

## Chapter 7.

# MEASURING THE SPEED OF MENTAL PROCESSING

At one time it was believed that thought was instantaneous and therefore attempts to measure its speed were futile. This belief was overturned in 1850 when Herman von Helmholtz measured the speed of neural transmission and showed that nerve impulses travel at speeds that are not particularly fast (around 30 m/s or about 100 kilometers per hour). Because the physical substrate of thought is presumably based in the nervous system, Helmholtz's measurement led people to believe that thoughts might also require a measurable amount of time to take place. With his measurement, known as the **subtraction method**, Helmholtz also introduced an important idea to psychological research: comparisons between carefully selected conditions might provide relatively direct measurements of otherwise inaccessible processes. Although contemporary psychologists now usually prefer a modification to the subtraction method known as the **additive factors method** (introduced by Ulric Neisser, Saul Sternberg and other cognitive psychologists in the 1950s and 1960s)<sup>1</sup>, the subtraction method still plays a key role in brain imaging studies of psychological functioning.

***The subtraction method.*** The basic idea of the subtraction method for reaction time is that a measure of the duration of a particular process can be found by obtaining two measurements of time that include the process and subtracting one from the other. The method applies to situations in which time can be directly measured for the completion of a physically inseparable series of events that each take time but cannot be measured individually. The challenge solved by the subtraction method is to measure the time taken by one or more of these distinct but inseparable events.



Herman von Helmholtz (1821-1894)

For example, deciding which of two racers is the faster runner is not the same as deciding who can run the faster race (the latter is determined by who finishes the

<sup>1</sup> The name *additive factors method* was invented by Sternberg in 1972.

race first). The time to finish the race depends on the time for the runner to respond to the starter pistol, get out of the starting blocks, reach top running speed, and then cross the finish line. As competitive racers know, if one racer gets out the blocks or reaches top running speed more quickly than another, that racer could win the race even if the other person has a higher top running speed. Thus, just timing the race does not provide a pure measure of running speed. Timing a race is a measure of non-running components combined with running components in a continuous flow of activity. The subtraction method eliminates the non-running components from the total time and allows running speed to be measured with only a stopwatch, activated when the starter pistol sounds and halted when a runner crosses the finish line.

In the case of the runner, the critical ingredient for the subtraction method is the ability to make two measurements that differ only because of running speed. This can be achieved by measuring the time for the runner to cross a finish line placed at, for example, 90 meters (m), as well as the time for the runner to cross a finish line placed at 100 m. Comparing these two times provides a relatively pure measure of running speed, because the time to reach 90 m differs from the time to reach 100 m only because of the additional distance. Suppose the time to cross the 90 m mark is 9.2 seconds (s), and the time to reach the finish line at 100m is 10 s. The actual running speed is the distance covered ( $100\text{m} - 90\text{m} = 10\text{ m}$ ) divided by the time to cover that distance ( $10\text{s} - 9.2\text{s} = .8\text{ s}$ ). The resulting measurement ( $10\text{m} / .8\text{s} = 12.5\text{ m/s}$  or about 28 mph) is the top running speed.

Helmholtz was the first person to use this method to measure the speed of the nerve impulse. He placed a stimulating electrode at one end of a sciatic nerve (dissected from a frog), placed a recording electrode on the nerve about 4 cm away, and recorded the time for the nerve impulse triggered by his stimulation to be detected by the recording electrode. Then he moved the recording electrode a little farther from the stimulating electrode, stimulated the nerve a second time, and recorded the time for this impulse to be detected. By subtracting the time for detection when the electrode was 4 cm away from the stimulator from the time for detection when the electrode was farther from the stimulator, Helmholtz arrived at a measure of the transmission speed of the nerve impulse: about 30 m/s.

***The subtraction method in psychology.*** Applied to questions about the speed or duration of a psychological process, the subtraction method requires assuming that completion of a given task involves a set of component processes that take place sequentially. With this assumption, the total time for completing the task is simply the sum of the times to complete each component process, just as the time to complete a race is the sum of the times to complete each distance segment. To measure the time to complete a specific component requires subtracting a time that does not include the component from a time that does. About a decade after Helmholtz's measurement of the nerve impulse, Franciscus Donders, a Dutch psychologist, developed a set of three tasks that he believed would allow the measurement of the time to make a decision, as distinct from the time to execute a response. This distinction between a decision time and a reaction time is evident in the case of a driver approaching an intersection just as the signal light turns yellow. The driver has to decide whether to brake and stop or accelerate through the intersection before the light turns red, and the time spent deciding what to do contributes to the total time to carry out either response. Donders' three tasks were the following:



Franciscus Donders (1818 - 1889)

- a *simple reaction time task*, consisting of a single signal stimulus to which the subject responds as quickly as possible by making a predetermined response, as in pressing a button as soon as a single, white light is turned on;
- a *go/no-go task*, consisting of two stimuli but with only one designated as a signal for responding, as in pressing a button when a green light is turned on but not pressing when a red light is turned on; and
- a *choice reaction time task*, consisting of two signal stimuli to which the subject makes two different responses, as in pressing a left hand button to a green light and a right hand button to a red light.

