

This excerpt from

Indirect Perception.
Irvin Rock.
© 1997 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.

Chapter 1

The Concept of Indirect Perception

Irvin Rock

Over a lifetime of research I have often had reason to ask a question about the immediate determinant of a particular perceptual phenomenon. The reason was that the meaning of the determinant under study was ambiguous. The first time I ran across this question concerned the altered appearance of objects and their consequent loss of recognizability when their orientation in a frontal plane was changed. The ambiguity concerned the meaning of "orientation." Was it the altered orientation of the object's image on the retina that mattered or the altered orientation of the object as it was perceived in the scene? Ordinarily, of course, both of these changes occur concomitantly, but I realized that they could be separated. One way of doing so was to require an observer to view a novel object from a tilted or inverted orientation (of the whole body or just the head), after having previously seen it from an upright position. Given information from gravity or from the visual frame of reference, the location of an object's top, bottom, and sides would appear unchanged in the scene, but the orientation of its retinal image would undergo a change. Conversely, one could tilt the object in the second viewing by the same angle as the body or head was tilted. In this case, the orientation of the object in the environment would change (and presumably would be perceived to do so), but the orientation of the retinal image would not.

My colleagues and I performed experiments along these lines to ascertain which meaning of orientation was relevant to the effect: the retinal or the phenomenal. The results of the experiments are described in chapter 9 and need not be discussed here except to say that what mattered for shape recognition was phenomenal orientation. It turned out that this same question could be asked about any number of other phenomena because a similar ambiguity arises in trying to understand perceptual organization, the perception of shape (apart from the effect of orientation on shape), the nature of masking, the perception of lightness, the perception of motion, and the perception of illusions. Each of these topics is addressed in a part of this volume.

Given the emergence of a clear question in science, curiosity alone is a sufficient reason for pursuing it and, at the beginning, I was driven to a

large extent by curiosity. However in each case, I also realized that the answer to the retinal/phenomenal question had important theoretical consequences. Consider the ambiguity that arises in the case of perceptual grouping. The Gestaltists proposed a number of “laws” of grouping, concerning why certain elements grouped with one another. One such law was that of proximity: all else being equal, elements that are nearer to one another than others would group into larger structures. This was taken to mean that the corresponding retinal images of the elements in the array, being more proximal to one another—and thus projecting representations to regions of the visual cortex that were nearer to one another, relatively speaking—tended to attract one another more strongly than cortical representations that were farther apart. This was the Gestalt view, although the Gestaltists were not explicit about it beyond the just-mentioned brain metaphor, because they were unaware of the retinal-phenomenal ambiguity.

The ambiguity arises when one realizes that the elements whose retinal images are closer to one another are also usually *perceived* to be closer together. The “law” could therefore be reformulated as follows: elements *perceived* to be closer together than others tend to be grouped into larger structures (for example, into columns or rows). Now, although it isn’t obvious when the array is in the frontal plane that the perception of proximity entails more or different processing than just the proximity given by the projection of their images to the visual cortex, the fact is that such further or different processing is implicated. To make this clear, imagine the array of elements to be small objects (e.g., tiny spheres) in a rectangular lattice such that they are closer together vertically than horizontally. When viewed head-on in the frontal plane, the observer tends to see columns rather than rows. Suppose now one views the entire array from the side so that, by virtue of perspective foreshortening, the projection of the spheres to the retina (and cortex) is altered. If one is far enough to the side, the sphere’s images will be closer together horizontally (see figure F.3). However, given adequate depth perception leading to constancy (i.e., veridical perception), the relative position of the spheres vis-a-vis one another will continue to appear as it did when viewed from directly in front. If so, the spheres would still appear to be closer together vertically, and thus one would continue to perceive columns rather than rows. This would occur as long as constancy held and the correct “law” of proximity is the reformulated one based on *perceived* proximity.

This experiment was performed by Brosgole and myself (see chapter 3). The results implicate perceived or phenomenal proximity rather than retinal proximity, as was the case for perception of shape after changes in perceived versus retinal orientation. The same question has now been asked about other Gestalt principles of grouping, and analogous experi-

ments have been performed with the same general result: grouping appears to be governed by perceived rather than retinal factors. Chapters 3, 4, and 5 describe these experiments.

Why is the realization of this ambiguity and a clarification of the question it poses important? The answer is that it tells us about the level or stage of processing that underlies the phenomenon of grouping by proximity. At the time of Wertheimer's classic paper on the laws of grouping in 1923, questions about "level" were rarely, if ever, asked. It was with the later advent of information processing as an approach to perception and cognition that such questions became central. However, it was implicit for Gestalt psychology that perceptual grouping occurred at a very early stage of processing because, until the scene before us became organized, no further perception could occur. The achievement of discrete, segregated units in the field was held to be the end product of organization, and thus organization was the *sine qua non* for all subsequent perception.

