On the ability to directly evaluate sensory ratios

Sergio Cesare Masin
University of Padua, Italy

Abstract The study explored whether participants could directly evaluate sensory ratios. The results indicate that they did not exhibit this ability for the extensive continuum of length and the intensive continua of brightness and brightness difference. Functional measurement analysis combined with a chronometric analysis of responses indicated that participants used mental counting for length and used difference evaluation for brightness and brightness difference. The finding that participants evaluated brightness differences under instructions to evaluate brightness ratios suggests that difference judgments may be a general human capability. It resolves questions raised by prior tests of Torgerson’s conjecture and by prior consistency tests of ratio evaluation.

The processes underlying metric responses are largely unknown despite the wide use of metric response measures in all fields of psychology, especially in social and personality psychology and in sensory psychophysics (Anderson, 2008; Gescheider, 1997; Marks, 1974; Wegener, 1982). A widespread and persistent assumption is that these processes are based on people’s ability to directly evaluate ratios of mental intensities (Fullerton & Cattell, 1892; Lim, 2011; Luce, 2002; Merkel, 1888; Narens, 2007; Richardson & Ross, 1930; Stevens, 1975).

Torgerson’s conjecture

Garner (1954a) found that participants subdivided a sensory interval in the same manner whether they were asked to subdivide it in terms of sensory differences or of sensory ratios. Using this finding, Torgerson (1961) hypothesized that people fail to distinguish sensory ratios from sensory differences. This hypothesis has been called “Torgerson’s conjecture” (Birnbaum & Veit, 1974). The following are tests of Torgerson’s conjecture.

For fixed sensory intensities, one may rank-order all the possible pairs of these intensities in terms of theoretical differences, theoretical ratios, judged differences, or judged ratios between the intensities in each pair. Theoretical differences and theoretical ratios imply different rank orders of pairs of intensities. Instead judged differences and judged ratios have been found to yield the same rank order of pairs of intensities in agreement with Torgerson’s conjecture (Birnbaum & Elmasian, 1977; Mellers, Davis, & Birnbaum, 1984; Parker & Schneider, 1974; Parker, Schneider, & Kanow, 1975).

For a number \( n \) and the sensory intensities \( \Psi_A, \Psi_B, \) and \( \Psi \), Ellermeier, Narens, & Dielmann (2003) had participants first produce the intensity \( n \cdot \Psi \) and then produce the intensity \( \Psi_C \) by adding the difference \( |\Psi_A - \Psi_B| \) to \( n \cdot \Psi \). They also had participants first produce the intensity \( \Psi_D \) by adding \( |\Psi_A - \Psi_B| \) to \( \Psi \) and then produce the intensity \( n \cdot \Psi_D \). As predicted by Torgerson’s conjecture, it turned out that \( \Psi_C = n \cdot \Psi_D \) as if participants did not distinguish ratios from differences.
These tests suggest that people do not discriminate sensory ratios from sensory differences. They fail to determine whether people evaluate only sensory ratios or only sensory differences. The following consistency tests help resolve this indeterminacy.

Given the sensory intensities $\Psi_A$, $\Psi_B$, and $\Psi_C$ ($\Psi_A < \Psi_B < \Psi_C$) Fagot and Stewart (1969) had participants directly evaluate the ratios $P_{AB} = \Psi_A / \Psi_B$, $P_{BC} = \Psi_B / \Psi_C$, and $P_{AC} = \Psi_A / \Psi_C$ and the differences $\Delta_{AB} = |\Psi_A - \Psi_B|$, $\Delta_{BC} = |\Psi_B - \Psi_C|$, and $\Delta_{AC} = |\Psi_A - \Psi_C|$. Consistent ratio and difference evaluations imply that $P_{AC} = P_{AB} \cdot P_{BC}$ and $\Delta_{AC} = \Delta_{AB} + \Delta_{BC}$, respectively. It turned out that $P_{AC} \neq P_{AB} \cdot P_{BC}$ and $\Delta_{AC} = \Delta_{AB} + \Delta_{BC}$. If Torgerson’s hypothesis that people cannot distinguish ratios from differences holds, then Fagot and Stewart’s (1969) results suggest that people are able only to evaluate sensory differences.

**Objection**

Recently, Luce (2012) has objected that Torgerson’s conjecture is probably false. His argument runs as follows.