Donders proposed that the time for the choice reaction task was composed of a simple reaction time (measured in the first task) and a stimulus discrimination time (measured in the second task) as well as the time for making a choice. Donders also suggested that measuring performance in all three tasks made it possible to find both the stimulus discrimination time and the choice time

by subtraction.

In equation form, Donders proposed the following:

$$\begin{array}{lcl}
 \text{Simple Reaction Time} & = & \text{Signal} \quad \text{Motor} \\
 & & \quad \quad + \\
 & & \text{Perception} \quad \text{Response} \\
 \\
 \text{Go/ No Go Reaction Time} & = & \text{Signal} \quad \text{Stimulus} \quad \text{Motor} \\
 & & \quad \quad + \quad \quad \quad + \\
 & & \text{Perception} \quad \text{Discrimination} \quad \text{Response} \\
 \\
 \text{Choice Reaction Time} & = & \text{Signal} \quad \text{Stimulus} \quad \text{Response} \quad \text{Motor} \\
 & & \quad \quad + \quad \quad \quad + \quad \quad + \\
 & & \text{Perception} \quad \text{Discrimination} \quad \text{Choice} \quad \text{Response}
 \end{array}$$

Subtracting the time for the Go/No Go task from the time for the Choice Reaction Time task, Donders believed, provided a pure measure of the *response choice* time. Similarly, he believed that subtracting the time for the Simple Reaction Time task from the time for the Go/ No Go Reaction task gave a pure measure of stimulus discrimination time.

***Problems with subtraction.*** As a practical matter, Donders' method ran into a serious problem when some studies found average times for the choice reaction task that were shorter than the average times for the stimulus discrimination task. This result is completely inconsistent with the idea that the choice task incorporates the stimulus discrimination task.

The major problem with Donders' version of the subtraction method appears to be the fact that, with repetition on almost any task, people become experts on it, and expert performance is simply not comparable to the performance of novices. Thus, unless the degree of expertise has been equated for a simple reaction time task and a choice reaction time task, the time for completing one task will bear no meaningful relation to the time for completing the other.

***The Additive Factors method*** To get around the problems created by comparing one task to another in the subtraction method, contemporary psychologists have developed the additive factors method. As in Donders' analysis, the choice reaction time task is considered to be a complex task that includes multiple underlying component processes. To measure these

processes, however, the additive factors method relies, not on different tasks, but on different variations of the choice reaction time task. Thus, the task does not change, but the specific information required for different choices does.

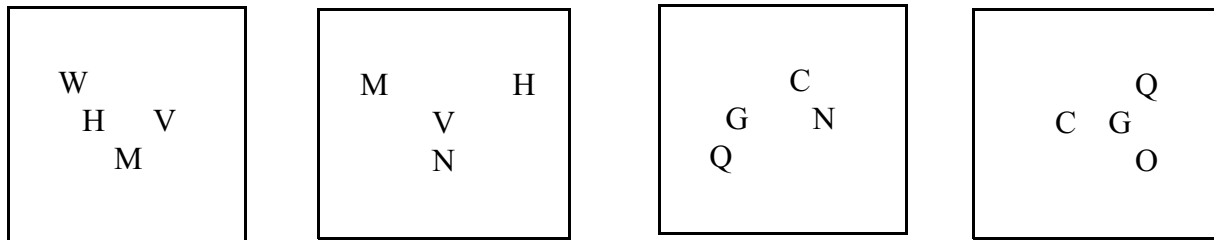
The logic of the additive factors method involves three assumptions about measurement.

- \* **Measuring a single small unit can be accomplished by measuring a *large, known quantity of small units then dividing by the number of units.*** For example, to measure the thickness of a single sheet of paper with an ordinary ruler, measure the thickness of a stack of paper sheets, then divide by the number of sheets. If a ream of paper, which is 500 sheets, is 2 inches thick, then each sheet of paper is  $2/500$  or .004 (four thousandths) inches thick. [To measure something that is four thousandths of an inch thick with a ruler that has markings only to an accuracy of  $1/16$  of an inch (which is .0625 inches) or even  $1/8$  of an inch (which is .125 inches) seems like a remarkable accomplishment.].
- \* **The total time to complete a response is literally the sum of the times of separate processing steps.** Thus, if it takes 200 msec to complete visual decoding, 300 ms to complete scanning, and 500 ms to make a detectable response, the total time to respond is  $200 + 300 + 500$ , or 1000 ms.
- \* **Some individual processing steps, such as pressing a response key, take the same amount of time regardless of the amount of information to be scanned.** For example, if response execution takes 500 ms when scanning 1 item, it should only take 500 ms when scanning 4 items. The times for steps that stay the same can be treated as constants.

