Psychonomic Bulletin & Review
2006, 13 (4), 636-642

Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion?

FUAT BALCI and CHARLES R. GALLISTEL
Rutgers University, New Brunswick, New Jersey

Meck and Church (1983) estimated a 5:1 scale factor relating the mental magnitudes representing number to the mental magnitudes representing duration. We repeated their experiment with human subjects. We obtained transfer regardless of the objective scaling between the ranges; a 5:1 scaling for number versus duration (measured in seconds) was not necessary. We obtained transfer even when the proportions between the endpoints of the number range were different. We conclude that, at least in human subjects, transfer from a discrimination based on continuous quantity (duration) to a discrimination based on discrete quantity (number) is mediated by the cross-domain comparability of within-domain proportions. The results of our second and third experiments also suggest that the subjects compare a probe with a criterion determined by the range of stimuli tested rather than by trial-specific referents, in accordance with the pseudologistic model of Killeen, Fetterman, and Bizo (1997).

Animals—both human and nonhuman—appear to represent abstract quantities, such as number, duration, distance, and area, by means of noisy mental magnitudes (e.g., Cordes, Gelman, Gallistel, & Whalen, 2001; Meck, Church, & Gibbon, 1985). These may be thought of as brain signals that are causally related to the dimension they represent and that enter into behavior-relevant computations, such as rates (number/duration), densities (area/number), means (sum of values/number of values), and so on. One question that arises from this consideration is the form of the function relating subjective (mental) magnitudes to objective magnitudes within a given dimension: Is it logarithmic (e.g., Dehaene, 2001), a power function (Stevens, 1964), or a scalar function (Gibbon, 1977)? A second question concerns the form of the noise (random variability) in the resulting magnitudes: Is it constant (additive), binomial (proportional to the square root of the magnitude represented), or scalar (proportional to the magnitude itself; see Fetterman & Killeen, 1995)? A third question—the focus of this work—concerns the scale factor relating the magnitudes for one dimension of experience (say, duration) to another (say, number). Knowledge of these scale factors would establish the brain's system of mensuration.

Meck and Church (1983) and Meck et al. (1985) estimated the scale factor relating the mental number scale to the mental duration scale by means of cross-domain transfer experiments, in which rats spontaneously transferred a duration discrimination to a number discrimination. In this article, we revisit their experiments, but with human subjects and with further controls.

They used the bisection method, in which subjects are initially familiarized with two reference values in a given domain (i.e., one short [S] and one long [L] duration). On the unreinforced probe trials, subjects are presented with intermediate values (i.e., $S \leq t_{probe} \leq L$ for temporal discrimination) and make a forced choice between the responses associated with the referents. The psychometric function thus obtained is the frequency of making the response associated with the greater referent, as a function of the value of the probe.

In their transfer experiment, Meck and Church (1983) trained rats to discriminate the signal durations of 2 and 4 sec by pressing two different levers. In a final phase of the experiment, the stimuli on the unreinforced probe trials were sequences of white noise bursts, alternating 1 sec on/1 sec off for 10, 12, 14, 16, or 20 cycles. Because the 19-sec duration of the smallest number of cycles greatly exceeded the longest of the reference durations (4 sec), the rats were assumed to respond on the basis of the mental magnitudes representing the number of bursts rather than on the basis of any measure of sequence duration. The rats did show spontaneous transfer on the probe trials, choosing the "short" response on small-number trials and the "long" response on large-number trials. Meck and Church indicated that, in pilot work, they found that only this range of numbers produced transfer, but, crucially, in light of the results we here report, they did not give data showing this. In their interpretation of the spontaneous transfer, they assumed that, on the transfer trials, subjects compared the mental magnitude representing the number of cycles (a discrete quantity) to the mental magnitudes representing the reference durations (continuous quanti-

This research was supported by NIMH Grant 1 R21 MH63866 to C.R.G. and was approved by Rutgers IRB. The rights of the subjects were protected, and signed consent was received from all subjects. The authors thank Russell Church for his helpful comments on the earlier draft of this manuscript. Correspondence should be addressed to F. Balci, Rutgers Center for Cognitive Science, Psych. Bldg. Addition, Busch Campus, Rutgers University, 152 Frelinghuysen Rd., Piscataway, NJ 08854-8020 (e-mail: fbalci@ruccs.rutgers.edu).

ties). They concluded that the mental magnitude representing a 1-sec duration has the same magnitude as the mental magnitude representing a count of 5.

