CHAPTER THREE

PERCEPTION AND COMPUTATION

In this chapter we indicate how the class of observers properly contains the class of Turing machines. We discuss the simulation of observers by Turing machines.

1. Turing observers

We begin with a brief review of Turing machine terminology. The theory of automata considers several characterizations of Turing machines. All characterizations are equivalent to defining a Turing machine as a language recognizer. Let Σ be the "terminal alphabet" of a Turing machine T; Σ is the set of all elementary symbols which can be input to T. Let Σ^* be the set of all strings of finite length of elements of Σ. The ***language recognized by*** T is the subset $L \subset \Sigma^*$ consisting in those strings which, when input, cause T to halt in an "accept state." The property of a subset L of Σ^* which allows it to be recognized by some Turing machine is called "recursive enumerability," and sets enjoying this property are called "recursively enumerable," abbreviated ***RE.***[1] More generally, given any countable set C we can define a subset $B \subset C$ to be RE in C if C can be embedded in some Σ^* in such a way that B corresponds to an RE language in Σ^* . In this sense we can speak of a Turing machine "recognizing a subset B of C." Intuitively, B is RE in C if there exists a procedure with this property: given an arbitrary element x of C, if $x \in B$ the procedure will determine this in finitely many steps. If $x \notin B$, however, the procedure may not halt. In fact, if B is RE in C, its complement $C - B$

[1] There are various ways to give a mathematical characterization of the collection all RE subsets of a given Σ^*, but we will not need to do so here.

may not be RE. If both B and $C - B$ are RE in C, we say simply that B is a *recursive* subset of C. This means intuitively that there is a recursive procedure which will determine in finitely many steps whether or not any given element of C is in B. A function f with a countable domain D and range R is called *recursive* or *Turing computable* if the graph Γ of f is RE in $D \times R$. The Turing machine which computes f is then the one which recognizes Γ. In fact, to compute $f(d)$ for $d \in D$ the machine can enumerate the elements of Γ until it reaches the unique one whose first component is d; the second component is then $f(d)$. Thus, Turing machines can also be characterized as computers of recursive functions. It can be shown that both the support and range of a recursive function are RE sets.

All Turing machines have sufficient structure to be viewed as observers. We describe below how the class of Turing machines is a subclass of the class of observers. Simply stated, this subclass consists of observers whose inferences are deductively valid, and the deduction in question is a Turing computation. This accounts for but a small subclass of observers; observers more generally perform inferences that are not deductively valid (while they have some degree of inductive strength). Moreover, even if the inferences of an observer are deductively valid, they need not be Turing computable.

Let T be a Turing machine, with terminal alphabet Σ, that recognizes the language $L \subset \Sigma^*$. We associate to T an observer (X, Y, E, S, π, η) as follows: we will view Σ^* as a measurable space whose σ-algebra is its full power set. Let $X = Y = \Sigma^*$, $E = S = L$, $\pi =$ the identity map on Σ^*. Then for $s \in S$, $\pi^{-1}\{s\}$ is just a copy of the point s, now considered as an element of E. $\eta(s, \cdot)$ must therefore be Dirac measure ϵ_s concentrated on this point. We will denote by ϵ the kernel defined by $\epsilon(s, \cdot) = \epsilon_s$. With this notation we can state how the class of Turing machines is a subclass of the class of observers.

1.1. The assignment

$$T \mapsto (\Sigma^*, \Sigma^*, L, L, \text{identity}, \epsilon)$$

embeds the class of Turing machines in the class of observers.

The observers which arise from Turing machines in this manner are called *Turing observers.* An observer (X, Y, E, S, π, η) is isomorphic to a Turing observer if and only if X is countable, E is an RE subset of X, and π is bijective.

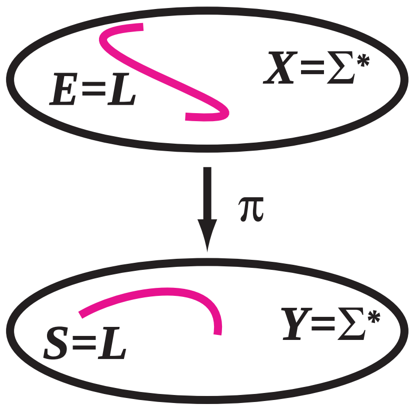

FIGURE 1.2. *A Turing observer.* $X = Y = \Sigma^*$. $E = S = L$. π *being bijective means that the Turing observer's conclusions are deductively valid.*

2. Turing simulation

Once we recognize Turing machines as a subclass of observers, we see that most observers are not Turing machines: perception is a more general concept than computation. However one can ask whether, for a given observer, there exist Turing machines which simulate that observer and, if so, how these machines are related. For an observer with uncountable X, Y, E, or S we ask whether there exist Turing machines which simulate discrete approximations of the observer. To study these questions we here define a canonical procedure for the simulation of discrete observers. In the next section we consider the issue of discretization.

