrates a sample run for a user interested in determining chance probabilities associated with special case intracategory touching. Figure 2 illustrates a sample run for a user interested in determining chance probabilities associated with immediate intercategory alternations. And Figure 3 illustrates a sample run for a user interested in calculating the mean run length for a given child's observed sequential touching behavior. Each of these applications is described in more detail below.

Setting up TouchStat 3.00 to generate Monte Carlo samples. As noted above, users of the sequential touching procedure are usually interested in determining whether the sequential touching behavior of a given child is indicative of the categorical understanding of that child with respect to the presented categories. As first noted by Mandler et al. (1987), making this determination requires that the child's observed touching pattern be compared in some way to touching patterns that would be expected by chance, and then determining whether the child's observed touching pattern is typical of randomly generated touch sequences, or differs from them in some systematic way. The purpose of the Monte Carlo approach is to generate a large number of simulated touching patterns which assume random, independent touches, and which can then be used as points of comparison for the observed touching pattern.

Of course, to determine the chance probability associated with a given observed touching sequence, the user must first specify several values to ensure that the Monte Carlo program is sampling in a way that parallels as much as possible the real-life situation that is being modeled. The user must tell the computer to simulate touch sequences that are structurally similar to, and constrained by the same boundaries as, the touches generated by the child. If the child were given 8 objects to touch, then the computer must also be given 8 objects to touch, at least virtually. If the child touched the 8 objects 30 times, then the computer must be told to touch the 8 objects 30 times virtually. The only thing the computer is not told, is which specific objects to touch. Rather, it is simply told to touch randomly. Of course, it is insufficient to have the computer generate only a single simulated touching sequence, because a single sample of a simulated touch sequence does a poor job at representing the sampling frame of all

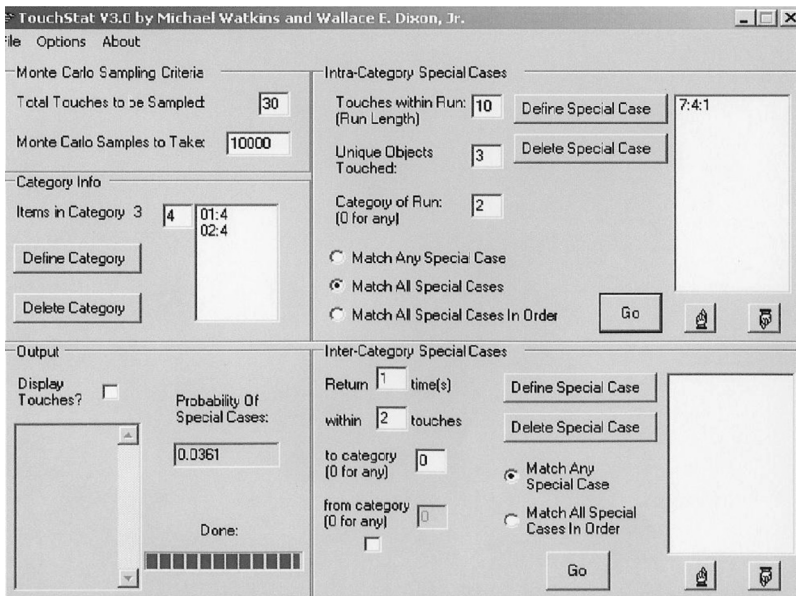


Figure 1. Image of TouchStat 3.00 interface showing intracategory Monte Carlo simulation.

possible touch permutations. Rather, 10,000, 100,000, or even 1,000,000 samples of simulated touching patterns should be generated to represent as close as possible the entire sampling frame.

When the child's touching pattern is compared to the Monte Carlo generated touch simulations, the researcher looks to determine the extent to which certain special and a priori -determined features of the child's touch sequence are represented among the simulated, randomly generated touching samples. Suppose for example that a

child engaged in 10 touches, and that 9 of these touches were to objects from a single category. Suppose further that among these 9 intracategorical touches were touches to each of the four objects in the category. In this case, the researcher wishes to know whether touch sequences like the child's are typical or representative of random touching behavior, or whether touch sequences like the child's are unusual and not very likely to occur as a result of random touching behavior. Once a child's touch sequence is deemed to be not very likely the result of random touch-

