Mental Representations; psychology of

C. R. Gallistel[1]
[1] Rutgers University, New Jersey, USA

Abstract

A mental representation is a system of symbols isomorphic to some aspect of the environment, used to make behavior-generating decisions that anticipate events and relations in that environment. A representational system has the following components: symbols, which refer to aspects of the environment, symbol processing operations, which generate symbols representing behaviorally required information about the environment by transforming and combining other symbols, representing computationally related information, sensing and measuring processes, which relate the symbolic variables to the aspects of the world to which they refer, and decision processes, which translate decision variables into observable actions. From a behaviorist perspective, mental representations do not exist and cannot be the focus of a scientific psychology. From a cognitivist perspective, psychology is the study of mental representations, how they are computed and how they affect behavior.

B00804307 67XXXXXX

Proofreader Please Note:
Margin marks (in the left margin as the example below) indicate a link from the text reference to an entry in the bibliography. These must be checked against bibliography, and where wrong they must be corrected.

B13 morality (Austin $\boxed{1954}$, p. 112). The second sense of morality is found to be indicative of changes of state and not part of the basis of law.

Encyclopedia of the Social and Behavioral Sciences
ISBN: 0-08-043076-7

Mental Representations; psychology of

sx0050

!QA: The copyeditor report gives this as University of California: please give correct affiliation

!QA:M

Mental representations were banned from scientific psychology by the behaviorists. They came back into psychology during the so-called cognitive revolution, when information processing models came to dominate psychological theorizing.

sx0015 They were banned by the behaviorists for two reasons. First, they are not directly observable; they must be inferred from their observable behavioral consequences. Radical behaviorists believed that inferred entities had no valid role to play in a scientific B17 B18 B19 psychology (Skinner 1938, 1950, 1990). Second, mental representations are not neurobiologically transparent: it has been and remains difficult to say how the entities and processes central to many kinds of hypothesized mental representations might be realized by currently understood neurobiological processes and structures. Not surprisingly, efforts to eliminate mental representations from psychological theorizing have often been driven by a desire to anchor psychological theorizing in neurobiology. (See, for example, B3 B9 B10 Edelman & Tononi 2000; Hull, 1930, 1952; Rumelhart B14 & McClelland, 1986.)

sx0020 Another difficulty is that it has not always been clear what cognitive psychologists understood by the term representation. This lack of clarity is due to the inherent complexity and abstraction of the concept. Although mental representations are central to prescientific folk psychology, folk psychology does not provide a rigorous definition of representation, any more than folk physics provides a rigorous definition of mass and energy. Representation, rigorously defined, is a mathematical and computational concept.

sx0025 The cognitive revolution was closely tied to the emergence of computer science because computer science created indubitably physical machines that unequivocally computed. This dispelled the widespread belief that computing was an inherently mental activity in the dualistic sense—mental and therefore not physical. More importantly, computer science led to a deeper understanding of what it meant—from a physical and mathematical perspective—to say that B20 something computed (Turing 1936). Computation became an object of mathematical thought rather than merely a tool of such thought.

sx0030 A representation, mental or otherwise, is a system of symbols. The system of symbols is isomorphic to another system (the represented system) so that conclusions drawn through the processing of the symbols in the representing system constitute valid inferences about the represented system. Isomorphic means ‘having the same form.’ The form in question is mathematical form, the forms of the equations specifying the relations among the symbols and among the things that the symbols represent. For example, Ohm’s law—$I = V/R$—which is the equation for the relation between current (I), voltage (V) and resistance (R) in an electrical circuit, has the same form as the equation for the relation between speed (S), force (F) and viscous resistance (R) in a mechanical system like a shock absorber—$S = F/R$. The identical form of these two equations is suggestive of the much broader isomorphism (mathematical equivalence) between electrical, mechanical, hydraulic and thermal systems that gives rise to linear systems theory in engineering.

sx0035 The symbols in the above two equations differ, but that is only to remind us that the physical variables they refer to differ. The important thing is that the equations that describe the two systems are the same. Because the forms of the relations the variables enter into are the same, we can represent a mechanical system with an electrical system (and vice versa). And we can represent either of them with a paper and pencil system that we endow with a suitable mathematical form. What matters in representations is form, not substance.