Our method is similar to theirs. In each of three experiments, we first trained human subjects to discriminate the signal durations of 2 and 4 sec of a visual stimulus. Then, in a transfer phase, the stimuli on the probe trials were sequences of flashed visual stimuli, with the number of flashes varying from trial to trial. In Experiment 1, we varied the range of numbers tested (5–10 or 10–20). In Experiment 2, we varied the proportions between the endpoints of the range of values tested, making the largest number three or four times bigger than the smallest. In these first two experiments, we deliberately placed the human subjects in the same situation as the rat subjects in the Meck and Church (1983) experiment, in that we left it to them to figure out what to do when confronted with these novel probe stimuli. The subjects were asked to give the "most appropriate answer." There was no feedback. In Experiment 3, we generated the flash sequences using a random-rate (Poisson) process, to partially deconfound the duration of a stimulus sequence and the number of flashes it contained. The subjects were explicitly instructed to respond on the basis of number (but there was no other feedback), and we analyzed for an effect of sequence duration on choice.

METHOD

Subjects

The subjects were undergraduate student, graduate student, and postdoctoral volunteers at Rutgers University. There were 4 subjects in Experiment 1, and there were 5 new subjects each in Experiments 2 and 3.

Apparatus

The temporal and numerical parameters, presentation of the stimulus, and recording of the responses were controlled with a Macintosh running OS 9.1. The experiments were written in MATLAB 5.3 using the Psychophysics Toolbox extensions (Brainard, 1997). The stimulus that signaled durations and numbers was a black square presented (flashed) on a white background.

Procedure

Training phase. The subjects were shown one presentation of two signal durations (2.0 or 4.0 sec) on the computer screen. They were told which one was the short signal and which was the long signal. Then, they were presented with five short and five long signals in random order. Subjects were asked to press key "a" to indicate if the signal presented was the short one and key "l" to indicate if it was the long one.

Duration-probe phase. In addition to the base durations, the subjects were presented with five intermediate signal durations at logarithmically equal intervals (2.2, 2.5, 2.8, 3.2, and 3.6 sec). Each of the seven durations was presented, in a random order, 15 times in Experiments 1 and 2 and 25 times in Experiment 3. The next presentation of a duration did not occur until all the durations had been presented. The subjects were again asked to press key "a" to indicate if the signal duration presented represented the short duration or to press key "l" if the signal duration presented represented the long duration. The subjects were not given feedback about their performance.

Number-probe phase. Instead of continuous signals, the subjects were presented with the flashing black square. In Experiments 1 and 2, it flashed on for 0.75 sec and off for 0.75 sec. The number of flashes varied from trial to trial within subjects, and the range of these numbers varied across subjects. In Experiment 1, numbers tested were 5, 6, 7, 8, and 10 flashes for 2 subjects and 10, 12, 14, 16, and 20 flashes for the other 2 subjects. In Experiment 2, the numbers tested were 5, 6, 7, 8, 10, 12, 14, 16, and 20 flashes for 3 subjects and 5, 6, 7, 8, 10, 11, 12, 13, and 15 flashes for 2 subjects. Each number was presented in random order 15 times. The next presentation of a number did not occur until all the numbers had been presented. In Experiments 1 and 2, the total duration of the sequence with the smallest number of cycles exceeded the longest duration of the two reference durations. At the beginning of this phase, the subjects were asked not to count the signals, were asked to say the word *the* at each presentation of the stimulus (to forestall inadvertent counting), and were asked to pay close attention to the presentations. They were again asked to give the most appropriate answer to the presentations by pressing the "a" and the "l" keys.

In Experiments 1 and 2, the subjects were not instructed to respond on the basis of number. In Experiment 3, they were, and the duration of the flashes and interflash durations were randomly drawn from an exponential distribution with an expected mean of 750 msec. The shortest of these was 75 msec, and the longest was 2,000 msec. Because of this maximally random variation in the durations of flashes and interflash intervals, the presentation duration for one number could be shorter than, the same as, or longer than the presentation duration of other larger or smaller numbers. In this final experiment, the numbers tested were 5, 6, 7, 8, 10, 12, 14, 16, and 20 flashes for 2 subjects and 5, 6, 7, 8, 10, 11, 12, 13, and 15 flashes for 3 subjects. A given number was presented 50–52 times in random order.