One begins by more formally defining Torgerson’s conjecture. Given the sensory intensities $\Psi_A$ and $\Psi_B$ ($\Psi_A < \Psi_B$) let a participant set a sensory intensity $\Psi_p$ such that

$$W\left(\frac{1}{2}\right) = \frac{\Psi_p - \Psi_A}{\Psi_B - \Psi_A}$$

(1)

with $W$ an unknown function, and let this same participant also set a sensory intensity $\Psi_e$ such that

$$1 = \frac{\Psi_e - \Psi_A}{\Psi_B - \Psi_e}.$$  

(2)

Torgerson’s conjecture states that

$$\Psi_p = \Psi_e.$$  

(3)

Garner (1954b) assumed that people are able to directly evaluate sensory ratios but proposed that these evaluations could misrepresent sensory ratios. Setting $W$ to be the identity function amounts to assuming that direct evaluations of sensory ratios represent sensory ratios consistently. Setting $W$ to be a non-identity function amounts to assuming that direct evaluations of sensory ratios systematically misrepresent sensory ratios.

Luce (2012) showed algebraically that $W$ needs to be the identity function for Eq.s 1–3 to hold. Empirical tests of axioms underlying Eq. 1 suggest instead that $W$ is a nonlinear function (Steingrimsson & Luce, 2007). Using this result, Luce (2012) concluded that Torgerson’s conjecture is probably false. This conclusion implies that the above tests of Torgerson’s conjecture were probably flawed.

**Consistency tests**

The following tests of consistency of ratio evaluation with multiplication were proposed by Narens (1996) and carried out by Ellermeier and Faulhammer (2000) and Zimmer (2005). As Luce (2012) has remarked, these tests start from the assumption that direct evaluations of sensory ratios consistently represent sensory ratios ($W$ assumed to be the identity function).
For a sensory intensity $\Psi$ and the numbers $a$, $b$, and $z = a \cdot b$, participants were asked to produce the sensory intensities

\[
\begin{align*}
\Pi_a &= a \cdot \Psi \\
\Pi_b &= b \cdot \Psi \\
\Pi_z &= z \cdot \Psi \\
\Pi_{ab} &= a \cdot \Pi_b \\
\Pi_{ba} &= b \cdot \Pi_a .
\end{align*}
\]

**Commutativity and multiplicativity conditions**

Eq.s 4–8 predict the commutativity condition that $\Pi_{ab} = \Pi_{ba}$. It turned out that $\Pi_{ab} = \Pi_{ba}$ indicating consistency of ratio evaluation with multiplication.

Eq.s 4–7 predict the multiplicativity condition that $\Pi_z = \Pi_{ab}$. It turned out that $\Pi_z < \Pi_{ab}$ for $a$ and $b$ larger than 1 (Ellermeier & Faulhammer, 2000) and $\Pi_z > \Pi_{ab}$ for $a$ and $b$ smaller than 1 (Zimmer, 2005), confirmed by Augustin (2008) and Augustin & Maier (2008). These results indicate that ratio evaluation is inconsistent with multiplication.

**Difficulty of interpretation**

Luce (2012) has remarked that the multiplicativity test was based on the assumption that $W$ was the identity function while empirical data suggest that $W$ is probably nonlinear (Steingrimsson & Luce, 2007). It follows that the result of this test that ratio evaluation is inconsistent with multiplication may have depended on misrepresentation of sensory ratios rather than have depended on participant’s ability to evaluate sensory ratios.

It also follows that the results of the commutativity test, which were based on the assumption that $W$ was the identity function, imply inconsistency of ratio evaluation with multiplication in the event that $W$ should be a nonlinear function.

**Hypotheses about ratio evaluation**

Considering that prior tests of Torgerson’s conjecture and prior consistency tests may not be entirely decisive about humans’ ability to directly evaluate sensory ratios, the present study tried to verify this ability by a different approach, that of functional measurement methodology coupled with a chronometric analysis of responses.

We begin with the common lay notion that people are easily able to evaluate ratios of perceived lengths. Given the task of evaluating the ratio of a perceived length $\Psi_T$ to a perceived or remembered standard length $\Psi_S$, one can make the following hypotheses.