sx0040 The symbols in an information processing system (a symbol system) have two fundamental properties: they refer to things outside the system and they enter into symbol processing operations. The symbol processing operations in the above examples are the operations of arithmetic (V divided by R) and the rewrite rules (rules of algebra) dictated by the principles that define the system of arithmetic operations.

sx0045 We believe that we understand simple electrical circuits because the inferences we draw from manipulating the symbols on paper correctly predict what we observe when we make the corresponding manipulations of the electrical circuit itself. Thus, for example, simple algebra, allows us to deduce from $I = V/R$ that $IR = V$. When we measure I and R and compute the numerical product of the two measurements, the number we get turns out to be the same number that we get when we measure V. Our paper and pencil representation of the electrical circuit, which includes both the symbols themselves and the rewrite rules that we observe in deriving $IR = V$ from $I = V/R$, correctly predicts the results of the measurements that we make on the circuit itself.

sx0050 The above example of a symbolic system contains three distinct contrivances—symbols, rules that govern the manipulation of those symbols, and measuring processes. The measuring processes relate the numerical values of the symbols to the voltages, resistances and currents to which they refer. Because these are obviously human contrivances, it might seem that representations are artifacts of a purely human manner of interacting with the world, requiring perhaps some form of consciousness. However, the same three contrivances are present in a process control computer. Such a computer is also a human contrivance, but it interacts with the world without human intervention. It measures physical variables using digitizing transducers, symbolizes those variables by means of bit patterns in its memory banks, manipulates those symbols in accord with the applicable rules of algebra and physics, and uses the results to control observable actions—all without human intervention.

sx0055 The cognitive revolution was predicated on the possibility that brains—both human and animal—are process control computers contrived by evolution through natural selection. They assess their environment through sensory or perceptual processes; they symbolize the results of these assessments by values stored in memory; they manipulate those values by means of the relevant mental operations (the operations of perception and thought); and they use the results (percepts, inferences and deductions) to control behavior. If so, then mental representations are the very stuff of psychology. A psychology without mental representations is no more possible than a physics without masses and energies.

Thus, from the standpoint of a cognitivist, psychology is the science of mental representations. The essential questions in psychology are: What representations does the mind compute? From what data does it compute them? How does it compute them? How does a given representation get translated into observable behavior?

1. *Information*

sx0060 An important development in the mathematical treatment of computation and representation was the rigorous definition and quantification of the information
B16 carried by signals (Shannon 1948). Signals are symbols that convey information from one location in space and time to another—like, for example, the nerve signals that carry information about the environment from sensors in the periphery to the brain. The amount of information conveyed by a signal is a function of the amount of information about the world already present at the site where the signal is received and processed. When a digitizing thermometer sends a bit pattern to a computer specifying the temperature of a fluid, the signal, that is, the transmitted bit pattern, conveys information about the environment to the computer. The less information the computer already has about the temperature, and the more precisely the bit pattern received specifies what the temperature is, the more information is
B13 conveyed by the signal (See Rieke et al., 1997 for the rigorous development of these ideas in the analysis of neural signaling.).

sx0065 This idea that the amount of information conveyed by a signal is measured by the amount by which the signal reduces the receiver's uncertainty about the state of the world is highly intuitive: If we already know that the temperature is 70°, then a signal indicating that it is 70° tells us nothing. This simple idea has, however, non-intuitive mathematical consequences. It implies, for example, that signaling presupposes prior knowledge on the part of the receiver regarding the range of possibilities. If the mind of the newborn baby is truly a blank slate, with no beliefs about what temperature its environment might have, then any signal that tells it what the temperature is—even if imprecisely—conveys an infinite amount of information. In information theory, no signal can convey an infinite amount of information in a finite amount of time. Thus, in order for us to acquire information about the world from our experience of it, we must have built into our information processing structures implicit representations of the range of environments that could be encountered. We must know in advance something about the world we are to experience.

sx0070 This implication of information theory, together with the consideration that the machinery of computation itself seems unlikely to arise from the impact of experience on a system not endowed with some initial computational capacity, gives a nativist cast to information processing theories of mind. If the brain is fundamentally an organ of computation devoted to the computation of the mental representations that enter into the decisions leading to actions, then it does not seem that it could get up and running without a non-trivial amount of genetically specified structure, much of which contains implicit knowledge about the world to be represented. That is why extreme empiricists tend to be anti-representational: they tend to reject the cognitivist assumption that the mental representations are the stuff of psychology.

