From Embodiment, Ego-space, and Action
edited by R.L. Klatzky, B. Mac Whinney,
& M. Behrmann
New York: Taylor & Francis, 2008

1

Measuring Spatial Perception with Spatial Updating and Action

Jack M. Loomis and John W. Philbeck

Measurement of perceived egocentric distance, whether of visual or auditory targets, is a topic of fundamental importance that is still being actively pursued and debated. Beyond its intrinsic interest to psychologists and philosophers alike, it is important to the understanding of many other topics which involve distance perception. For example, many complex behaviors like driving, piloting of aircraft, sport activities, and dance often involve distance perception. Consequently, understanding when and why errors in distance perception occur will illuminate the reasons for error and disfluency in these behaviors. Also, the understanding of distance perception is important in the current debate about the "two visual systems," one ostensibly concerned with the conscious perception of 3-D space and the other with on-line control of action. Similarly, determining whether nonsensory factors, such as intention to act and energetic state of the observer, influence perceived distance, as has been claimed (e.g. Proffitt, Stefanucci, Banton, & Epstein, 2003; Witt, Proffitt, & Epstein, 2004, 2005) depends critically on the meaning of distance perception and how it is to be measured. Still another topic where measurement of distance perception is critical is spatial updating (the imaginal updating of a target perceived only prior to observer movement) involving observer translation. Being able to measure the accuracy of spatial updating depends upon being able to partial out errors due to misperception of the initial target distance (Böök & Gärling, 1981; Loomis, Klatzky, Philbeck, & Golledge, 1998; Loomis, Lippa, Klatzky, & Golledge, 2002; Philbeck, Loomis, & Beall, 1997). Finally, measurement of distance perception is important for the development of effective visual and auditory displays of 3-D space. Indeed, developing virtual reality systems that exhibit naturally appearing scale has proven an enormous challenge, both for visual virtual reality (Loomis & Knapp, 2003) and for auditory virtual reality (Loomis, Klatzky, & Golledge, 1999), and there has been a spate of recent research articles concerned with understanding the causes for uniform scale compression in many visual virtual environments (e.g., Creem-Regehr, Willemsen, Gooch, & Thompson, 2005; Knapp, 1999; Knapp & Loomis, 2004; Sahm, Creem-Regehr, Thompson, & Willemsen, 2005; Thompson, Willemsen, Gooch, Creem-Regehr, Loomis et al., 2004). Virtual reality systems that successfully create a realistic sense of scale will enjoy even greater aesthetic impact and user acceptance and will prove even more useful in the training of skills, such as safe road crossing behavior by blind and sighted children.

Indirectness of Perception

Naïve realism is the commonsense view that the world we encounter in everyday life is identical with the physical world that we come to know about through our schooling. Following decades of intellectual inquiry, philosophers of mind and scientists have come to an alternate view referred to as "representative realism"—that contact with the physical world is indirect and that what we experience in everyday life is a representation created by our senses and central nervous system (e.g., Brain, 1951; Koch, 2003; Lehar, 2003; Loomis, 1992; Russell, 1948; Smythies, 1994). Indeed, this representation, generally referred to as the phenomenal world, is so highly consistent and veridical that we routinely make life-depending decisions without ever suspecting that the perceptual information upon which we are relying is once removed from the physical world. The high degree of functionality of the perceptual process accounts for its being self-concealing and for the reason that most laypeople and indeed many scientists think of perception as little more than attention to aspects of the environment.

The representational nature of perceptual experience is easy to appreciate with color vision because the mapping from physical stimulation to perceptual space entails a huge loss of information, from the many dimensions of spectral lights to the three perceptual dimensions of photopic color vision. In order to appreciate the representational nature of perception more generally it is helpful to keep in mind such perceptual phenomena as diplopia, binocular stereopsis elicited by stereograms, geometric visual illusions, and motion illusions; such phenomena point to a physical world beyond the world of appearance. Although experiencing such phenomena momentarily reminds us of the representational nature of perception, we too easily lapse back into naïve realism when driving our cars, engaging in sports activity, and interacting with other people. It is quite an intellectual challenge to appreciate that the very three-dimensional world we experience in day-to-day life is an elaborate perceptual representation. Indeed, many people seem to be naïve realists when it comes to visual space perception, for they think of visual space perception largely as one of judging distance. But, visual space perception is so much more than this—it gives rise to our experience of the surrounding visual world, consisting of surfaces and objects lying in depth (e.g., Gogel, 1990; Howard & Rogers, 2002; Loomis, Da Silva, Fujita, & Fukusima, 1992; Marr, 1982; Ooi, Wu, & He, 2006; Wu, Ooi, & He, 2004). Virtual reality makes the representational nature of visual space perception obvious (Loomis, 1992), for the user experiences being immersed within environments which have no physical existence (other than being bits in computer memory). Teleoperator systems are useful for drawing the same conclusion. Consider a visual teleoperator system consisting of a head-mounted binocular display and externally mounted video cameras for driving the display. The user of such a teleoperator system experiences full presence in the physical environment while being intellectually aware that the visual stimulation comes only indirectly by way of the display. Because the added degree of mediation associated with the display pales in comparison with the degree of mediation associated with visual processing, the representational nature of perception when using a teleoperator points to the representational nature of ordinary perception.