Interestingly enough, in the years that followed Wertheimer's publication, the assumption that grouping occurred at a very early stage remained an article of faith by subsequent investigators and later was explicitly stated by those favoring an information-processing approach. The interest had shifted to attention, and it was held that perceptual grouping was *pre-attentive*. It was argued that attention required the prior phenomenal existence of units in the field to which it could be deployed or not, as the case may be (e.g., Neisser 1967; Treisman 1986).

However, given the findings that grouping conforms to *perceived* structure and therefore must emerge at a later stage following depth perception and constancy operations, important new questions arise. Given the logical argument that perceptual organization *must* occur at an early stage on the one hand and, given the *finding* that Gestalt grouping does *not* occur at an early stage on the other, then what principle or principles govern initial perceptual organization? Do the Gestalt laws play any role in the final organization of the scene or are they mere laboratory epiphenomena? The entire question of grouping has to be reopened and reexamined, as Palmer and Rock (1994a, 1994b) have recently done.

So here we see, in one example at least, the importance of (1) recognizing that there is an ambiguity in the meaning of a determinant (in this case "proximity"), (2) attempting to disambiguate the meaning, and (3) attempting to clarify the process underlying the phenomenon in question. Clearly the "retinal" answer to the question points in one theoretical direction and the "phenomenal" answer points in a very different theoretical direction.

I will give only one other example here of the retinal-phenomenal ambiguity in the determination of a phenomenon in order to bring out the theoretical importance of knowing the answer. It concerns the perception

of lightness: the appearance of surfaces along the white-gray-black continuum. For roughly the last half century most investigators of this topic have been favorably disposed to the theory that it is not the absolute intensity of light (or luminance) reflected by a surface that governs its apparent lightness, but the ratio of the intensities from adjacent regions (e.g., Land 1977; Wallach 1948). This theory seemed to explain not only lightness perception, but also lightness constancy: the fact that the apparent lightness of a surface does not change with increments or decrements of the intensity of light that illuminates the surface. Since such changes of overall light intensity typically affect adjacent regions equally, the ratio of the light they reflect remains more or less constant despite widely varying levels of illumination.

The ambiguity in this case concerns the meaning of "adjacency." Adjacency could mean the contiguity of the retinal images of two (or more) regions. However, it could also mean the *perceived* contiguity of the regions giving rise to these retinal images. The two meanings of adjacency can be teased apart, although, more often than not, it is true that two adjacent retinal regions will result from two regions in the environment that are adjacent to one another. However, one has only to think of separating the regions in depth to realize that such spatially separated regions can still project to the retina in such a way as to be side by side, so to speak.

It makes a good deal of intuitive and theoretical sense to believe that the ratio of the luminance of two adjacent *retinal* regions would determine perceived lightness and therefore, lightness constancy, whether or not they are reflected from environmentally adjacent regions. A priori, it does not seem very likely that the ratio of *phenomenally* adjacent regions would have any such effects merely because they *appear* to be next to one another. Be that as it may, as suggested above, one can easily create a situation in which region 1 is at a different depth or in a different depth plane than region 2 while their retinal images are nonetheless adjacent. In this case we can arrange to have retinal adjacency but not phenomenal adjacency as shown in figure 1.1.

It has been found that when such an arrangement is created and one allows the difference in depth or planarity to be perceived, the luminance ratio between regions 1 and 2 no longer governs perceived lightness. By the simple maneuver of eliminating cues to depth, while holding everything else about the display and the observer's position constant, a drastically different outcome regarding perceived lightness occurs because now the two regions *are* perceived to be phenomenally adjacent. Experiments of this kind will be described in chapter 22 of this volume (and see Gilchrist 1980) so no further discussion is required now except to emphasize that here again it is the phenomenal factor, not the retinal, that matters.

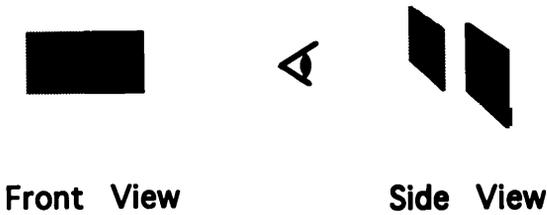


Figure 1.1

Retinally adjacent regions can be either phenomenally adjacent or separated in depth.