2. *Decision Processes*

sx0075 Symbols are translated into observable behavior by means of control variables and decision processes. A control variable specifies a parameter of an action, for example, the angle that is to be maintained with respect to a directional reference stimulus like the sun illustrative example below). (see illustrative example below A decision variable is a computed symbolic value representing some aspect of the current environment that merits a response just in case it exceeds some criterion, called the decision criterion. The analysis of decision processes in modern psychology has been heavily influenced by statistical decision theory, which treats the structural and formal features of decisions made in the face of ambiguous B8 information (Green & Swets 1966). The information about the world carried by symbols is ambiguous for two reasons: First, the processes that generate the symbolic values are inherently and inescapably noisy. The temperature of the fluid cannot be exactly known and hence it cannot be known with certainty whether an environmental variable actually does exceed some criterion; it can only be known with varying degrees of probability. Thus, decisions are inherently statistical sx0080 in nature. Optimal decision processes must take account of the statistical uncertainty about the true value of critical variables. Second, one and the same sensory input may be generated by more than one state of the world. For example, radically different arrangements of surfaces in the three-dimensional environment can produce identical images when projected on to the two-dimensional retina. In computing a representation of the three-dimensional environment that generated these inputs, the brain must be sensitive to the relative likelihoods of various three-dimensional configurations, given the two-dimensional projection that has been sensed. The analysis of optimal decision-making under these conditions brings in another aspect of statistical decision theory, Bayesian inference B12 (Knill & Richards 1996).

3. *Illustrative Examples*

The development of mathematical analyses of information processing and decision making inspired a psychology focused on mental representations. What has sustained it are the many examples of human and animal behavior that imply an underlying representation. Some the simplest and most illuminating examples are found in learned behavior in non-human animals that depend on underlying representations of abstract but basic properties of the world like distance, direction, duration and time of day (phase of the day-night cycle). sx0085

A honeybee forager, when she returns to the hive after discovering or revisiting a source of rich nectar, does a dance that symbolizes the solar bearing (direction relative to the sun) and distance of the source from the hive. Foragers that have followed her while she danced later leave the hive and fly in the indicated direction for the indicated distance before they begin to look for the source. Because the dance directly symbolizes direction and distance and because the witnesses to the dance base their own flight directions and distances on what they have observed, it seems inescapable that the direction and distance of the source must be represented in the system that controls bee behavior, the bee brain. In this case, the representational nature of mental processes is manifest in a behavior that is itself representational. (See Gallistel B4 1998 for a recent review of insect navigation and bee dancing, emphasizing the information processing implications.) sx0090

The dance is in the form of a figure eight. It is performed on the vertical surface of the interior of the hive out of sight of the sun. When running the middle bar of the eight (the part common to the two circles), the dancing bee waggles rapidly from side to side. The angle of this waggle run with respect to the vertical symbolizes the direction of the source relative to the sun, while the number of waggles symbolizes the distance. sx0095

The angle of the waggle run with respect to the vertical changes during the day so as to take into account the changing direction of the sun, even under sx0100 conditions where the dancers have not seen the sun nor anything else in the sky that indicates where the sun is for hours or even days. Both the dancer and her audience are able to represent compass direction (direction relative to the earth's surface) by reference to the sun's compass direction, first, because they have learned the solar ephemeris, the compass direction of the sun as a function of the time of day, and, second, because they possess a circadian clock. The circadian clock symbolizes the time of day. It is a cyclical molecular process within nerve cells (Gekakis et al., B6 B15 1998; Sehgal 1995), with approximately the same period as the day-night cycle that is synchronized to the sun's cycle every dawn and dusk by signals coming from photoreceptors. Because this biochemical cycle sx0105 within cells is synchronized with the day-night cycle, phase within this biochemical cycle—the momentary concentrations of the different molecules whose concentration varies cyclically—indicate the phase of the earth's rotational cycle, that is, the time of day.

The solar bearing symbolized by the direction of the sx0110 waggle run is computed from the representation of two different aspects of the bee's previous experience. One set of experiences are those from which she learns B2 the solar ephemeris (Dyer & Dickinson 1996). The other is the foraging experience from which she learns the compass direction of the source from the hive. The solar bearing is the angular difference between the compass direction of the source from the hive and the current direction of sun (as given by the solar ephemeris).