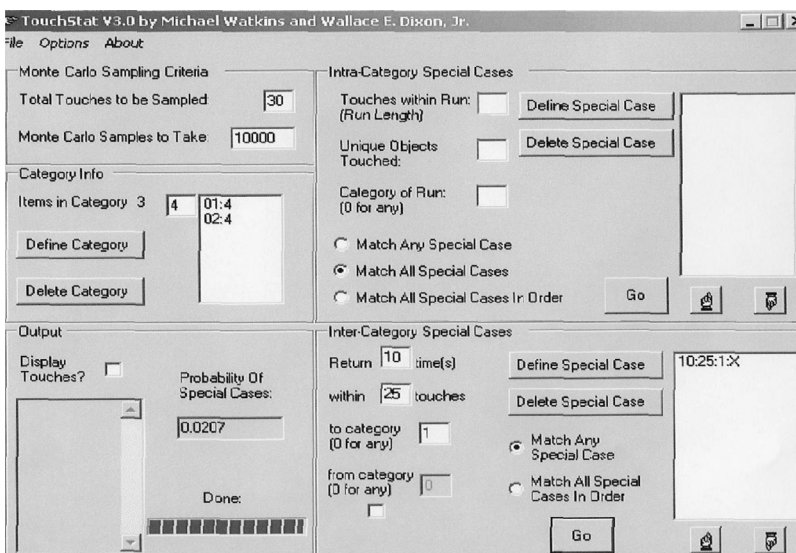


Figure 2. Image of TouchStat 3.00 interface showing intercategory Monte Carlo simulation.

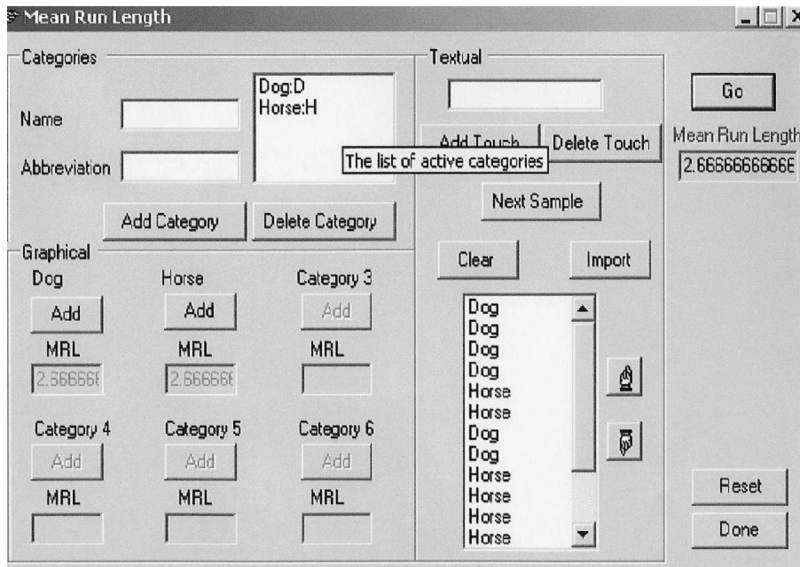


Figure 3. Image of TouchStat 3.00 interface showing mean run length calculation.

ing, it is conventional to assume that the child's touching was not random, and that it was instead due to his or her recognition of the presence of categories.

Intracategorical Touching

When using TouchStat 3.00 to determine chance probabilities associated with one or more special *intracategorical* touching sequences exhibited by a child, the researcher specifies a number of settings within the "IntraCategory Special Cases" module. The purpose of this step is to delimit the virtual sampling context within TouchStat 3.00 so that it parallels structurally the actual sampling context presented to the child. If the child were presented 8 objects comprising two categories, TouchStat 3.00 needs to be virtually presented with 8 objects comprising two categories. As well, in order for TouchStat 3.00 to simulate random touching with parameters that correspond to the child's actual touching, TouchStat 3.00 needs to be informed about how many objects were touched by the child. If the child engaged in 30 total touches, then TouchStat 3.00 should be told to simulate strings of 30 touches. Once these settings have been specified, the researcher is free to have TouchStat 3.00 generate as many simulated touching sequences as she desires. We would recommend taking 100,000 Monte Carlo samples, because it would be a reasonably representative sample of all possible touch permutations without drawing extensively on microprocessor resources.