What is the importance of realizing the ambiguity of “adjacency”? The theoretical implications are profound. If what matters for the determination of perceived lightness by the ratio of the luminances of adjacent regions is that the *retinal images* of these regions be adjacent, one kind of theory is implicated. In fact, based on this interpretation of adjacency (which has been implicit or explicit for well over a century), one might think that a process of lateral inhibition between retinally adjacent regions is responsible for the outcome (e.g., Cornsweet 1970). Indeed *contrast* has been explained in terms of lateral inhibition, and as the luminance of region 1 increases when it is seen in sunlight instead of shadow, the luminance of region 2 increases equally, and thus its lateral inhibition of region 1 increases. Thus the net change in rate of firing of neurons from region 1 would be said to remain about the same. Hence constancy! Other such lower-level or sensory kinds of theory of lightness constancy are also possible. But if what matters is that the adjacency must be *phenomenal*, then a very different kind of theory seems required, one that is more central and that entails cognitive operations. For example, the particular lightness of each region must be inferred by where its luminance stands relative to that of other regions, particularly relative to the region of highest luminance that is taken to be white and thus serves as an anchor for the various regions of differing luminance (see Gilchrist 1995).

So much for illustrations of ambiguity concerning the meaning of a determinant of a particular perceptual phenomenon and for why resolving that ambiguity is so important. As can be seen from these illustrations and many others to be discussed in this volume, however, the ambiguity is generally resolved in favor of the phenomenal rather than the retinal aspect of the stimulus conditions. This fact has further, deep implications that form the central thesis of this volume. If it is the *perceived* character of the stimulus array that matters in regard to the phenomenon under consideration—be it perceived orientation, perceived proximity, perceived adjacency, or whatever—then such perception becomes an earlier stage of processing that culminates in a later, final perception. Perceived orientation is the precondition for perceived shape, perceived proximity is

the precondition for the perception of grouping by proximity, perceived adjacency is the precondition for the perception of lightness, and so forth.

Such conclusions are important because they show that *one perception depends upon another, prior perception*. We thus have a perception → perception chain of causation which means, among other things, that the final perception, the one we are generally seeking to explain, is *indirect*. It cannot be said to be the *direct* result of a proximal stimulus, as James J. Gibson (1950, 1966, 1979) maintained, no matter how complex or higher-order we take that proximal stimulus to be. As to a possible philosophical objection to the claim that one perception can cause another, there is a simple rejoinder. The initial perception is itself caused by events occurring in the brain (although at this stage of our knowledge of the brain we usually don't know precisely what neurophysiological process accounts for it), and so is the final perception for that matter. So the perception → perception chain of causality can be translated to a brain-event-1 → brain-event-2 chain of causality. No philosophical problems are entailed by this formulation of the "indirect perception" thesis because all the events in question are ordinary physical processes.

However, this translation into talk about brain events should not be taken to weaken the claim I am making here. One might want to argue that the claim simply reduces to the argument that a somewhat longer, more complex sequence of brain events underlies (or causes) a particular perception than has heretofore been recognized for the phenomenon under consideration. But that would be to miss the point. The first brain event in the sequence is the neural correlate of the initial percept; the second brain event is the neural correlate of the final percept. They are qualitatively and ontologically distinct events. Moreover, one must not forget about the intentional "content" if one chooses to translate the argument about a perception → perception chain of causation into the language of neurophysiology. So, for example, if the initial perception is one of the phenomenal orientation of an object, then that "content" must somehow be represented by the initial brain event; the same holds true for the "content" of the later perception of the object's shape and its representation in the second brain event.

That the content of the initial perception is crucial—rather than a particular pattern of sensory stimulation—is seen by the fact that there are often many routes to its achievement. For example, if the perceived orientation of an object is the important first stage in the achievement of its unique phenomenal shape, then we need to recognize that there are various stimulus conditions or proximal stimuli that can lead to that perceived orientation. The direction of gravity alone can be responsible because experiments have shown that a luminous object in an otherwise dark room will (1) appear to have a top, bottom, and sides as governed by

gravity, and (2) appear to have a shape that is, in part, a function of (1). The visual frame of reference alone can be responsible for the outcome because experiments have shown a powerful effect on perceived orientation (and therefore on perceived shape) of a large tilted frame even while gravity remains unchanged. Egocentric visual coordinates can lead to these effects too because a change of an object's orientation in a *horizontal plane* (which eliminates gravitational information), viewed through a circular aperture that eliminates external reference frames, will lead to an altered perception of its top and bottom, and therefore to an altered perception of its shape.