### ***An illustration of the additive factors method: Visual scanning***

Looking for a friend in a crowded hall, we scan our visual field, shifting our attention from face to face, region to region, until we find the target of our search. The time to shift attention from one face to another is an example of an internal process that cannot be directly observed but which can be measured using the additive factors method. To illustrate the method, this section will describe the logic and the steps involved in measuring scanning time, based on a laboratory task in which the subject is looking for a target letter in displays of letters that may or

may not contain the target letter. Suppose the target letter is “N”. Figure 7-1 gives examples of displays of 4 letters, some of which contain the target and some of which do not.



**Figure 7-1** Displays of letters for a visual scanning task to find the target letter “N”.

In the task, a particular display is shown and remains visible until the subject presses either a key marked “Present”, to indicate that an “N” is present (as in the second and third displays in the figure) or a key marked “Absent”, to indicate that the display does not contain an “N” (as in the first and fourth displays in the figure). The time between the onset of the stimulus display and the completion of the key press is the directly observable reaction time (RT).

However, the observed RT is not simply a measure of how fast the subject scanned the 4 letters in the display. The observed RT includes the time needed to encode the display initially, defined as encoding time (ET), the time needed to scan the display, defined as scan time (ST), the time needed to make a decision, defined as decision time (DT), and the time needed to execute a movement, defined as motor time (MT). If encoding, scanning, deciding, and making a movement are assumed to be separate, independent steps in the total process, and if these steps are assumed to occur in sequence, then the observed RT can be expressed as the sum of the times for each of these steps, as shown in Equation 7-1:

$$RT_4 = ET_4 + ST_4 + DT_4 + MT_4 \quad \text{Eq. 7-1}$$

This equation states that the observed reaction time to find one letter among 4 alternatives ( $RT_4$ ) is the sum of times to encode, to scan, to decide, and to make a movement when there are 4 alternatives (the subscript indicates that these times are for the case of 4 alternatives).

Now, suppose the task is changed so that the subject has to find an "N" among a set of 16 letters, but that in all other respects the task remains the same as the 4-letter task just described. In this case, the subject would scan a display of 16 letters until finding an "N" and pressing the response key marked "Present" or not finding the letter and pressing the key marked "Absent". In this case the observed RT can likewise be calculated from the times for the steps of encoding, scanning, deciding and making a movement, as shown in Equation 7-2:

$$RT_{16} = ET_{16} + ST_{16} + DT_{16} + MT_{16} \quad \text{Eq. 7-2}$$

Common sense leads us to expect that finding an "N" among 16 letters is harder than finding it among 4 letters, hence  $RT_{16}$  is expected to be greater than  $RT_4$ . But, in terms of the equations, where does the additional time come from? That is, which among the 4 separate steps takes more time, when the display contains 16 instead of 4 items?

Consider the encoding time. If "encoding time" refers to nothing more than the time to convert the visual display from visual energy to neural activation, then encoding time should not depend on display size, and  $ET_{16} = ET_4$ .

Similarly, if movement time refers to just the time to initiate a movement, then movement time also should not depend on display size, and  $MT_{16} = MT_4$ .

Finally, if "decision time" refers to the time to decide what overt response to make, then decision time should not depend on display size either, and  $DT_{16} = DT_4$ .

According to this argument, the only step that takes more time between responding to a 4-letter display and responding to a 16-letter display is the "scanning time". There are more letters to scan with 16 letters than with 4 letters, and, if scanning a single item takes a fixed amount of time, then more time is needed to scan more items. In other words, if the reaction time for the 4 item display is subtracted from the reaction time for the 16 item display, the difference,  $RT_{16} - RT_4$ , is a direct function of the extra time needed to scan 16 as compared to 4 items:

$$RT_{16} = ET_{16} + ST_{16} + DT_{16} + MT_{16} \quad (2)$$

$$- \quad RT_4 = ET_4 + ST_4 + DT_4 + MT_4 \quad (1)$$

$$\text{-----}$$

$$RT_{16} - RT_4 = ST_{16} - ST_4 \quad (3)$$

Assuming that the difference,  $ST_{16} - ST_4$ , is due simply to the time to scan an additional 12 items in the 16-item display, this equation leads to a direct measure of the scanning rate for 12 items:

$$RT_{16} - RT_4 = ST_{12} \quad (4)$$

Furthermore, assuming that scanning each item takes a fixed amount of time, the time to scan 12 items is simply the scanning time for 1 item multiplied by 12:

$$ST_{12} = ST_1 \times 12 \quad (5)$$

Rearranging and dividing by 12 gives the following equations for the time to scan a single item:

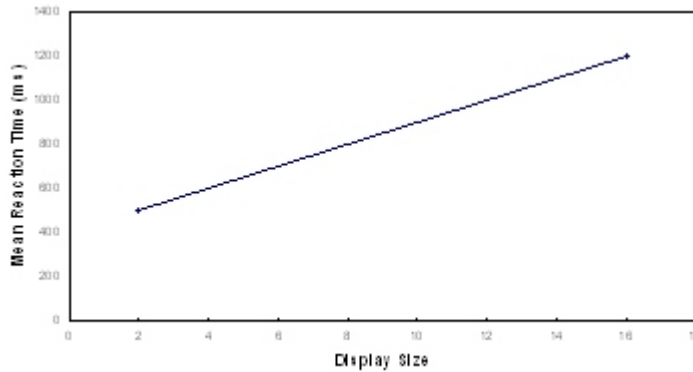
$$\begin{aligned} ST_1 &= ST_{12} / 12 \\ &= (RT_{16} - RT_4) / 12 \end{aligned}$$

This example shows how to measure the scanning time for a single item, using two different display sizes. Including more display sizes, such as displays with 8 items, permits additional measurements of scanning time. If displays with 8 letters were included, then, following the argument just presented for displays of 4 and 16 letters, it would be possible to calculate an  $ST_8$  from  $RT_{16}$  and  $RT_8$  and an  $ST_4$  from  $RT_8$  and  $RT_4$ , to yield additional estimates of  $ST_1$ . Presented with several different estimates of  $ST_1$ , we need a way to choose the "best" estimate.

***Using slope to measure processing speed*** As an illustration of how to get a best estimate of



scanning time, consider the pictorial representation of the data shown in Figure 7-2, which plots RT on the y-axis and the number of items in the display (or display size) on the x-axis. This representation is based on the preceding example, which yields 2 points, one for  $RT_4$  and one for  $RT_{16}$ .



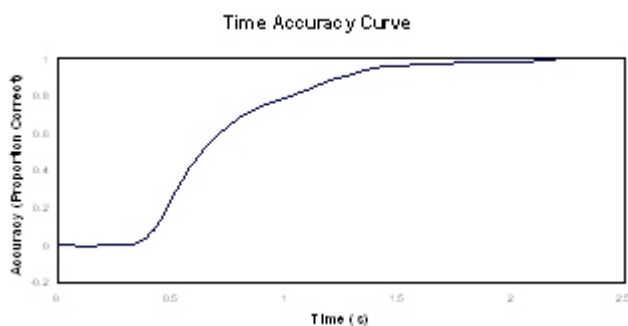
The previous discussion reached the conclusion that  $RT_{16}$  is greater than  $RT_4$  because of the additional scanning time needed in the case of  $RT_{16}$ . With a pictorial representation, this conclusion is equivalent to saying that the amount of increase on the y-axis, as display size goes up from 4 to 16 on the x-axis, is due to visual scanning time. In mathematics, any number that expresses a change on the y-axis relative to a change on the x-axis is a "slope". Thus, the measure of scanning time expressed in Equation 5 is related to the slope of the line that connects the data points for  $RT_{16}$  and  $RT_4$ .

If more than 2 points are plotted (for example, if an  $RT_8$  is also plotted), chances are good that the points will not all fall exactly on a single line. While it is possible to connect any *two* points with a straight line and find a slope for that line, different pairs of points would produce different lines, each with a different slope. What is needed is a way to draw a single line that takes into account all the points. Fortunately, finding a single line to represent multiple points is a familiar problem in statistics, where the accepted solution is to calculate a "line-of-best-fit" or "regression" line. Thus, if we find a line-of-best-fit, the slope of that line can be used to provide

a measure of scanning time.

*Speed-accuracy tradeoffs.* The **speed-accuracy tradeoff** refers to the fact that everyone can “trade” speed for accuracy or vice versa. When it is important to do a task accurately, then we tend to go more slowly, whereas when it is important to complete a task quickly, we are willing to accept some errors. The importance of this common sense fact is not always fully appreciated when reaction time measures of performance are studied in research. However, one implication of the possibility for tradeoffs between speed and accuracy is that comparisons between two conditions on the basis of the time to complete the task can only be meaningful when the error rates are the same (consider typing speed).

The nature of the tradeoff is illustrated graphically in Figure 7-1, which plots accuracy on the y-axis as a function of time since the onset of a signal stimulus, such as test probe. For a short initial period, no change in accuracy occurs, because a minimum amount of time is required before any information is available to guide a decision. Once some information has been extracted, accuracy starts to rise fairly quickly as the elapsed time continues to increase, but eventually the level of accuracy reaches a plateau, and further increases become smaller and smaller.



**Figure 7-3.** Schematic diagram of a time accuracy tradeoff.