There does not appear to be a way to account for the sx0115 bee's behavior without endowing her brain with the capacity to symbolize the time of day, compass direction and distance. She has also to have the capacity to learn functions like the solar ephemeris. A function is a set of paired symbols, an input symbol and an output symbol. The input symbol in the solar ephemeris represents the time of day, while the output symbol represents the compass direction of the sun. A function may be realized by means of a look-up table, which stores the possible pairs of input and output symbols, but this can make large demands on memory. Alternatively, a function may be generated by a neuronal process that transforms an input signal into an output signal. In that case, the relation between the input and the output of this process must have the same mathematical form as the solar ephemeris itself.

Finally, the bee brain must be able to compute an angular difference, a symbol representing the difference between the compass direction given by its solar ephemeris function and the compass direction of the source. This latter symbol is retrieved when needed sx0120 from the memory generated at the time the bee found the source. The result of this computation, the symbol representing the angular difference between the directions represented by two other symbols, represents the solar bearing of the source. It is this angle that we observe when a dancer makes her waggle run. A psychology focused on mental representations rests on the claim that there is no way to explain this robust and reliable fact about bee behavior except by an appeal to the kind of information processing just described.

A second example of the fundamental role that sx0125 information processing plays in the control of behavior comes from the extensive studies of conditioned behavior in the common laboratory animals—the rat, the pigeon, and the rabbit. In Pavlovian conditioning, the experimenter repeatedly presents temporally paired elementary stimuli. For example, using rabbits as subjects, the experimenter may repeatedly present a tone followed at a short latency by an annoying puff of air directed at the sclera of the eye or an annoying shock to the skin around the eye. The tone is called a conditioned stimulus (CS), because it elicits observable behavior only after conditioning, while the puff or shock is called an unconditioned stimulus (US), because it elicits observable behavior in the absence of any conditioning. When a US has reliably followed a sx0130 CS, the subject responds to the CS in anticipation of the US. In the present example, the rabbit blinks when it hears the tone. This blink is called the conditioned response. It is so timed that the moment of peak closure more or less coincides with the moment when the US is expected. If the US sometimes comes at a latency of 0.4 seconds and sometimes at a latency of 0.9 seconds, the rabbit learns to blink twice, with the first blink peaking at about 0.4 seconds and the second B11 at about 0.9 seconds (Kehoe et al., 1989).

Evidently, the rabbit measures and remembers the sx0135 durations of the intervals between the onsets of the tone and the onsets of the US. How else can we explain the fact that it matches the latency of its response to the latency of the US? The rabbit must possess a B20 memory like the memory that Alan Turing (1936) sx0140 placed at the heart of his mathematical abstraction of a computing device, the so-called Turing machine. This notional machine has a memory to which it writes and from which it reads symbols. If the rabbit did not have a memory in which it could store a symbol representing the CS-US latency and from which it could subsequently retrieve that symbol, its ability to match its conditioned response to that latency would be inexplicable.

It is a general property of conditioned behavior that sx0145 the latency of the conditioned response is proportional B5 to the CS-US latency (Gallistel & Gibbon 2000). Moreover, from the nature of the variability in conditioned response latencies, it appears that the decision to about when to make a conditioned response following the onset of a CS must be based on the ratio between the remembered CS-US interval and the interval elapsed since the onset of the current CS B7 (Gibbon et al., 1984). Thus, when the tone sounds, the rabbit retrieves from memory a symbolic value representing the CS-US interval, measures the time elapsed since the onset of the tone, to generate a constantly growing signal whose momentary magnitude represents the duration of the currently elapsed interval, computes the ratio of the two values, and responds when the ratio exceeds a critical value.