After TouchStat 3.00 generates its touch simulations, it then scans them for instances of the special touching behavior of the target child, as defined by the researcher. That is, the researcher specifies the kinds of touching patterns she determines to be special, and which TouchStat 3.00 should look for. If the target child generated a run of 9 intracategorical touches, within 10 total touches, the researcher can have TouchStat 3.00 generate 100,000 samples of 10-touch simulations, and then have it count how many times a run of at least 9 intracategorical touches occurred. If the

target child touched all four exemplars within the category within his 9 intracategory touches, the researcher can further specify that TouchStat 3.00 look only for those cases in which 9 intracategory virtual touches happened *and* in which all 4 different objects within the category were touched. Thus, the intracategorical run length, the number of unique objects touched within the intracategorical run, and the category within which the intracategorical run took place can be specified in this module. As in the literature, an intracategorical run of touches, or more simply *a run*, is defined here as a touch sequence to objects within a single category; *unique objects touched* is defined as the number of different objects touched within a category during the run; and *category of run* is defined as the specific category in which the run took place.

Figure 1 illustrates the chance probability of a touch sequence in which, out of a total of 30 touches, a child touched 7 objects in a row from Category 1, and within those 7 touches, touched 4 different objects. The chance probability associated with *at least* this touch sequence turned out to be $p = .0361$. In other words, a touching event this extreme was found to take place by chance 361 times out of the 10,000 Monte Carlo 30-touch sequences modeled by the user. By Mandler et al.'s (1987) criteria of $p < .10$, this touch sequence is not very likely to have happened by chance, which permits the user to conclude that the child exhibiting the modeled touch sequence is a categorizer.

Figure 1 also shows that the user is in the process of defining a second special case, in this instance with a run of 10 touches to items from Category 2, with at least three unique Category 2 objects being touched. The user need only press the "Define Special Case" button, and then the "Go" button, and then TouchStat 3.00 would generate another Monte Carlo sampling run consisting of 10,000 samples of 30 simulated touches each, and would count up the number of simulated touch sequences in which at least *both* of the special cases occurred.

Intercategorical Touching

Initial applications of the sequential touching procedure linked categorization behaviors only to intracategorical touching. Recently researchers have argued that *intercategorical* touching, what we have been calling immediate intercategory alternations (or IIAs), also represents a type of categorization because engaging in this behavior requires children to recognize systematic similarities and differences between objects from the two categories (Oakes & Plumert, 2002; Thomas & Dahlin, 2000). TouchStat 3.00 also permits the researcher to determine chance probabilities associated with these IIAs. To do so, she must define three parameters from within the “Inter-Category Special Cases” module (note that the input format is different from that of the intracategory module). For IIAs, the user would again have a special touch string in mind in which the target child appeared to be exhibiting immediately alternating touches between the two categories. But this time the user specifies how many times the child returned to a particular category (in the “*return x times*” parameter), and also specifies the number of touches within which the child returned to that category after having left it (in the “*within x touches*” parameter). Finally, the user specifies the category to which the child is continually returning in the “*to category x*” parameter.

Figure 2 illustrates the chance probability of a touch sequence in which, out of a total of 30 touches, a child returned 10 times to Category 1 within a run of 25 touches. The chance probability of at least this event was found to be .0207, given a Monte Carlo sampling frame of 10,000 samples of 30-touch sequences. This pattern of touching could be taken as evidence that the child is making a systematic distinction between the two categories of objects, even if the form of the distinction is different than the child who categorizes through intracategorical touching.¹

Using TouchStat 3.00 to Calculate Mean Run Lengths

TouchStat 3.00 has several additional features of potential interest to users of the sequential touching task. Perhaps most useful is the “Mean Run Length” module (accessible through the File pull-down menu) which allows users to calculate MRLs (see Figure 3). Users have the option of either entering an individual child’s touch sequence through the data entry interface or by cutting and pasting text from a preexisting data file. For manual entry, touches can be entered either by defining and pressing virtual buttons, or by entering data textually. TouchStat calculates both total MRL across all categories and individual MRLs for each of the categories.

To calculate mean run lengths of a target child’s touching sequence, the researcher brings up the MRL module, and begins by defining the number of categories of interest that were presented to the child. In a typical two-category, four-exemplar case involving dogs and horses, for example, the user would first type in a descriptive label such as “horse” in the *Name* slot, and then an abbreviation such as “h” in the *Abbreviation* slot. When “Add Category” is pressed, the TouchStat 3.00 MRL module knows that at least one category exists, and demonstrates that knowledge by des-

ignating a button as the *horse* button. The same procedure would be followed to establish the “dog” category, and to designate a dog button. Note that the number of items in each category is irrelevant here, since TouchStat does not need to know the number of exemplars in a category to determine mean run lengths in that category. TouchStat allows up to six categories to be defined within this module.