RESULTS AND DISCUSSION

The subjects transferred the discrimination from the duration phase to the number phase (Figure 1). Importantly, the transfer occurred regardless of the absolute values that anchored the ends of the twofold range of numbers tested. If transfer depended on a direct comparison of magnitudes representing number to magnitudes representing duration, it should occur for at most one range, the range for which the mental magnitudes for number and duration were the same. This was the logic behind the Meck and Church (1983) interpretation, which allowed them to infer the scale factor relating mental magnitudes for number to mental magnitudes for duration. The scale factor determines the range for which transfer will occur.

When transfer occurs regardless of range, it must presumably be mediated not by the direct comparison of mental magnitudes representing different kinds of quantities (number and duration) but rather by the cross-domain similarity of within-domain proportions. On this model, discrimination behavior is based on the proportion that a stimulus bears to reference value(s) established by the range of stimuli tested in the transfer condition. These proportions are unitless magnitudes—they do not represent an amount of something—which suits them to mediate cross-domain transfer.

On the basis of the results of Experiment 1, we assumed that the probability of giving a certain kind of response (say, a "long" response) was determined by a comparison between two subjective quantities (one representing the probe and one a standard of comparison). When one changes the proportions between the endpoints of the tested range, as we did in Experiments 2 and 3, the form