sx0150 Just as bees can compute an angular difference from directions (compass angles) stored in memory, so rats can compute a temporal difference from durations stored in memory, as shown by experiments using what is called backward conditioning. In a backward conditioning experiment, the US precedes the CS. For example, a tone CS comes on 1 second after a shock US ends. Under these conditions, subjects respond weakly or not at all to the tone, because it no longer gives advanced warning of the US. Although they do not respond to the tone, they learn the (negative) interval between it and the shock. This is shown by also teaching them a forward temporal relation between a light and the tone. When they have also been taught that the onset of the light predicts the onset of the tone after a latency of 5 seconds, then they respond B1 strongly to the light (Barnet et al., 1997). From their sx0155 representation of the tone-shock interval (= −1 second) and their representation of the light-tone interval (= +5 seconds), they appear to have computed the expected light-shock interval (4 seconds). Consequently, they react fearfully to the light, even though it has never been followed by shock. Its only connection to shock is by way of the tone, but they do not react fearfully to the tone itself, because it has always followed the shock. The predictive relation of the light to the shock has been inferred by computations performed with the symbols that represent the two durations.

sx0160 As these illustrative examples show, animals are able to function effectively in a complex world because their brains construct mental representations of behaviorally important aspects of that world—spatial relations, temporal relations, numerical relations, social relations, and so on. Wherever there is regularity and form in the world, animals represent that regularity and that form in order to exploit it for their own ends. The most basic mechanism of life itself—the genetic mechanism—is a mechanism for copying, transmitting, and processing information. A significant fraction of that information specifies the immensely complex structure of the brain, an organ dedicated to the processing of information about the animal's environment.

Bibliography

b1 Barnet R C, Cole R P, Miller R R 1997 Temporal integration in second-order conditioning and sensory preconditioning. *Animal Learning and Behavior* **25**: 2 221–233

b2 Dyer F, Dickinson J A 1996 Sun-compass learning in insects: Representation in a simple mind. *Current Direction in Psychological Science*. **5**3: 67–72 b3 Edelman G, & Tononi G 2000 *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books/Allen Lane, New York

b4 Gallistel C R 1998 Brains as symbol processors: the case of insect navigation. In: Sternberg S, Scarborough D (eds.) *Conceptual and methodological foundations*. Vol. 4 of *An invitation to cognitive science*. 2nd edn., 1–51. MIT Press, Cambridge, MA

b5 Gallistel C R, Gibbon J 2000 Time, rate and conditioning. *Psychological Review* **107**: 289–344

b6 Gekakis N, Staknis D, Nguyen H B, Davis F C, Wilsbacher L D, King D P, Takahashi J S, Weitz C J 1998 Role of the CLOCK protein in the mammalian circadian mechanism. *Science*. **280**: 1564–1570

!QA: Is this an acronym or should it be l.c.?

b7 Gibbon J, Church R M, Meck W H 1984 Scalar timing in memory. In: Gibbon J, Allan L (eds.) *Timing and time perception* Vol. 423, 52–77. New York Academy of Sciences, New York

b8 Green D M, Swets J A 1966 *Signal Detection Theory and Psychophysics*. Wiley and Sons, New York

b9 Hull C L 1930 Knowledge and purpose as habit mechanisms. *Psychological Review* **37**: 511–525

b10 Hull C L 1952 *A Behavior System*. Yale University Press, New Haven, CT

b11 Kehoe E J, Graham-Clarke P, Schreurs B G 1989 Temporal patterns of the rabbit's nictitating membrane response to compound and component stimuli under mixed CS-US intervals. *Behavioral Neuroscience* **103**: 283–295

b12 Knill D, Richards W (eds.) 1996 *Perception as Bayesian Inference*. Cambridge University Press, New York

b13 Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W 1997 *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA

b14 Rumelhart D E, McClelland J L (eds.) 1986 *Parallel Distributed Processing*. MIT Press, Cambridge, MA

b15 Sehgal A 1995 Molecular genetic analysis of circadian rhythms in vertebrates and invertebrates. *Current Opinion in Neurobiology* **5**6: 824–831

b16 Shannon C E 1948 A mathematical theory of communication. *Bell Systems Technical Journal* **27**: 379–423 623–656

b17 Skinner B F 1938 *The Behavior of Organisms*. Appleton-Century-Crofts, New York

b18 Skinner B F 1950 Are theories of learning necessary? *Psychological Review* **57**: 193–216

b19 Skinner B F 1990 Can psychology be a science of mind? *American Psychologist* **45**: 1206–1210

b20 Turing A M 1936 On computable numbers, with an application to the *Entscheidungs*problem. *Proceedings of the London Mathematical Society* 2nd series. **42**: 230–265

C. R. Gallistel