Once categories have been defined, the “Add” button for each defined category becomes enabled. Users can then begin entering sequential touching data by either pressing the *Add* button associated with a category, or typing textual data into the “Textual” box and pressing the “Add Touch” button. Additionally, properly formatted text can be copied and pasted directly into the MRL module data window. After all touches have been entered, pressing the “Go” button produces mean run lengths separately for each defined category and overall across all categories.

Expected run lengths for a variety of categorical contrasts can also be estimated from within TouchStat 3.00, using the *Import* feature contained within the MRL module. Being able to estimate expected MRLs within TouchStat 3.00 turns out to be quite advantageous, especially to the extent that the original formula for determining expected mean run length, as published in Mandler et al. (1987) is wrong (also noted by Thomas & Dahlin, 2000). Mandler et al. indicate that the expected run length for two categories of four objects each is 1.75. Their formula, attributed to Jeffrey O. Miller, is

$$E [\text{run length}] = \frac{n}{n-1} \times \frac{n-1}{2n-1} \frac{1}{1 - \left(\frac{n-1}{2n-1}\right)^2}. \quad (1)$$

However, this equation produces a value of 0.70 when $n = 4$ (i.e., $n =$ number of objects per group), which is less than the shortest possible run length of 1.00. The correct formula² for the expected run length is much simpler,

$$E [\text{run length}] = 2 - \frac{1}{n}, \quad (2)$$

which, because it produces the 1.75 value described by Mandler et al., leads us to believe that the formula published in their article was the result of a printing error (a conclusion also reached by Thomas & Dahlin, 2000). Nevertheless, even this simplified formula holds true only for the case in which there are two categories represented and the number of items is equivalent across the categories. Should researchers be interested in calculating expected run lengths in broader applications, such as when there are more than two categories or when the number of items differs across categories, the expected run length can still be determined through the Monte Carlo sampling procedures employed by TouchStat 3.00.

Determining expected run lengths using TouchStat 3.00 requires generating samples from within the Monte Carlo sampling module, importing the resulting data into the mean run length module, and calculating MRLs as usual. The resulting MRLs would represent MRLs obtained through random touching obtained through Monte Carlo generated samples, and so represent expected mean run lengths. These

expected run lengths, then, would be the point against which children's actual MRLs could be compared and evaluated.

Suppose, for example, that we were interested in testing the Mandler et al. (1987) claim that the expected run length in the two-category, four-item case *really is* 1.75. In this instance, we could generate, say, 5,000 samples of 100 touches each, resulting in a half million total touches. We could then determine the MRL within this half-million, random-touch sequence, and in this instance, the MRL *becomes the expected run length* against which the observed touching patterns of target children can be compared. To simulate this event in TouchStat 3.00, we would define two categories containing four items each, we would check the *Display Touches* box so that the simulated touches are output to a temporary file, and we would define at least one *Special Case*.³ When the simulation runs, TouchStat prints the details of the 500,000 touches to the *Display Touches?* output window (which may take several moments). The user then opens the MRL module from within the *File* pull-down menu, and from within the MRL module, clicks on the *Import* button. At this point, TouchStat 3.00 transfers the 500,000 randomly simulated touches from the Monte Carlo generation module to the MRL module. Then, the user simply clicks the *Go* button to have MRLs calculated separately for Category 1, Category 2, and overall.

Having simulated this event ourselves, we determined that Mandler et al. were accurate in their formulations. We found that the Monte Carlo based MRLs were indeed 1.75; within Category 1, within Category 2, and overall, when there were two categories of 4 objects each represented. We also determined MRLs to be 1.80, 1.80, and 1.80, respectively, when the number of objects per category was 5; and 1.83, 1.83, 1.83, respectively when the number of objects was 6. All of these values are consistent with what would be expected from Equation 2 as well. The important point here is that TouchStat 3.00 can be used to determine expected run lengths for as yet unexplored categorical arrangements.