One hypothesis is that the evaluation of the ratio $\Psi_T / \Psi_S$ yields the real number

\[
R = \frac{m}{\Psi_S} \cdot \Psi_T + n
\]

with $\Psi_S$, $m$, and $n$ constant (Stevens, 1936). The constants $m$ and $n$ express the idea that direct ratio evaluations linearly misrepresent true sensory ratios due to constant errors.

Another hypothesis is that $R$ is related to $\Psi_T$ curvilinearly and is closely expressed by

\[
R = p \cdot \Psi_T^b + q
\]
with \( p \) and \( q \) constant, the exponent \( \beta \neq 1 \), and \( \Psi_S \) absorbed in \( p \) (Attnave, 1962; Rule & Curtis, 1982). Eq. 10 is one of various possible ways to express the idea that direct ratio evaluations misrepresent true sensory ratios non-linearly.

A third hypothesis is that people mentally count the fractional number of times \( R \) that \( \Psi_S \) is contained in \( \Psi_T \) implying that

\[
R = \frac{1}{\Psi_S} \cdot \Psi_T + k
\]

with \( \Psi_S \) and \( k \) constant (Krantz, 1972). The constant \( k \) expresses the possible effect of constant errors. The hypothesis that \( \Psi_S \) is constant during mental counting receives support from the finding that, under proper conditions, mental counting yields judgments of lengths in flat surfaces which are consistent with the laws of geometry (Masin, 2012).

Experiment 1 tested the validity of Eqs. 9–11 by functional measurement methodology (Anderson, 1982). This methodology for testing ratio evaluation is discussed in Anderson (1974, pp. 266–271). Given a factorial design in which the factors are \( \Psi_T \) and \( \Psi_S \), Eqs. 9–11 predict that plotting \( R \) against \( \Psi_T \) for each \( \Psi_S \) yields a fan-shaped pattern of diverging factorial curves. More specifically, Eq. 10 predicts that these curves are non-straight curves and Eqs. 9 and 11 predict that these curves are straight lines. If obtained factorial curves form a fan of diverging straight lines then Eq. 10 is refuted and Eqs. 9 and 11 are supported.

Experiment 1 also aimed at distinguishing the processes underlying Eqs. 9 and 11. Using the same stimuli, one group of participants was asked to evaluate \( \Psi_T / \Psi_S \) ratios and another group was asked to mentally count the number of times each \( \Psi_S \) was contained in each \( \Psi_T \). Since ratio evaluation consists in one single operation for each \( \Psi_T / \Psi_S \) ratio, its duration is invariant with \( \Psi_T \) and \( \Psi_S \). Instead, since mental counting consists in one or more single operations of counting one \( \Psi_S \), its duration increases with \( \Psi_T \) and as \( \Psi_S \) decreases. Ratio evaluation is supported if response time is invariant with \( \Psi_T \) and \( \Psi_S \). Mental counting is supported if response time increases with \( \Psi_T \) and as \( \Psi_S \) decreases.

### Experiment 1

**Method**

**Participants** There were 30 university students divided in two equally numerous groups. Stimuli and presentation conditions were the same for each group.

**Stimuli** In a normally lighted room, the stimuli were 0.5-mm thick horizontal lines displayed on a frontal-parallel black monitor screen (Philips Brilliance 190B). The whole screen was covered with an tightly adherent opaque black cardboard with a 250 × 200 mm rectangular hole. The stimuli were concentric with the presentation area of the screen visible through this rectangular hole. Participants had their chin rested 600 mm away from the screen. A microphone was placed in front of the their mouth.

Each trial consisted in presenting one standard stimulus followed by a test stimulus. The standard stimulus was red. It had a length of 6, 9.5, or 15.5 mm and a duration of 1.5 sec. The test stimulus was white. It had a length of 20, 30, 40, or 50 mm and remained visible until the trial terminated. The interstimulus interval was 1.5 sec.
There was one trial for each factorial combination of the lengths of the standard and test stimuli. The resulting series of 12 trials was presented four times consecutively each time with trials in random order. Responses obtained in the first series of 12 trials were excluded from analysis of group data.

The participant’s response was vocal. It activated the microphone placed in front of the participant’s mouth. For each test line a computer recorded the time elapsed from the onset of the test line to the activation of the microphone by the participant’s response. This activation terminated the stimulus and opened a small window in the lower part of the presentation area within which the number pronounced by the participant appeared as it was typed by the experimenter. The participants were told that they needed to respond on the microphone to allow this window to open for the typing of the response. Nothing else was said to the participant about this microphone. The participant was allowed to take as long as needed to respond, as in the procedure adopted by Hartley (1977).