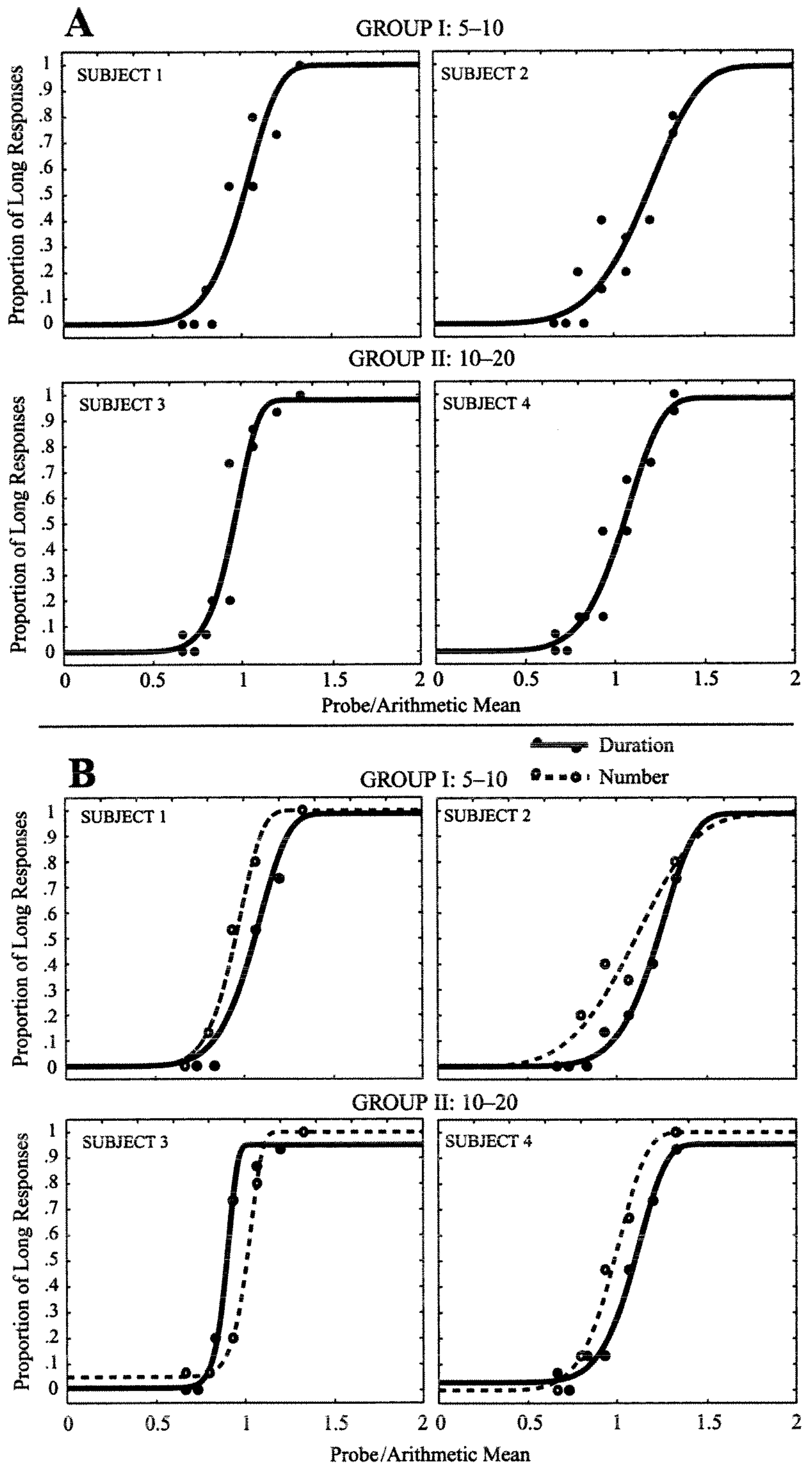

Figure 1. The subject-by-subject psychometric functions for duration and number discrimination in Experiment 1. The *x*-axis is normalized by the midpoint of the range of stimuli tested. The curves fit to the data are Weibull functions. (A) The duration and number data were pooled before fitting the model. (B) We fit the model independently for duration and number data. The solid lines are the psychometric functions for duration discrimination, and the dashed lines are the psychometric functions for number discrimination.

and location of the function obtained in the transfer phase will depend on what the subjects' standard of comparison was. We considered several simple models (Table 1): In the first, the standard of comparison was the arithmetic mean (AM model). In the second, it was the geometric mean (GM model). In the third, we assumed that they compared the difference between the probe and the minimum point of the range with the difference between the minimum and maximum points of the range (DIFF model). For each of these, we also considered the possibility that, in the transfer condition, the subjects might have somewhat misestimated the range being tested and/or that the noise in the mental magnitudes on which transfer responding was based was substantially different from the noise on which duration responding was based.

There must be a stochastic stage in a model, because responding "long" is only probabilistically related to the probe number. We assumed that, whatever the underlying comparison was, the Weibull function

$$P=\gamma-(1+\gamma-\lambda)\left(1-2^{-(cp/L)^{S}}\right)$$

would adequately describe the mapping from the results of the comparison to the observed frequencies of "long" responses. P is the predicted proportion of "long" responses; γ is the lower asymptote for this proportion, which should be 0, but because of occasional lapses in the subjects' attention to the task, may not be; λ is the upper asymptote, which should be 1, but may not be; cp is the hypothesized comparison proportion (see Table 1); L is the value of the comparison proportion at which the estimated proportion of "long" responses is .5; S is the shape parameter. We chose the Weibull function because it can take on different forms, depending on the value of the S parameter.

To compare the explanatory power of the different models, we found for each model the Weibull parameters that maximized the binomial likelihood of our data by fitting a Weibull function to the data using the binomial loss function. The maximum likelihood (ML) for a given model is the product of the binomial likelihoods of the observed proportions, given the likelihood-maximizing values of the model's parameters. For each of the three models of the comparison, we fitted the duration and number data either pooled (P) or separately (UP). In pooling the duration and number data before fitting the model, we assume that the subjects' estimates of the referents (arithmetic mean, geometric mean, or range of values tested) were accurate and that the sources of noise were the same in the transfer condition as in the pretransfer conditions. In that case, one set of parameter values for the Weibull function will satisfactorily explain both data sets. When fitting a model to the duration and number data separately, we assume that the subjects' estimates of the referents differed between the training and the transfer task or that the sources of noise were different, in which case two sets of values for the Weibull parameters are needed, one for the duration data and one for the number data. The first case (same parameter estimates for both data sets) is nested within the second (separate parameter estimates for the duration data and number data). To estimate the relative likelihoods of the three models under each of the two cases (P and UP), we used the Schwarz criterion (Schwarz, 1978):

$$S=\log(ML_1)-\log(ML_2)-\frac{1}{2}(d_1-d_2)\log(n),$$

where ML_1 and ML_2 are the maximum likelihoods of the two models, d_1 and d_2 are the numbers of parameters in the two models (4 and 8 for the P and UP comparisons, respectively), and n is the number of observations (trials). The Schwarz criterion (S) gives an approximation to the log of the Bayes factor. The Bayes factor quantifies the explanatory power of the models, vis-à-vis the data, that is their relative explanatory power (see Glover & Dixon, 2004).