How one conceives of perception determines how one goes about measuring perceived distance. For the researcher who accepts naïve realism, perceiving distance is simply a matter of judging distance in "physical space." Under this conception, one can simply ask the observer how far away objects are and then correct for any judgmental biases, such as reporting 1 m as 2 m. In contrast, for researchers who adhere to the representational conception, the measurement of distance perception is a major challenge, inasmuch as one is attempting to measure aspects of an internal representation. Because one starts with behavior of some kind (e.g., verbal report, action) and because there can be distortions associated with the readout from internal representation to behavior, measurement of perception depends on a theory connecting internal representation to behavior, a theory that is best developed using multiple response measures (e.g., Foley, 1977; Philbeck & Loomis, 1997).

Some Methods for Measuring Perceived Distance

Verbal report and magnitude estimation are two traditional methods for measuring perceived distance (Da Silva, 1985). Figure 1.1 gives the results of a number of studies using verbal report for target distances out to 28 m (Andre & Rogers, 2006; Foley, Ribeiro, & Da Silva, 2004; Kelly, Loomis, & Beall, 2004; Knapp & Loomis, 2004; Loomis et al., 1998; Philbeck & Loomis, 1997). The data sets are generally well fit by linear functions with 0 intercepts, but the slopes are generally less than 1.0. The mean slope is 0.80.

Concerns about the possible intrusion of knowledge and belief into such judgments (Carlson, 1977; Gogel, 1974) have prompted the search for alternative methods. So-called indirect methods make use of other perceptual judgments thought to be less subject to intrusion by cognitive factors and then derive estimates of perceived distance by way of theory. Several of these methods rely on so-called percept-percept couplings. Space perception researchers have long known that perceptual variables often covary with one another (Epstein, 1982; Gogel, 1984; Sedgwick, 1986). In some cases these covariations may be the result of joint determination by common stimulus variables, but in other cases variation in one perceptual variable causes variation in another (Epstein, 1982; Gogel, 1990; Oyama, 1977); such causal covariations are referred to as percept-percept couplings. The best known coupling is that between perceived size and perceived egocentric distance and is referred to as size-distance invariance (Gilinsky, 1951; McCready, 1985; Sedgwick, 1986). Size-distance invariance is the relationship between perceived size (S') and perceived egocentric distance (D') for a visual stimulus of angular size α: $S' = 2D' \tan(\alpha/2)$. A special case of size-distance invariance is Emmert's Law—varying the perceived distance of a stimulus causes perceived size to vary proportionally, with angular size held constant. Another coupling of perceptual variables is that between the perceived distance of a target and its perceived motion (Gogel, 1982, 1993). Gogel demonstrated that the perceived motion of an object can be altered by mere changes in its perceived distance while keeping all other variables constant. He developed a quantitative theory for this coupling between perceived distance and perceived motion and applied it in explaining the apparent motion of a variety of stationary objects, such as depth-reversing figures and the inverted facial mask (Gogel, 1990). The existence of percept-percept couplings is methodologically important, for these couplings can be used to measure perceived distance in situations where the researcher wishes observers not to be aware that perceived distance is being measured. Judgments of perceived size and perceived motion have been used to measure perceived distance (e.g., Gogel, Loomis, Newman, & Sharkey, 1985; Loomis & Knapp, 2003) and to demonstrate the effect of an experimental manipulation on perceived distance (e.g., Hutchison & Loomis, 2006a).