A similar argument holds for cases where the perceived depth of an object is the first step in the perception of some other property such as grouping or lightness. There are many different kinds of cues to depth, and it is easily shown that one can be substituted for another to achieve the initial percept. So, in summary, it is not appropriate or parsimonious to reduce perception to some proximal stimulus, as theorists favoring a direct theory of perception might wish to do, since what matters is the phenomenally perceived property, however that may be achieved.

A possible criticism of the claim presented here of one perception depending upon a prior perception is that the two perceptions may be temporally co-occurring rather than successive. In the case of lightness perception as affected by perceived adjacency in the third dimension, for example, one might argue that all one is entitled to say is that a given depth perception *is intimately associated with* a given lightness perception rather than that a given depth perception *causes* a given lightness perception. However, if the claim I am making were only one of association, then a form of symmetry should follow:¹ manipulating perceived lightness should be just as likely to yield changes of perceived depth as manipulating perceived depth should yield changes of perceived lightness. Yet that is not generally the case. If one introduces stereoscopic information that two regions are not in the same plane, change in the perceived lightness of one or both regions will occur; however, introducing a change in the perceived lightness of two such regions does not usually affect or create perceived depth. To be sure, there are cases where luminance differences based on attached shadows can create an impression of depth, as in the homogeneous surface lightness of a sculpture or a crumpled handkerchief. But here it is luminance differences that have the effect on perceived depth, not perceived lightness differences.

In other examples, irreversibility of the chain of causation is even clearer. Does it make sense to say that the achievement of grouping determines the perception of proximity or similarity? Does it make sense to say that certain perceived distortions in the size and shape of a three-dimensional wire cube, when the cube rotates (or we walk around it), cause perceptual

reversal to occur? Rather, it is well known that these distortions occur because the perceived reversal of the cube necessitates a reinterpretation of the sizes and shapes of the retinal images of the cube. Ordinarily the far face yields a smaller image than the near face and is corrected for appropriately on the basis of perceived depth. When a regular cube is seen with reversed depth, the face yielding the *larger* image is perceived as farther away, so that constancy operations enlarge this face even more, yielding the perception of a highly irregular cube.

Finally, it would seem that even if, despite my arguments above, one could not establish a unidirectional sequence of events, it remains indisputable that a complex interactive relationship must obtain between a certain perceived character of the proximal stimulus and a further perception. Given the evidence to be presented in this volume, it is simply not possible to argue that perception is *direct*, if one takes this to mean the determination of perception by the proximal stimulus without mediating perceptual events.

Two Kinds of Indirect Perception

There seems to be another kind of indirect perception that ought to be distinguished from the kind I have been discussing thus far. In the kind I have been discussing, the two perceptions that occur are not on different levels or stages of processing, nor do they imply a lower-level–higher-level sequence. They have equal status. Thus, for example, grouping *per se* is not a higher stage of processing than perceived proximity or similarity. In such situations grouping *depends* upon or makes use of perceived proximity or similarity. Moreover, the occurrence of the first perception does not lead automatically to the second perception. The first perception is a necessary condition for the occurrence of the second, but it isn't a sufficient condition.

So, for example, achieving the perception of the directional coordinates of an object does not, *ipso facto*, lead to the perception of the object's phenomenal shape. It is a necessary step because assigning such coordinates differently will give rise to a very different phenomenal shape. But without focusing attention on the object, no consciously perceived shape may occur at all. On the other hand, the awareness that an object is present and the awareness of where its top and bottom are, no doubt do occur without focused attention.

The other kind of indirect perception refers to instances in which the final experience is the end result of a two-stage sequence. The existence of the first stage is in part speculative because often one is not fully aware of it. A good example of this kind of indirectness is amodal completion. In figure 1.2 one sees a circle, part of which is occluded by a square per-

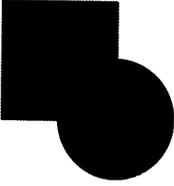


Figure 1.2
Amodal completion of a circle partly occluded by a square.

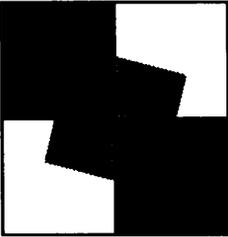


Figure 1.3
Phenomenal translucency of a square region through which a black and white checkerboard is perceived.

ceived to be in front of it. I have suggested elsewhere that, prior to the occlusion/completion perception, another occurs—however fleetingly—in which one sees literally what is represented in the figure: namely, a three-quarter circle adjacent to a rectangle in a kind of mosaic arrangement. There is now some experimental evidence supporting the existence of this literal stage (e.g., Gerbino and Salmaso 1985; Sekuler and Palmer 1992).