Figure 2 shows the relative likelihood profiles for each subject in each of the three experiments. There is a column for each model. The numbers that appear within the columns are subject IDs. Whichever model has a subject's number at the Max level (top of the column) is the best model for that subject. The vertical position of that subject's number in the other columns shows how much less likely those models are. If a subject's number does not appear in a column, that model was more than 1,000-fold less likely—hence, "out of contention."

For all but 1 subject (Subject 1 in Experiment 3), a model in which the probe value was compared with an estimate of the probe range's central tendency—either its arithmetic mean or its geometric mean—was either the best model or "not unlikely" (close to the best). Thus, almost all subjects appeared to extract an estimate of the midpoint of the range being probed and based their decision on the proportion between the probe and this midpoint. This finding is consistent with the pseudologistic model (PLM) developed by Killeen, Fetterman, and Bizo (1997). According to PLM,

Table 1
Models of the Comparison Proportion

AM	GM	DIFF
$N_p/[(N_{max}+N_{min})/2]$	$N_p/(N_{max}*N_{min})^{1/2}$	$(N_p-N_{min})/(N_{max}-N_{min})$

Note—N_p is the number of flashes in the probe. The denominators of the first two models are the subject's estimates of the arithmetic and geometric means of the ranges, respectively. The subject responds on the basis of the proportion (ratio) between the probe and this estimate of central tendency. In the third model, the subject responds on the basis of where, proportionately speaking, the probe falls within the range of values tested.

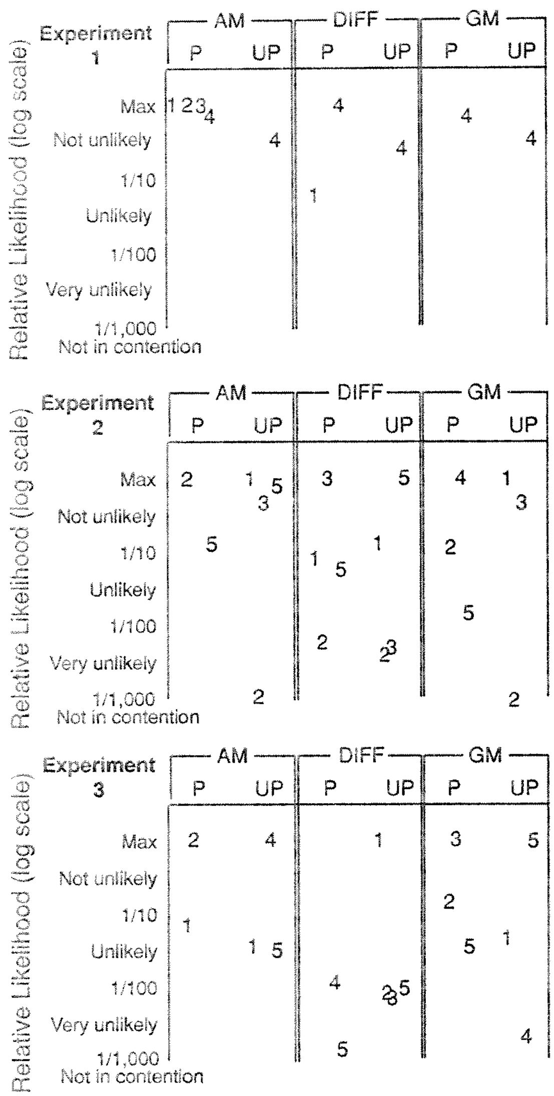

Figure 2. Likelihood profiles. The three types of models (AM, DIFF, and GM; see Table 1), together with the assumption that the data can be either pooled (P) or not (UP), give the six columns, one for each model. The numbers within the columns are subject IDs.

probe values are compared with the bisection point ($T_{1/2}$) rather than the referents that define the endpoints of the range. Allan (2002a) and Allan and Gerhardt (2001) found that the standard of comparison did not change even when the referents changed from trial to trial. Their subjects used the range of referents encountered over many trials to establish a standard that was employed on every trial, regardless of the referents given on that trial.