Another indirect method of measuring perceived distance involves judgments of collinearity and relies on the perception of exocentric direction. A visible pointer is adjusted by the observer to be aligned with the target stimulus (Wu, Klatzky, Shelton, & Stetten, 2005); these authors used the method to measure the perceived distance of targets within arm's reach under the assumption that the pointer is perceived correctly. Application of the method to the measurement of large perceived distances seems promising, but the method will have to compensate for systematic biases in exocentric direction perception (Cuijpers, Kappers, & Koenderink, 2000; Kelly et al., 2004).

Still other indirect methods rely on judgments of perceived exocentric extent and attempt, by way of theory, to construct scales of perceived distance. The best known example is the work by Gilinsky (1951) and, more recently, Ooi and He (2007). In their experiments, observers constructed a set of equal-appearing intervals on the ground extending directly away from the observer. The more distant intervals had to be made progressively larger in order to appear of constant size. Assuming that perceived egocentric distance over the ground plane to a given point is the concatenation of the equal appearing intervals up to that point, the derived perceived distance can be associated with the corresponding cumulative physical

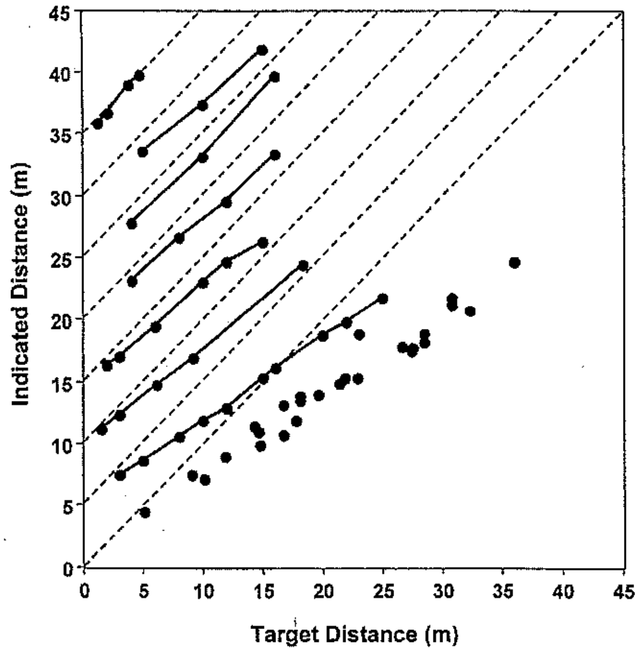

Figure 1.1 Summary of verbal reports of distance for visual targets. The data from the different studies have been displaced vertically for purposes of clarity. The dashed line in each case represents correct responding. Sources from top to bottom: Experiment 3 of Philbeck and Loomis (1997), experimental condition of Experiment 2A of Loomis et al. (1998), calibration condition of Experiment 2A of Loomis et al. (1998), results for gymnasts in Experiment 1 of Loomis et al. (1998), results of full field of view condition of Knapp & Loomis (2004), mean data from the control conditions in the 3 experiments of Andre and Rogers (2006), results of Kelly, Loomis, and Beall (2004), and egocentric distance judgments of Foley et al. (2004).

distance. The derived function of perceived egocentric distance is compressively nonlinear even within 10 m and under full-cue conditions. Because the derived function is noticeably discrepant with other functions to be discussed here and because there are process interpretations for doubting that the derived function is indeed a measure of perceived distance, we do not discuss it further.

Methods Based on Action and Spatial Updating

Given the importance of distance perception and the lack of consensus about how to measure it, researchers have occasionally proposed new measurement procedures. Here, we focus on relatively new methods for measuring perceived distance that rely on action, sometimes with the involvement of spatial updating. The typical procedure begins with the stationary observer viewing or listening to a target stimulus. After this period of "preview," further perceptual information about the target is removed by occluding vision and hearing, and the observer attempts to demonstrate knowledge of the target's location by some form of action (e.g. pointing, walking, or throwing a ball). *Visually directed pointing* was a term coined by Foley and Held (1972) to refer to blind pointing with the finger to the 3-D location of a visual target that had been previously viewed. This type of response has been used in other studies to measure the perceived locations of visual targets within arm's reach (e.g., Bingham, Bradley, Bailey, & Vinner, 2001; Foley, 1977; Loomis, Philbeck, & Zahorik, 2002). For more distant targets, ball or bean bag throwing has been used (Eby & Loomis, 1987; He, Wu, Ooi, Yarbrough, & Wu, 2004; Sahm et al., 2005; Smith & Smith, 1961). Another form of visually directed action, blind walking (sometimes called "open loop walking"), has been used to study the perception of distances of "action space" (distances beyond reaching but within the range of most action planning; Cutting & Vishton, 1995); here the observer typically views a target on the ground and attempts to walk to its location without vision. These various forms of open-loop behavior, along with others to be discussed, are referred to collectively as "perceptually directed action."