Let $O = (X, Y, E, S, \pi, \eta)$ be the observer to be simulated. The objective of the simulation is the computation of $\eta(s, A)$, for all sensorial points s and $A \in \mathcal{X}$. However, such a computation is meaningful as stated only when S and $\mathcal{X}$ are countable sets. Let us assume that X is countable and that $\mathcal{X}$ is just 2^X. S is then countable (since $S \subset Y$ with $\pi: X \to Y$ surjective), but of course $\mathcal{X}$ is uncountable in general. The natural way to handle this difficulty is to restrict our attention to the recursively enumerable subsets A of X. In fact, let $\mathcal{A}$ denote the collection of these subsets. $\mathcal{A}$ itself is countable, as is well-known, so that with our restriction we can view the objective of the simulation as the computation of the function $\eta(\cdot,\bullet)$ whose domain is now the countable set $S \times \mathcal{A}$. Moreover the question of the computability of η then takes a much simpler form, as follows. Let A be any subset of X. The infinite sum

$$\sum_{x \in A} \eta(s, \{x\})$$

converges abstractly to $\eta(s, A)$, but without a procedure for enumerating the elements of A the sum has little computational meaning. If A is recursively enumerable, however, there is an effective procedure for enumerating these elements; hence there is an effective procedure for approximating the value $\eta(s, A)$ provided that for each $x \in A$, $\eta(s, \{x\})$ is computable. In this way, the restriction of our attention to sets $A \in \mathcal{A}$ leads us to consider the question of the computability of the $\eta(s, \{x\})$ for all $x \in X$. We now define canonical simulation.

Definition 2.1. Let O be an observer whose X is countable[2] and whose $\mathcal{X} = 2^X$. We will associate to O the function $f\colon S \times X \to \mathbf{R}$ defined by

$$f(s, x) = \eta(s, \{x\})$$

The *canonical Turing simulator* of O is the machine T which recognizes S in Y and then computes f.

It is clear that T exists if and only if O satisfies the requirements:
(i) S is recursively enumerable in Y.
(ii) f is recursive.

Generally, X and S are uncountable, so by definition there is no canonical simulator for these observers. But even when everything is countable, the conditions (i) and (ii) above will not be satisfied in general, so simply by making a discrete approximation to an observer we cannot expect that it will have a Turing simulation. However, at least in certain instances of interest to vision researchers, discrete approximations may allow Turing simulation. For these reasons and others it is essential to have a general theory of discretization of observers. We now give some indication of this.

---

[2] This can be generalized to include observers whose X is not necessarily countable, but whose measures $\eta(s,\cdot)$ on X are "atomic." We will not develop this generalization here; but it is discussed in Bennett, Hoffman and Prakash (1987).

3. Discretization

Our purpose in this chapter is to illustrate some ideas, and not to present a complete theory. Accordingly we will restrict attention to Euclidean configuration spaces X and premise spaces Y (with their standard Borel algebras), and assume that $\pi\colon X \to Y$ is a projection. Let $O = (X, Y, E, S, \pi, \eta)$ be an observer with $X = \mathbf{R}^{n+m}$, $Y = \mathbf{R}^n$, and π projection, say onto the first n coordinates. In order to effect a discretization, we assume an additional datum—a finite measure λ on S. Intuitively, λ and η come from the same source, namely a probability measure ρ on E which expresses the actual probabilities of distinguished configurations in a specific universe.[3] In this case the natural choice for λ is $\pi_*(\rho)$, just as the natural choice for η is a version of the regular conditional probability distribution (rcpd) of ρ with respect to π. In our case, since λ and η are assumed given, we can simply define the measure ρ on E by $\rho = \lambda\eta$:

$$\rho(A) = \int_S \lambda(ds)\eta(s, A), \quad A \in \mathcal{E}.$$

We will describe a canonical procedure for discretization in terms of this ρ. This procedure will result in observers with countable configuration spaces.

Let δ be a simultaneous partition of X and Y by measurable subsets of nonzero Euclidean volume. Let X_δ and Y_δ denote the sets whose elements represent the distinct subsets of the respective partitions. *We will assume that for $\bar{y} \in Y_\delta$, $\pi^{-1}(\bar{y})$ is a union of elements of X_δ.* (For example, we can partition X and Y into hypercubes whose edges have length d, and whose vertices have coordinates which are integer multiples of d. The resulting sets of hypercubes are then the X_δ and Y_δ. See Figure 3.1.) Given this assumption, π induces a map $\pi_\delta\colon X_\delta \to Y_\delta$. Let E_δ denote the collection of those sets $\bar{x}$ in X_δ such that $\rho(E \cap \bar{x}) > 0$. Let $S_\delta = \pi_\delta(E_\delta)$. As a consequence of these definitions, if $\bar{e} \in E_\delta$ then $\rho(\bar{e}) > 0$, and if $\bar{s} \in S_\delta$, $\lambda(\bar{s}) > 0$. We will define below a kernel η_δ (depending on the original kernel η and on δ) such that $O_\delta = (X_\delta, Y_\delta, E_\delta, S_\delta, \pi_\delta, \eta_\delta)$ is an observer. We can think of this O_δ as a "δ-discretization" of O.