Procedure

One group was asked to report how many times the length of the test stimulus was longer than the length of the standard stimulus (ratio evaluation). The other group was asked to count the number of standard stimulus lengths needed to be added to one another to obtain the length of the test stimulus (mental counting).

Statistical analysis

Unless otherwise specified, statistical significance in analyses of group data was tested using repeated measures $4 (Ψ_T) × 3 (Ψ_S)$ analyses of variance with three replicates. In all factorial plots the parameter is the value of the standard stimulus, with triangles and circles referring to smallest and largest standard stimuli, respectively.

Results and discussion

Let $R$, $Ψ_T$, and $Ψ_S$ denote the numerical response, the perceived length of the test stimulus, and the remembered length of the standard stimulus, respectively. In Fig. 1 the top diagrams show mean $R$, and the bottom diagrams show mean response time, plotted against the physical length of the test stimulus separately for ratio evaluation and mental counting instructions.

In the top diagrams the results show that factorial curves formed a fan-shaped pattern of almost straight lines, for ratio evaluation and for mental counting. Eq.s 9 and 11 predict a fan-shaped pattern of diverging straight lines. As is also shown by the statistics reported below, the results support Eq.s 9 and 11. They reject Eq. 10 since this equation predicts a fan-shaped pattern of non-straight curves.

The results in the bottom diagrams allow distinguishing between the processes underlying Eq.s 9 and 11. They show that mean response time increased with $Ψ_T$ and as $Ψ_S$ decreased, for ratio evaluation and for mental counting. The results for mental counting agree with prior results (Hartley, 1977, 1981). The results support only Eq. 11 in that Eq. 9 predicts one single evaluation operation of constant duration while Eq. 11 predicts multiple elemental operations of counting one $Ψ_S$ whose number increases with $Ψ_T$ and as $Ψ_S$ decreases.
Without completely excluding that participants had the ability to evaluate length ratios, the results suggest that participants did not exhibit this ability when they were asked to use it. Plausibly, participants mentally counted units of length both under the instructions to mentally count length units and under the instructions to evaluate length ratios.

With response time as the dependent variable, a 2 (method) × 4 (ΨT) × 3 (ΨS) analysis of variance showed that the effect of method was significant \[ F(1,28) = 4.8, p < .05 \] with all interactions involving method being not significant. It could be that mental counting instructions increased response time through a decisional process as that which increases reaction times when instructions stress response accuracy.

**Statistics for ratio evaluation** For numerical responses the linear-linear component of the interaction was significant \[ F(1,14) = 34.5, p < .001 \] and the residual components not significant \[ Fs(1,14) < .6 \] except for the quadratic-quadratic component \[ F(1,14) = 7.6, p < .05 \]. The quadratic-quadratic component was not significant after excluding the data for the length of 50 mm from the analysis \[ F(1,14) = .4 \] suggesting some end effect. This effect may be related to the recent finding that participants autonomously replace the experimenter-defined standard with a mental standard of their own choice when they mentally count the standard in test lines about 10 times longer than the experimenter-defined standard (Masin & Anali, 2011).

For response times, \( \Psi_T \) \[ F(3,42) = 12.4, p < .001 \] and \( \Psi_S \) \[ F (2,28) = 4.7, p < .05 \] were significant and the interaction was not significant \[ F(6,84) = 1.9 \].

**Statistics for mental counting** For numerical responses the linear-linear component of the interaction was significant \[ F(1,14) = 50.5, p < .001 \] with all residual components not significant \[ Fs(1,14) < 3.1 \]. For response times, \( \Psi_T \) \[ F(3,42) = 12.2, p < .001 \] and \( \Psi_S \) \[ F(2,28) = 12.0, p < .001 \] and the interaction \[ F(6,84) = 3.0, p < .05 \] were significant.

**Individual analyses** To maximize statistical power, individual analyses used all four replicates per factorial combination—using three replicates produced essentially the same results. The linear trend of \( \Psi_T \) was significant for all participants. The quadratic trend of \( \Psi_T \) was not significant for all participants except for 1 participant in the ratio evaluation group. Table 1 reports the significance of the individual interactions and interaction components. In agreement with mean data, there were many interactions with significant linear-linear component.