Many studies have used blind walking to assess the accuracy of perceiving the distances of targets viewed on the ground under full-cue conditions, for distances up to 28 m (Andre & Rogers, 2006; Corlett, Byblow, & Taylor, 1990; Corlett & Patla, 1987; Creem-Regehr et al., 2005; Elliott, 1987; Elliott, Jones, & Gray, 1990; Knapp & Loomis, 2004; Loomis et al., 1992; Loomis et al., 1998; Messing & Durgin, 2005; Rieser, Ashmead, Talor, & Youngquist, 1990; Steenhuis & Goodale, 1988; Thomson, 1983; Wu et al., 2004). Figure 1.2 shows many of the results, with the data sets shifted vertically for purposes of clarity. Except for two data sets, perceived distance is proportional to physical distance with no evidence of systematic error (slopes of the best fitting linear functions are generally close to 1, and intercepts are near zero). In contrast, when the same task, modified for audition, is used to study distance perception of sound-emitting sources heard out-of-doors, systematic errors are observed over the

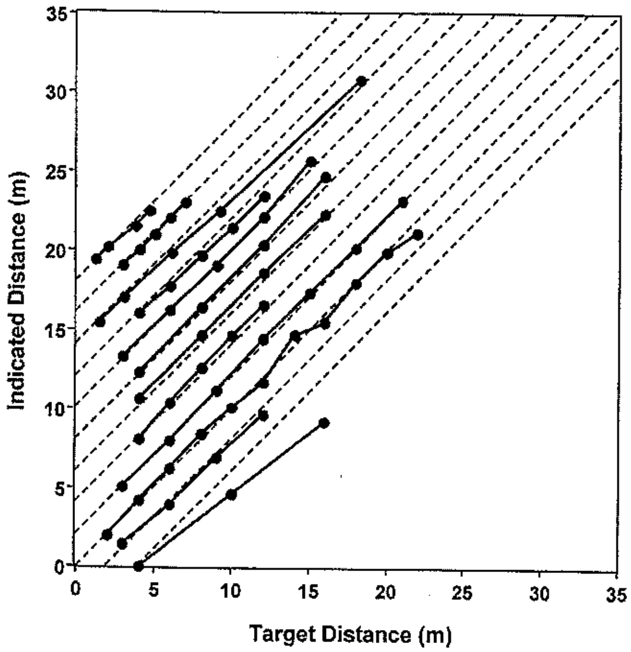

Figure 1.2 Summary of blind walking results for vision. The data from the different studies have been displaced vertically for purposes of clarity. The dashed line in each case represents correct responding. Sources from top to bottom: Experiment 3 of Philbeck and Loomis (1997), Experiment 1 of Wu, Ooi, and He (2004), mean data from the control conditions in the 3 experiments of Andre and Rogers (2006), results of Elliott (1987), average of two groups of observers from Experiment 1 of Loomis et al. (1998), Experiment 2b of Loomis et al. (1998), Experiment 1 of Loomis et al. (1992), Thomson (1983), Rieser et al. (1990), Steenhuis and Goodale (1988), and Experiment 2a of Loomis et al. (1998).

same range of distances (Ashmead, DeFord, & Northington, 1995; Loomis et al., 1998; Speigle & Loomis, 1993). Figure 1.3 shows representative results; this time, the data sets have not been shifted vertically. The best linear functions have slopes close to 0.5, indicating response compression relative to the stimulus range, and there is considerable variability in the intercepts. We should mention, however, that in a recent review of these and other results obtained using other response measures including verbal report, Zahorik, Brungart, and Bronkhorst (2005) fit power functions to the data and generally found exponents less than 1.0, the interpretation being that perceived auditory distance is a compressively nonlinear function of source distance. Still, for the range of distances in Figure 1.3, the conclusion that they are linear functions with roughly constant slope but