DISCUSSION

We developed TouchStat 3.00 in an effort to assist researchers whose use of the sequential touching procedure promotes our understanding of the development of nonverbal categorization behavior in children. Although much work has yet to be done toward the development of even newer and better methodologies for assessing children's categorical development, we hope that by providing TouchStat 3.00 to the developmental research community, we are adding to the developmental researcher's arsenal of tools, and that features built into TouchStat 3.00 may stimulate researchers to expand the sequential touching procedure beyond its traditional boundaries. Because innovations included in TouchStat 3.00 were based on suggestions made by users of previous versions, it represents an improvement over earlier versions. However, we hope to continually improve TouchStat, and so invite researchers to continue sending us their ideas for improvement.

AUTHOR NOTE

Portions of this article were previously presented at the Biennial Meetings of the International Conference on Infant Studies in May, 2004, in Chicago. Special acknowledgement goes to Chris Robinson for his helpful comments on earlier versions of TouchStat v. 3.00. Comments and suggestions for improvement to TouchStat 3.00 are especially welcome. Copies of TouchStat 3.00 can be obtained at no charge via direct download from www.isisweb.org/touch_stat_pgm.htm. Correspondence concerning this article can be sent to W. E. Dixon, Jr., Department of Psychology, P.O. Box 70649, East Tennessee State University, Johnson City, TN 37614 (e-mail: dixonw@etsu.edu).

REFERENCES

- DIXON, W. E., JR., WOODARD, T., & MERRY, M. S. (1998). TouchStat: A Monte Carlo program for calculating sequential touching probabilities. *Behavior Research Methods, Instruments, & Computers*, **30**, 592-604.
- GOPNIK, A., & MELTZOFF, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, **58**, 1523-1531.
- GOPNIK, A., & MELTZOFF, A. N. (1992). Categorization and naming: Basic-level sorting in eighteen-month-olds and its relation to language. *Child Development*, **63**, 1091-1103.
- JOHNSTON, K. E., MADOLE, K. I., BITTINGER, K., & SMITH, A. (2001). Developmental changes in infants' and toddlers' attention to gender categories. *Merrill-Palmer Quarterly*, **47**, 563-584.
- LEVY, G. D. (1999). Gender-typed and non-gender-typed category awareness in toddlers. *Sex Roles*, **41**, 851-873.
- MANDLER, J. M., & BAUER, P. J. (1988). The cradle of categorization: Is the basic level basic? *Cognitive Development*, **3**, 247-264.
- MANDLER, J. M., FIVUSH, R., & REZNICK, J. S. (1987). The development of contextual categories. *Cognitive Development*, **2**, 339-354.
- OAKES, L. M., & PLUMERT, J. M. (2002). Variability in thirteen-month-old infants' touching patterns in the sequential-touching task. *Infant Behavior & Development*, **25**, 529-549.
- RAKISON, D. H., & BUTTERWORTH, G. (1998). Infants' use of parts in early categorization. *Developmental Psychology*, **34**, 49-62.
- RAKISON, D. H., & COHEN, L. B. (1999). Infants' use of functional parts in basic-like categorization. *Developmental Science*, **2**, 423-431.
- RICCIUTI, H. N. (1965). Object grouping and selective ordering behavior in infants 12 to 24 months old. *Merrill-Palmer Quarterly*, **11**, 129-148.
- SHORE, C., DIXON, W. E., & BAUER, P. J. (1995). Measures of linguistic and non-linguistic knowledge of objects in the second year. *First Language*, **15**, 189-202.
- SMITH, P. H., WHITMORE, J., SHORE, W. J., ROBINSON, C. W., & DIXON, W. E., JR. (1999). Infants' responses to objects representing levels of word knowledge and categorical information. *Infant Behavior & Development*, **22**, 511-526.
- STARKEY, D. (1981). The origins of concept formation: Object sorting and object preference in early infancy. *Child Development*, **52**, 489-497.
- SUGARMAN, S. (1981). The cognitive basis of classification in very young children: An analysis of object-ordering trends. *Child Development*, **52**, 1172-1178.
- THOMAS, H., & DAHLIN, M. P. (2000). Inferring children's categorizations from sequential touching behaviors: An analytical model. *Psychological Review*, **107**, 182-194.

NOTES

1. Note that we do not take a position on the meaning of immediate intercategory alternations, just that it is possible to determine their chance likelihoods using TouchStat 3.00.

2. Thanks to William B. Grasman, Professor of Mathematics, Heidelberg College, Tiffin, OH, for this equation.

3. TouchStat 3.00 requires that at least one *Special Case* be defined in order to run, but the details of the defined *Special Case* have nothing to do with our present efforts and will have no effect on them.