### Experiment 2

The above results indicate that participants mentally counted length units under the instruction to evaluate length ratios. Perceived length is an extensive attribute. One may also ask whether ratio evaluation and mental counting are possible for intensive attributes. Experiment 2 sought to answer this question for perceived brightness.

**Method**

**Participants** There were 30 university students different from those used in Experiment 1. They were divided in two equally numerous groups.
Stimuli Experiment 2 was a repetition of Experiment 1 with the following changes. The experimental room was dark. Stimuli were disks of diameter 25 mm. The standard and test stimuli appeared respectively above and below the center of the presentation area with a gap between these stimuli of 25 mm. The luminance of the standard stimulus was 1, 1.9, or 4 cd/m$^2$ and that of the test stimulus was 7.8, 14.9, 25.8, or 48.6 cd/m$^2$.

Procedure One group was asked to report how many times the brightness of the test stimulus was larger than the brightness of the standard stimulus (ratio evaluation). The other group was asked to count the number of standard stimulus brightnesses needed to be added to one another to obtain the brightness of the test stimulus (mental counting).

Results and discussion

Let $\Psi_T$ and $\Psi_S$ denote the perceived test stimulus brightness and the remembered standard stimulus brightness, respectively. In Fig. 2 the top and bottom diagrams show mean $R$ and mean response time, respectively, plotted against the luminance of the test stimulus separately for ratio evaluation and mental counting instructions.

The top diagrams show that obtained factorial curves were almost parallel, for ratio evaluation and for mental counting. With $\Psi_T$ and $\Psi_S$ defined as in this section, the results reject Eqs. 9–11 since each equation predicts diverging factorial curves.

The parallelism theorem of functional measurement states that a linear model of information integration holds if the observable response is a linear scale of the implicit response and if the factorial plot of observable responses exhibits a pattern of parallelism (Anderson, 1982). The results in the top diagrams show parallelism of curves suggesting a linear process of information integration occurring when participants evaluated $\Psi_T / \Psi_S$ ratios. A plausible interpretation of this linear process is that it yielded the real number

$$R = c_0 + c_1 \cdot |\Psi_T - \Psi_S|$$  \hspace{1cm} (12)

with $c_0$ and $c_1$ constants.

The bottom diagrams show that mean response time was nearly invariant with $\Psi_T$ or $\Psi_S$ for ratio evaluation and for mental counting instructions. These results support Eq. 12 since this equation expresses one single operation of constant duration.

The results suggest that participants were unable to evaluate brightness ratios or to mentally count brightness units. An indication that participants were unable to evaluate brightness ratios had been previously obtained using ratio production (Masin, 2007). For given values of a standard brightness, participants produced multiples of these values for different predefined ratios. They then rated each produced brightness. Supporting Eq. 12, these ratings yielded factorial curves that were almost parallel.

With response time as the dependent variable, a $2$ (method) $\times$ $4$ ($\Psi_T$) $\times$ $3$ ($\Psi_S$) analysis of variance showed that the effect of method $[F(1,28) = 2.2]$ and all interactions involving method were not significant.

Statistics for ratio evaluation For numerical responses, $\Psi_T$ $[F(3,42) = 23.4, p < .001]$ and $\Psi_S$ $[F(2,28) = 11.0, p < .001]$ were significant and the interaction was not significant $[F(6,84) = 0.5]$. For response times, $\Psi_T$ $[F(3,42) = 1.3]$ and $\Psi_S$ $[F(2,28) = .3]$ and the interaction $[F(6,84) = .8]$ were not significant.
Statistics for mental counting For numerical responses, \( \Psi_T [F(3,42) = 61.4, p < .001] \) and \( \Psi_S [F(2,28) = 10.2, p < .001] \) were significant and the interaction was not significant \( [F(6,84) = .2] \). For response times, \( \Psi_T [F(3,42) = .6] \) and \( \Psi_S [F(2,28) = .3] \) and the interaction \( [F(6,84) = 1.5] \) were not significant.

Individual analyses The linear trend of \( \Psi_T \) was significant for all participants. The quadratic trend of \( \Psi_T \) was significant for 4 participants in the ratio evaluation group and for 2 in the mental counting group. Table 2 reports significance of individual interactions and interaction components. In agreement with mean data, there were few cases with significant interaction and significant linear-linear component.