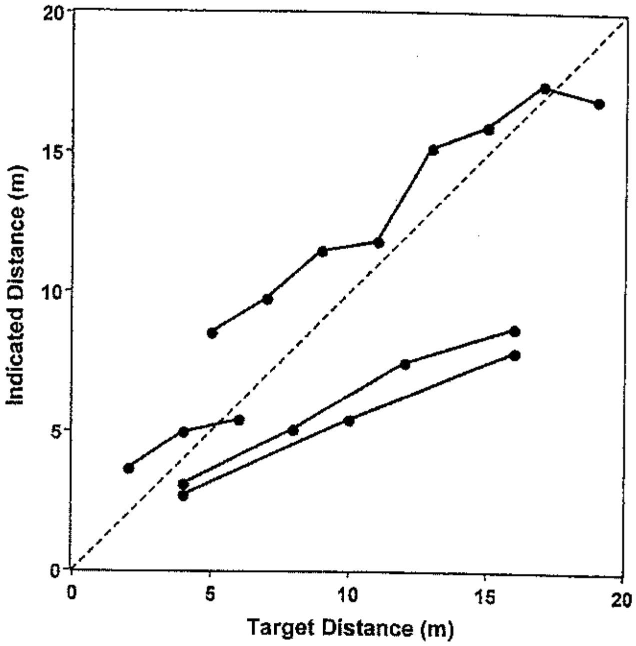

Figure 1.3 Summary of blind walking results for audition. These are the actual data and have not been displaced vertically for purposes of clarity. The dashed line represents correct responding. Sources from top to bottom: Ashmead et al. (1995), Speigle and Loomis (1993), Experiment 1 of Loomis et al. (1998), and Experiment 2a of Loomis et al. (1998)

varying intercept is justified. The source of the variation in intercept is a mystery.

With vision, systematic errors do arise. Sinai, Ooi, and He (1998) and He et al. (2004) have found that when the ground surface is interrupted by a gap, visual targets resting on the ground are mislocalized even with full cue viewing. Larger systematic errors occur when visual cues to distance are minimal. Figure 1.4 gives the results of a study by Philbeck and Loomis (1997) in which blind walking responses and verbal report were obtained under two conditions: reduced cues (luminous targets of constant angular size at eye level in the dark) and full-cues (the same targets placed on the floor with normal room lighting). When cues were minimal, both types of judgment showed large systematic errors, and when cues were abundant, both types of judgment showed small systematic errors. This study also showed that when the verbal responses were plotted against the walking responses in these and two other conditions, the data were

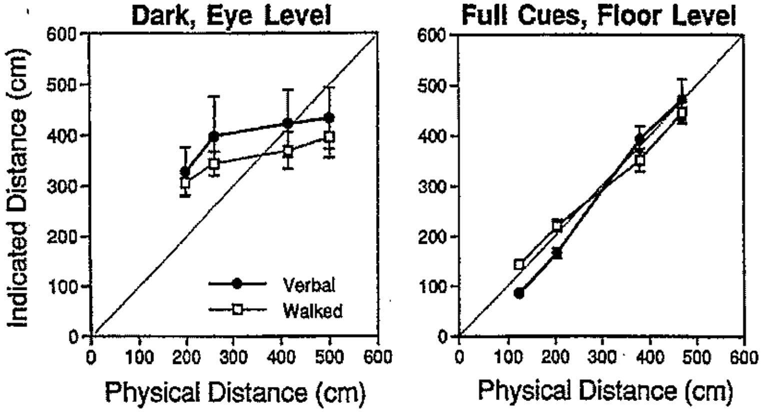

Figure 1.4 Results of an experiment using both verbal report and visually-directed blind walking in reduced-cue and full-cue conditions. Adaptation of Figure 5 from Philbeck, J. W. & Loomis, J. M. (1997). Comparison of two indicators of visually perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 72–85.

well fit by a single linear function, suggesting that variations in the two response measures are controlled by the same internal variable, visually perceived distance.

A related experiment was concerned with measuring the perceptual errors in visual virtual reality (Sahm et al., 2005). Observers performed blind walking and bean bag throwing to targets in both a real environment and a virtual environment modeled on the real environment. Prior to testing, observers were given feedback only about their throwing performance in the real environment. The results are given in Figure 1.5. The fact that the transition from the real to virtual environment produces the same errors in walking and throwing supports the claim that the two actions, one that involves locomotion and the other that does not, are controlled by the same internal variable, visually perceived distance. In addition, the results provide further support for the growing consensus that current virtual reality systems produce under perception of distance (Knapp, 1999; Thompson et al., 2004).