Another example of this kind of indirectness of perception is that of phenomenal translucency. In figure 1.3 one perceives a checkerboardlike background of black and white regions with a tilted translucent square in front, through which one also sees the background. But, prior to that perception, another may occur in which the tilting square consists of four different regions of differing perceived lightness. Because this perception correlates perfectly with the differing luminances in the entire figure, I have suggested that in this case as well, one might describe it as “literal.” By way of contrast, the amodal completion in figure 1.2 and the translucency “solution” in figure 1.3 are more creative and depart from the proximal stimulus in a one-to-one correspondence. I have suggested the terms *world mode* or *constancy mode* to describe these presumably later perceptions.

I imagine the reader can appreciate that the sequence from perception 1 to perception 2 in these last examples is different from the sequence of two perceptions discussed throughout the early part of this chapter. But how to characterize the difference? First, the sequence in these examples represents two perceptions arising from the same proximal stimulus, which therefore can be characterized as ambiguous. Second, in these new examples, the transition is from a lower-level, earlier-stage processing to a higher-level, deeper, final stage of processing. In fact, once the world mode perception occurs—in cases where it did not do so immediately—it is more or less irreversible. It is all but impossible not to see it this way just as it seems to be psychologically irreversible not to recognize, let us say, a word, even though it must be the case that there was a moment in time prior to recognition in which it was not yet identified. There clearly is a strong *preference* for the world mode solution. However, I hasten to add that such a preference does not rule out the occurrence of the literal mode stage *along with* the world mode stage. So, for example, in perceiving the amodally completed circle in figure 1.2, one can continue to be aware that the circle is not completely visible in the retinal image.

There is much more that can be said about this second kind of sequence from one perception to another, but hopefully what I have said will suffice to distinguish it from the first kind of sequence. Yet it seems correct to say that in both cases, perception is indirect. In the first case—let us call it type A—the indirectness comes about because it is simply not the proximal stimulus per se that can be said to yield the perception we are trying to explain. That stimulus must first give rise to a certain perception, be it of orientation, of proximity, of depth, etc., before the final perception can be expected to occur. The second kind of indirectness—call it type B—comes about because some ambiguous stimuli are first perceived on the basis of their literal correspondence to the proximal stimulus. However, that perception or “solution” is inadequate for some reason, so that perception is superceded. (Notice that being *superceded* is not the same as being *replaced* because the prior interpretation still exists.) One reason for the inadequacy of the first perception is that that perception requires acceptance of a coincidental or accidental relation between the proximal stimulus and the state of affairs in the world. But to pursue this important question would carry us too far afield.

Although the focus of this book concerns type A indirectness and virtually all the examples to be covered are of this type, it was important to make the distinction to avoid possible confusion. However, the realization that there is type B indirectness certainly strengthens the main thesis—namely that perception, more often than not, cannot be understood as a direct “response” of the visual system to the proximal stimulus. In fact, it becomes a challenge to find examples in which perception is direct,

and having found them—assuming they exist—to ask what they have in common.

Note

1. I wish to thank Steve Palmer for pointing out the importance of asymmetries of this sort in establishing perception-perception causal chains.

References

- Cornsweet, T. N. 1970. *Visual perception*. New York: Academic Press.
- Gerbino, W., and D. Salmasso. 1985. Un'analisi processuale del completamento amodale. *Giornale Italiano de Psicologia* 12:97–121.
- Gibson, J. J. 1950. *The perception of the visual world*. Boston: Houghton Mifflin.
- Gibson, J. J. 1966. *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gilchrist, A. 1980. When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics* 28:527–38.
- Gilchrist, A. 1995. Local and global processes in surface lightness perception. *Perception & Psychophysics* 57:125–35.
- Land, E. 1977. The retinex theory of color vision. *Scientific American* 237:108–28.
- Neisser, U. 1967. *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Palmer, S. E., and I. Rock. 1994a. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1(1): 29–55.
- Palmer, S. E., and I. Rock. 1994b. On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin and Review* 1:515–19.
- Sekuler, A. B., and S. E. Palmer. 1992. Visual completion of partly occluded objects: A micro-genetic analysis. *Journal of Experimental Psychology: General* 121:95–111.
- Treisman, A. 1986. Properties, parts, and objects. In *Handbook of perception and human performance*, vol. 2, edited by K. R. Boff, L. Kaufman, and J. P. Thomas, 35-1–35-70. New York: Wiley.
- Wallach, H. 1948. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology* 38:310–24.
- Wertheimer, M. 1923. Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, 4:301–350. Partial translation in *A sourcebook of Gestalt psychology*, edited by W. D. Ellis, 71–81. New York: The Humanities Press, 1950.

This excerpt from

Indirect Perception.
Irvin Rock.
© 1997 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.