Experiment 3

The results of Experiment 2 suggest that participants evaluated the difference \( |\Psi_T - \Psi_S| \) when they were asked to evaluate the ratio \( \Psi_T / \Psi_S \). This ability to evaluate brightness differences suggests that participants could also have the ability to directly evaluate ratios of brightness differences. Indeed, the ability to evaluate ratios of sensory differences is assumed in several current models of the rating response with end anchors (Masin & Busetto, 2010) and in recent studies of brightness judgment (Steingrimsson, 2009, 2011). The following experiment was designed to test the existence of this ability for brightness.

Method

Participants There were 30 university students different from those used in Experiments 1 and 2. They were divided in two equally numerous groups.

Stimuli Experiment 3 was a repetition of Experiment 2 with the only change that each stimulus of Experiment 2 was replaced by a pair of horizontally aligned achromatic disks of diameter 25 mm horizontally separated by a gap of 5 mm. The standard stimulus had luminances of disks of 0.5 and 1, 0.5 and 1.9, or 0.5 and 4 \( \text{cd/m}^2 \). The test stimulus had luminances of disks of 0.5 and 7.8, 0.5 and 14.9, 0.5 and 25.8, or 0.5 and 48.6 \( \text{cd/m}^2 \). For each trial, the luminances were assigned to the left or right disk randomly.

Procedure One group was asked to report how many times the brightness difference in the test stimulus was larger than the brightness difference in the standard stimulus (ratio evaluation). The other group was asked to count the number of standard brightness differences needed to be added to one another to obtain the brightness difference in the test stimulus (mental counting).

Results and discussion

Let \( \Psi_T \) and \( \Psi_S \) denote the perceived test brightness difference and the remembered standard brightness difference, respectively. In Fig. 3 the top and bottom diagrams show mean \( R \) and mean response time, respectively, plotted against the absolute luminance difference in the test stimulus, for ratio evaluation and mental counting instructions.

The top diagrams show that obtained factorial curves were almost parallel, for ratio evaluation and for mental counting. With \( \Psi_T \) and \( \Psi_S \) defined as in this section, the results reject Eq.s 9–11 since each equation predicts a fan-shaped pattern of diverging factorial
curves. The results support Eq. 12 since they obey the parallelism theorem of functional measurement (Anderson, 1982).

The results in the bottom diagrams show that mean response time was essentially invariant with $\Psi_T$ or $\Psi_S$ for ratio evaluation and for mental counting. These results also support Eq. 12 since this equation expresses one single operation of constant duration.

With response time as the dependent variable, a 2 (method) $\times$ 4 ($\Psi_T$) $\times$ 3 ($\Psi_S$) analysis of variance showed that the effect of method [$F(1,28) = 3.4$] and all interactions involving method were not significant.

**Statistics for ratio evaluation** For numerical responses, $\Psi_T$ [$F(3,42) = 34.8, p < .001$] and $\Psi_S$ [$F(2,28) = 4.4, p < .05$] were significant and the interaction was not significant [$F(6,84) = 1.9$]. For response times, $\Psi_T$ [$F(3,42) = .8$] and $\Psi_S$ [$F(2,28) = .5$] and the interaction [$F(6,84) = .3$] were not significant.

**Statistics for mental counting** For numerical responses, $\Psi_T$ [$F(3,42) = 26.7, p < .001$] and $\Psi_S$ [$F(2,28) = 21.5, p < .001$] were significant and the interaction was not significant [$F(6,84) = 1.3$]. For response times, $\Psi_T$ [$F(3,42) = 2.1$] and $\Psi_S$ [$F(2,28) = 1.2$] and the interaction [$F(6,84) = 1.1$] were not significant.

**Individual analyses** The linear trend of $\Psi_T$ was significant for all participants except for 1 in the mental counting group. The quadratic trend of $\Psi_T$ was not significant for all participants except for 4 in the ratio evaluation group and for 3 in the mental counting group. Table 3 reports significance of individual interactions and interaction components. There were few cases with significant interaction and significant linear-linear component.