Triangulation Methods

The similarity of the walking measures and verbal reports above might be taken as evidence of a simple strategy for performing blind

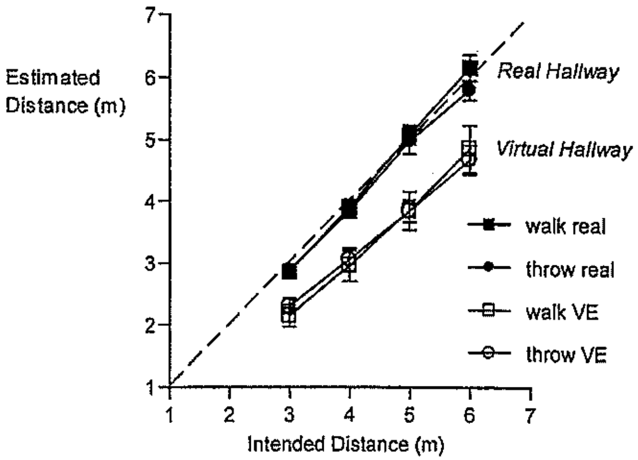

Figure 1.5 Results of an experiment using both visually-directed blind walking and visually directed throwing in real and virtual environments. Reprinting of Figure 3 of Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B., & Willemsen, P. (2005). Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception, 2*, 35–45.

walking—while perceiving the target, estimate its distance in feet or meters, and then, with vision and hearing occluded, walk a distance equal to the estimate.

Whereas the blind walking task might well be performed using this simple strategy, there are other closely related tasks that cannot. Foremost are triangulation tasks that require the observer to constantly update the estimated location of the target while moving about in the absence of further perceptual input specifying its location. Figure 1.6 depicts three triangulation tasks that have been used. In "triangulation by pointing," the observer views (or listens

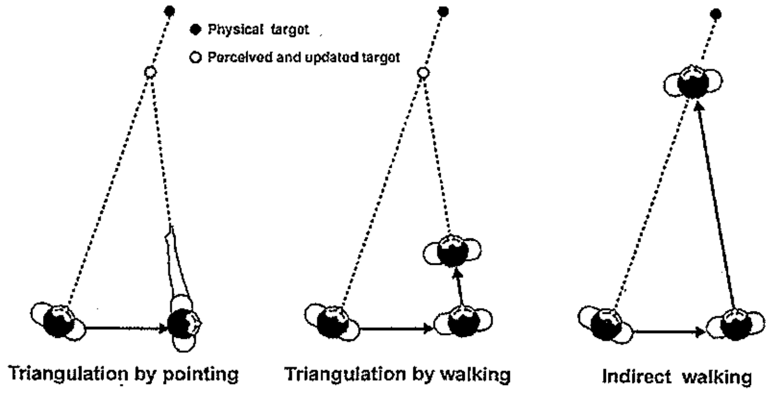

Figure 1.6 Three triangulation methods (see text for explanation).

to) a target and, then without further perceptual information about its location, walks along a straight path to a new location (specified by auditory or haptic signal) and then points toward the target. The pointing direction is used to triangulate the initially perceived and spatially updated target location. In one variant, the arm orientation was monitored continuously as the observer walked along a straight path (Loomis et al., 1992) after viewing a target on the ground up to 5.7 m away; the average pointing responses were highly accurate. "Triangulation by walking" (also, "triangulated walking") is similar to triangulation by pointing except that after the initial straight path, the observer turns and walks a short distance toward the target. The walking direction after the turn is used to triangulate the perceived and updated target location. Finally, in the "indirect walking" version of triangulation, the observer walks to a turn point (specified by auditory or haptic signal), and then attempts to walk the rest of the way to the updated target location. Figure 1.7 gives the results of a number of experiments using the different triangulation methods to measure the perceived distance of targets viewed under full-cue conditions (Fukusima, Loomis, & Da Silva, 1997; Knapp, 1999; Loomis et al., 1998; Philbeck et al., 1997; Thompson et al., 2004); as in Figure 1.2, the data sets have been vertically shifted for clarity. Although the data are more variable than with blind walking (Figure 1.2), they indicate overall that perceived distance is proportional to target distance with little systematic error.