The midpoint estimates varied across subjects in our experiments. For some subjects, the data were best explained by assuming that their estimate of the midpoint was the arithmetic mean, whereas for others, they were much better explained by assuming that their estimate of the midpoint was the geometric mean. This is consistent with the findings of Allan (2002b).

In Experiments 1 and 2, the subjects might have based their transfer responses on the durations of the sequences rather than on the numbers of flashes they contained. Experiment 3 partially deconfounded variation in number and variation in sequence duration. To test whether, despite our instructions to rely on number, the subjects based their judgments at least in part on duration in the transfer condition, we analyzed for the effects of cumulative flash duration and the total duration of the flash sequence on the probability of a "long" response. We sorted the data according to one or the other of these stimulus durations to see whether the probability of a "long" response on a trial increased with rank of that trial in the duration sort. The plots of the frequency of "long" responses as a function of rank within one or the other duration sorts all appeared flat on casual inspection (Figure 3). We used the chi-square test to check that the proportion of "long" responses in the longer half of a duration sort was not significantly greater than was the proportion in the first half. The results of this test for independence did not approach significance except for Subject 4, for whom both tests were significant, with p values in the .01–.03 range. On closer inspection, a slight upward trend can be discerned in the data for Subject 4 (see Figure 3). Note, however, that the proportion of "long" responses was not much less even at the relatively very short stimulus durations.

GENERAL DISCUSSION

Human subjects, like rat subjects, can transfer a duration discrimination to a number discrimination. In human subjects, however, the transfer is based on similarity in the proportion that a probe (transfer) stimulus bears to a reference established by the range of transfer stimuli rather than on a direct comparison between mental magnitudes representing number and mental magnitudes representing duration. Whether this alternative interpretation applies to nonhuman subjects as well depends critically on whether the transfer in the latter case is range specific. If this transfer is to be used to infer the scale factor relating the mental magnitudes that represent number to the mental magnitudes that represent duration, it must be shown that the transfer is range specific.

The finding that, more often than not, human subjects use the arithmetic mean favors the assumption that the mental magnitudes representing duration and number are scalar functions of (i.e., proportional to) the objective magnitudes, as opposed to the assumption that they are logarithmic functions of the corresponding objective magnitudes (Dehaene, 2001). On the first assumption, the subjective arithmetic mean may be computed mentally in the usual way[1]—by summing the subjective endpoints of the range and dividing the sum by 2. Then, on any given trial, the mental magnitude for the probe value divided by the midpoint estimate gives the mental magnitude on which the subject's decision is based. On the logarithmic compression assumption, what to assume is less clear, because there is no natural way to extract a representation of the objective arithmetic mean from subjective magnitudes proportional to the logarithms of the objective quantities

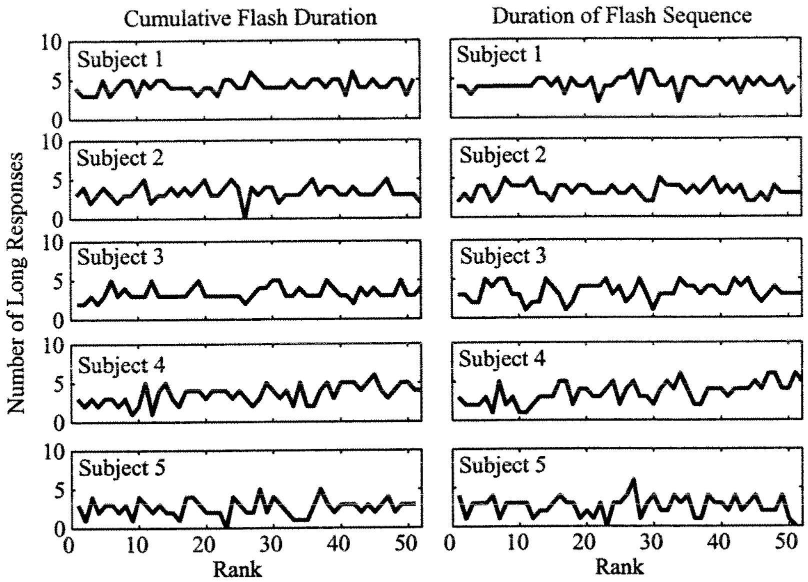

Figure 3. The frequency of "long" responses as a function of rank in a duration sort. Left panels: data sorted according to cumulative flash duration. Right panels: data sorted according to the total duration of the flash sequence.

whose mean is sought. Summing logarithmically compressed mental magnitudes and dividing by 2 gives the logarithm of the geometric mean, not the arithmetic mean.