**General discussion**

**Length (Fig. 1)**

**Reality of mental counting** For perceived length, Eq. 11 predicts that curves relating $R$ to $\Psi_T$ for each $\Psi_S$ form a fan of diverging straight lines. In agreement with Eq. 11, mean $R$ for mental counting yielded a fan-shaped pattern of straight lines.

The reality of mental counting was supported by the increase of mean response time with increasing $\Psi_T$ and decreasing $\Psi_S$, in conformity with the implication that the number of single counts of $\Psi_S$ increases with $\Psi_T$ and as $\Psi_S$ decreases (Hartley, 1977, 1981).

**Nonoccurrence of ratio evaluation** For ratio evaluation instructions, mean response time increased with $\Psi_T$ and as $\Psi_S$ decreased suggesting that participants mentally counted length units under instructions to evaluate length ratios. In Figs. 2 and 3, the invariance of response time with $\Psi_T$ and $\Psi_S$ for ratio evaluation of brightness or brightness difference further supported this possibility.

These results suggest that participants did not exhibit the ability to evaluate length ratios when they were asked to use it. Participants may have only seemingly evaluated ratios considering that mentally counting length units produces results equivalent to those that occur if participants are really able to evaluate length ratios (Eq.s 9 and 11).

**Brightness and brightness difference (Figs. 2 and 3)**

**Reality of difference evaluation** In the top diagrams in Figs. 2 and 3 the parallelism of factorial curves supports Eq. 12. It supports the possibility that participants evaluated
absolute differences in brightness (Fig. 2) and absolute differences in brightness difference (Fig. 3) both when they were instructed to evaluate ratios of brightnesses or of brightness differences and when they were instructed to mentally count units of brightness or of brightness difference, respectively.

These results agree with Torgerson’s conjecture and Fagot and Stewart’s (1969) consistency tests. They suggest that the reason why ratio evaluation and difference evaluation are based on a single judgment operation is that participants use the difference operation for both evaluations.

Nonoccurrence of ratio evaluation and of mental counting

In the top diagrams the patterns of factorial curves suggest that participants did not evaluate ratios of brightnesses or ratios of brightness differences, otherwise factorial curves would have diverged rightward. In the bottom diagrams the patterns of factorial curves suggest that participants did not mentally count brightnesses or brightness differences, otherwise mean response times would have increased with $\Psi_T$ and as $\Psi_S$ decreased.

Consistency tests

The present results agree with the commutativity and multiplicativity tests discussed in the introduction, which were based on the assumption that direct evaluations of sensory ratios consistently represent sensory ratios ($W$ assumed to be the identity function). To see this, consider that Eq. 12 maintains that Eqs 4–8 should be replaced respectively by

$$\Pi_a = \Psi + c_0 + c_1 \cdot |\Psi_a - \Psi|$$  \hspace{1cm} (13)
$$\Pi_b = \Psi + c_0 + c_1 \cdot |\Psi_b - \Psi|$$  \hspace{1cm} (14)
$$\Pi_z = \Psi + c_0 + c_1 \cdot |\Psi_z - \Psi|$$  \hspace{1cm} (15)
$$\Pi_{ab} = \Pi_b + c_0 + c_1 \cdot |\Psi_{ab} - \Pi_b|$$  \hspace{1cm} (16)
$$\Pi_{ba} = \Pi_a + c_0 + c_1 \cdot |\Psi_{ba} - \Pi_a|$$  \hspace{1cm} (17)

with $\Psi_a$, $\Psi_b$, $\Psi_z$, $\Psi_{ab}$, and $\Psi_{ba}$ the produced intensities in the event that $c_0 = 0$ and $c_1 = 1$.

Commutativity

While the equality $\Pi_{ab} = \Pi_{ba}$ agrees with ratio evaluation, it implies also that $|\Psi_b - \Psi| + |\Psi_{ab} - \Pi_b| = |\Psi_a - \Psi| + |\Psi_{ba} - \Pi_a|$ which agrees with the hypothesis that participants evaluated sensory differences.