A Model of Perceptually Directed Action

Because blind walking and the triangulation methods just mentioned rely on actions that occur after the percept has disappeared, it might be argued that these methods cannot be used to measure perception because the action depends upon postperceptual processes (e.g., Proffitt et al., 2006). However, we maintain that a valid measurement method is one for which the variations in the indicated values (those resulting from the measurement process) are coupled to variations in the variable being measured and for which a calibration between the two has been established (Hutchison & Loomis, 2006b). As with any measurement device (e.g., a thermometer with an electronic display), the indirectness of the mechanism between the variable being measured and the indicator has no bearing on whether the indicated values are proper measures of the variable of interest, here perceived distance. In the case of perceptually directed action, what is required is a theory linking the indicated value to perceived distance. Action can be used to measure perception provided that the postperceptual processes introduce no systematic biases or, if they do, that the biases can be corrected for by way of calibration. Of course, as with any measurement device or method, the precision of measurement will ultimately be limited by random noise associated with each of the subsequent processes, even if systematic biases can be eliminated by calibration.

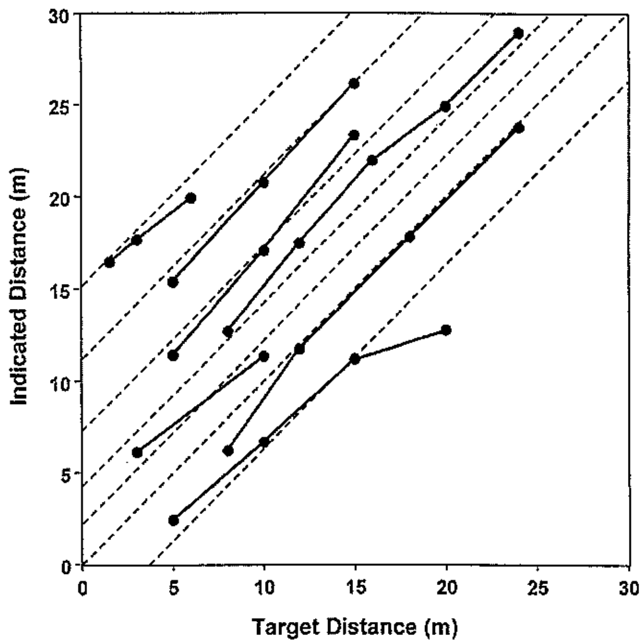

Figure 1.7 Summary of triangulation results for vision using triangulation by pointing, triangulation by walking, and indirect walking, all obtained under full-cue conditions. The data from the different studies have been displaced vertically for purposes of clarity. The dashed line in each case represents correct responding. Sources from top to bottom: results of indirect walking by Philbeck et al. (1997), results of triangulation by walking in real environment (Thompson et al., 2004), outdoor results of Knapp (1999), average of two conditions from Experiment 3 (triangulation by walking) of Fukusima et al. (1997), Experiment 3 (direct and indirect walking) of Loomis et al. (1998), average of two conditions from Experiment 4 (triangulation by walking) of Fukusima et al. (1997), and average of two conditions from Experiment 2 (triangulation by pointing) of Fukusima et al. (1997).

Here, we present a model of perceptually directed action that links the perceptual representation to the observed behavior. The model

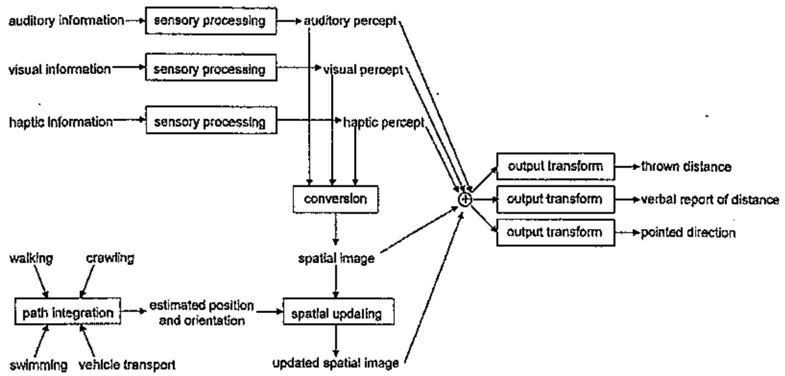

Figure 1.8 A block diagram of perceptually directed action (see text for explanation).