One could imagine that the brain uses a lookup table to find the logarithm of the quantity that is equal to the sum of the mental magnitudes representing the limits of the range (Dehaene, 2001). It would then subtract from this log a mental magnitude equal to the logarithm of 2 in order to get the mental magnitude that represents the objective arithmetic mean. Finally, the proportion between a probe and the estimate of central tendency would be represented by the arithmetic difference between the two logarithms (the first representing the probe, the second representing the arithmetic mean). On this account, one would also need to assume that the logarithmic compression in the duration and number domains had the same base. Unless the bases for the logarithmic compression in the two domains are the same, the mental magnitudes representing equal proportions within the two domains will not be equal. Thus, a model that consistently works with subjective quantities proportional to the logarithms of the objective quantities appears somewhat ungainly.

In Experiment 3, we confirmed that transfer can occur from proportions among durations to proportions among numbers. By means of a relative likelihood analysis, we established that, for some subjects—those for whom a pooled model was by far the most likely (see Figure 2)—a model that assumes that the psychometric function for discriminating the number probes has parameter values identical to the psychometric function for discriminating the duration probes is much more likely than a model that assumes different parameter values. This kind of conclusion—that the null model is more likely than a model that assumes an unspecified difference—can be reached only through relative likelihood analysis. In the present case, the null model is the most interesting because it implies that the sources of noise in the two cases are quantitatively the same—a point stressed by Meck and Church (1983). It is less clear what to conclude when the unpooled fits are best for subjects. The need to use different parameters could arise either from different amounts of noise or from misestimation of the referents. We did not pursue this further because it involves elaborating still more models.

It is obviously of interest to determine whether cross-domain transfer can be based on within-domain proportions in nonhuman animals. In closing, we stress that it is of more or less equal interest whether transfer occurs on the basis of direct comparisons between mental magnitudes representing number and mental magnitudes representing duration or on the basis of unitless within-domain proportions. In the first case, the transfer phenomenon would make it possible to establish the brain's system of mensuration, the relation between the scale factors relating different kinds of objective quantity to the mental magnitudes that are the common currency of mental mensuration. In the second case, the phenomenon shows the profoundly abstract character of the brain's representation of experience. From a formal perspective, proportions are simply real numbers (or, at any rate, rational numbers). These results also emphasize the readiness with which the brain unconsciously extracts both arithmetic and geometric estimates of central tendencies (see Ariely, 2001).

REFERENCES

Allan, L. G. (2002a). Are the referents remembered in temporal bisection? *Learning & Motivation*, **33**, 10-31.

Allan, L. G. (2002b). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology*, **55B**, 43-60.

Allan, L. G., & Gerhardt, K. (2001). Temporal bisection with trial referents. *Perception & Psychophysics*, **63**, 524-540.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, **12**, 157-163.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, **8**, 698-707.

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, **12**, 244-246.

Fetterman, J. G., & Killeen, P. R. (1995). Categorical scaling of time: Implications for clock-counter models. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 43-63.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, **84**, 279-325.

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, **11**, 791-806.

Killeen, P. R., Fetterman, J. G., & Bizo, L. A. (1997). Time's causes. In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: Psychological and neurobehavioural analyses* (Advances in Psychology, Vol. 120, pp. 79-131). Amsterdam: Elsevier.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 320-334.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology*, **11**, 591-597.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Stevens, S. S. (1964). Concerning the psychophysical power law. *Quarterly Journal of Experimental Psychology*, **16**, 383-385.

NOTE

1. These computations are not conscious ones. They are the kind of unconscious computations that give subjects estimates of, for example, the mean of the diameters of a set of circles (Ariely, 2001). Subjects do not realize they have such an estimate until asked to make judgments based on it, and they have no introspective insight into where their estimates of central tendency come from.

(Manuscript received February 9, 2005;
revision accepted for publication October 24, 2005.)