Multiplicativity

Eqs 6 and 7 express that ratio evaluation predicts that $\Pi_z = \Pi_{ab}$. While the findings that $\Pi_z < \Pi_{ab}$ for $a$ and $b$ larger than 1 and that $\Pi_z > \Pi_{ab}$ for $a$ and $b$ smaller than 1 disagree with ratio evaluation, they agree with difference evaluation. For $c_0 > 0$, Eq. 15 states that participants produce $\Pi_z$ by adding $c_0$ to $\Psi$ once, while Eq. 16 states that they produce $\Pi_{ab}$ by adding $c_0$ to $\Psi$ twice—once to produce $\Pi_b$ and once to produce $\Pi_{ab}$. This double addition of $c_0$ may explain why $\Pi_z < \Pi_{ab}$ for $a$ and $b$ larger than 1 and why $\Pi_z > \Pi_{ab}$ for $a$ and $b$ smaller than 1.

General conclusion

When asked to directly evaluate sensory ratios, the results indicate that participants did not exhibit this ability for any of the sensory attributes studied. The results for length suggest that participants used mental counting in place of ratio evaluation and those for
brightness and brightness difference suggest that participants were not able to use mental counting. The results for these dimensions support the hypothesis that participants used difference evaluation in place of ratio evaluation. While the present results offer empirical evidence indicating that people do not directly judge ratios when asked to do so, they suggest that difference judgments may be a general human capability at least with intensive dimensions. Considering the importance of this result for the understanding and modeling of the processes underlying metric responses, the present methodological approach will need to be assessed in the future with other stimulus dimensions.

Acknowledgment I wish to thank Giuseppe Toffan for technical assistance with recording of response times.

References


### Table 1

Results of analysis of individual interactions from Experiment 1. Filled cells show significance level of interaction and of interaction component with L and Q meaning linear and quadratic, respectively. Empty cells signify non-significance at the 0.15 level.

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Ratio Evaluation Interaction</th>
<th>L-L</th>
<th>L-Q</th>
<th>Q-L</th>
<th>Q-Q</th>
<th>Mental Counting Interaction</th>
<th>L-L</th>
<th>L-Q</th>
<th>Q-L</th>
<th>Q-Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.09</td>
<td>.09</td>
<td>.06</td>
<td></td>
<td></td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.09</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.10</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.06</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.07</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.01</td>
<td>.07</td>
<td>.08</td>
<td></td>
<td></td>
<td>.06</td>
<td>.00</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>.02</td>
<td></td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.05</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.00</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>.07</td>
<td>.01</td>
<td>.10</td>
<td></td>
<td>.01</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.04</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.02</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td></td>
<td>.01</td>
<td>.00</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.05</td>
<td>.15</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.09</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.00</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Results of analysis of individual interactions from Experiment 2. Filled cells show significance level of interaction and of interaction component with \( L \) and \( Q \) meaning linear and quadratic, respectively. Empty cells signify non-significance at the 0.15 level.

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Ratio Evaluation Interaction</th>
<th>( L-L )</th>
<th>( L-Q )</th>
<th>( Q-L )</th>
<th>( Q-Q )</th>
<th>Mental Counting Interaction</th>
<th>( L-L )</th>
<th>( L-Q )</th>
<th>( Q-L )</th>
<th>( Q-Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.04</td>
<td>.11</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.07</td>
<td></td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>.04</td>
<td>.05</td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.11</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
<td></td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Results of analysis of individual interactions from Experiment 3. Filled cells show significance level of interaction and of interaction component with \( L \) and \( Q \) meaning linear and quadratic, respectively. Empty cells signify non-significance at the 0.15 level.

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Ratio Evaluation Interaction</th>
<th>( L-L )</th>
<th>( L-Q )</th>
<th>( Q-L )</th>
<th>( Q-Q )</th>
<th>Mental Counting Interaction</th>
<th>( L-L )</th>
<th>( L-Q )</th>
<th>( Q-L )</th>
<th>( Q-Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.09</td>
<td>.12</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>.09</td>
<td>.10</td>
<td></td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>.07</td>
<td>.05</td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>.04</td>
<td>.07</td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1 Results for length. Mean response and mean response time plotted against physical length for three different standard lengths, for ratio evaluation and for mental counting instructions. Triangles and circles refer to the shortest and longest standard.
**Fig. 2** Results for brightness. Mean response and mean response time plotted against luminance for three different standard luminances, for ratio evaluation and for mental counting instructions. Triangles and circles refer to the smallest and largest standard.
Fig. 3 Results for brightness difference. Mean response and mean response time plotted against luminance difference for three different standard luminance differences, for ratio evaluation and for mental counting instructions. Triangles and circles refer to the smallest and largest standard.