involves a number of processing stages (Figure 1.8). For similar models, see Böök and Gärling (1981), Loomis et al. (1992), Medendorp, Van Asselt, and Gielen (1999), and Rieser (1989). First, the visual, auditory, or haptic stimulus gives rise to the percept, which may or may not be coincident with the target location (Figure 1.9). Accompanying the percept is a more abstract and probably more diffuse "spatial image," which continues to exist in representational space even after the percept ceases. There is evidence that the spatial images from different modalities are functionally equivalent, perhaps even amodal in nature (Avraamides, Loomis, Klatzky, & Golledge, 2004; Klatzky, Lippa, Loomis, & Golledge, 2003; Loomis, Lippa et al., 2002). We assume that the spatial image is coincident with the percept, but future research may challenge this assumption; for now, it appears that error in the percept is carried over to the spatial image. When the actor begins moving, sensed changes in position and orientation (path integration) result in spatial updating of the spatial image (Böök & Gärling, 1981; Loarer & Savoyant, 1991; Loomis et al., 1992; Loomis, Klatzky, Golledge, & Philbeck, 1999; Rieser, 1989; Thomson, 1983). At any point in the traverse, as depicted in Figures 1.8 and 1.9, the observer may be asked to make some nonlocomotion response, such as pointing at or throwing to the target or verbally reporting the remaining distance. The response processes clearly are different for different types of response. An important assumption, to be discussed later, is that the response is computed in precisely the same fashion whether based on the concurrent percept of the target or on the spatial image of the target (whenever the percept is absent). This assumption is depicted in Figure 1.8 by the convergence of percepts, initial spatial images, and updated spatial images onto the output transforms for different types of responses.

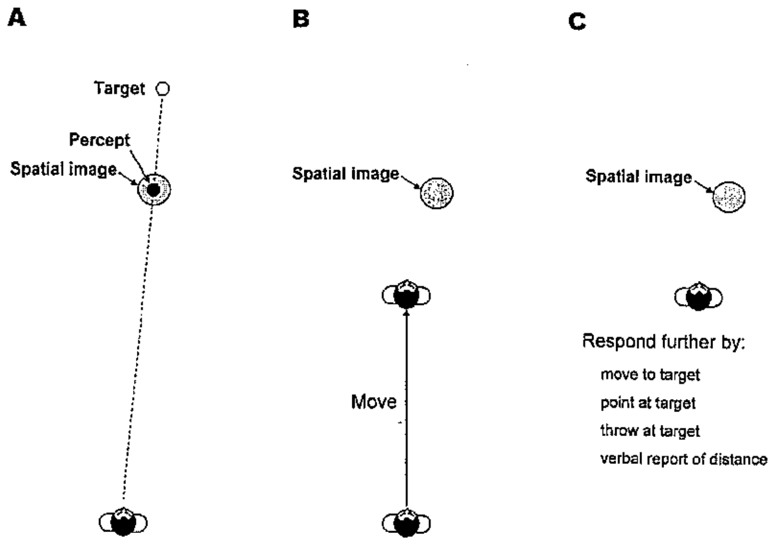

Figure 1.9 From left to right, depiction of 3 successive moments during perceptually directed action. A. The observer perceives a target closer than its physical distance. Accompanying the perceived target is a more abstract and spatially diffuse spatial image. B. With the stimulus and its percept no longer present, the observer moves through space, updating the egocentric distance and direction of the target. If path integration is accurate (as depicted here), the spatial image remains stationary with respect to the physical environment. C. After moving, the observer can make another response to the updated spatial image, by continuing to move toward it, by pointing at it, by throwing at it, or by making a verbal report of the distance remaining.

Not depicted in Figure 1.8 is a nonperceptual input to the creation of a spatial image. Loomis and his colleagues (Avraamides et al., 2004; Klatzky et al., 2004; Loomis, Lippa et al., 2002) have shown that once a person forms a spatial image, whether based on a spatial percept or based on spatial language, subsequent behaviors (like spatial updating and exocentric direction judgments) appear to be indifferent to the source of the input, suggesting that spatial images based on different inputs might be amodal. The implication is that the spatial image produced by vision, hearing, or touch, can, in principal, be modified by higher-level cognition so as not to be spatially coincident with the percept. Whether this dissociation between percept and spatial image ever occurs remains to be determined, but the evidence to be reviewed is consistent with the assumption of coincidence.