So that we can outline our intentions, let us assume for the moment that η_δ has already been defined. Our intention is to compare the various discretizations (for different δ) with each other and with the original observer. To this end, we give a canonical embedding of the discrete spaces E_δ and S_δ in the original X and Y. More precisely, we associate to each $\bar{e} \in E_\delta$ a point in X and

---

[3] Such a λ arises naturally in the discussion of noisy perceptual inferences (cf. **2**–4).

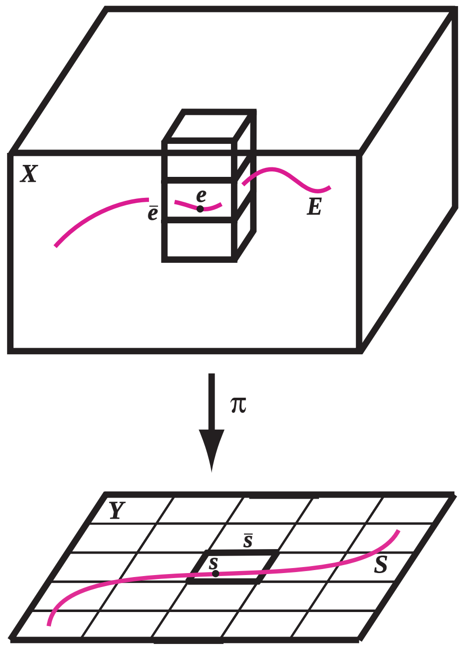

FIGURE 3.1. *A discretization of an observer.*

to each $\bar{s} \in S_\delta$ a point in Y by means of mappings $\alpha\colon E_\delta \to X$ and $\beta\colon S_\delta \to Y$, such that the diagram

$$\begin{array}{ccccc} E_\delta & \xrightarrow{\alpha} & E'_\delta & \subset X \\ \Big\downarrow{\scriptstyle \pi_\delta} & & \Big\downarrow{\scriptstyle \pi} & \\ S_\delta & \xrightarrow{\beta} & S'_\delta & \subset Y \end{array}$$

commutes. Here we have put $E'_\delta = \alpha(E_\delta)$, $S'_\delta = \beta(S_\delta)$; these are countable and hence measurable. When this is done, any kernel on E_δ relative to S_δ via π_δ can be transported, using α and β, to a kernel on E'_δ relative to S'_δ via π. In particular, η_δ may be transported in this manner to η'_δ. In this sense we can then consider an observer $O'_\delta = (X,\ Y,\ E'_\delta,\ S'_\delta,\ \pi,\ \eta'_\delta)$. We think of O'_δ as a geometric embedding of O_δ into the original spaces X and Y, and as a discrete *approximation* of the original observer $O = (X, Y, E, S, \pi, \eta)$. E'_δ and S'_δ do not actually lie on E and S in general, but converge to E and S as the partition δ gets arbitrarily fine.

To achieve this let us first consider how to embed S_δ in Y. Given the subset of Y represented by the element $\bar{s} \in S_\delta$, we may find its center of mass with respect to the measure λ (restricted to $\bar{s}$). This center of mass will not, in general, lie in S, but it is the natural punctual representative of $\bar{s}$ in Y. Recalling that for $\bar{s} \in S_\delta$ $\lambda(\bar{s}) > 0$, we may now define the embedding $\beta\colon S_\delta \to Y$ by

$$\beta(\bar{s}) = \int_{\bar{s}} s \; \lambda_{\bar{s}}(ds) \tag{3.2}$$

with

$$\lambda_{\bar{s}}(ds) = \frac{1}{\lambda(\bar{s})} 1_{\bar{s}}(s)\lambda(ds), \quad \bar{s} \in S_\delta.$$

That is, $\lambda_{\bar{s}}$ is the normalized restriction of λ to the hypercube $\bar{s}$.

Similarly, we wish to define a center-of-mass embedding for E_δ using appropriate measures on X. For purposes of finding the center of mass of $\bar{e}$ in E_δ, it may seem natural to use the normalization of the restriction of ρ to $\bar{e}$. However, as we shall see below, a slightly different choice of measure on $\bar{e}$ is much better suited to the task at hand. To this end, let $\rho_{\bar{e}}$ be the normalized restriction of ρ to $\bar{e}$, that is,

$$\rho_{\bar{e}}(C) = \frac{\rho(C \cap \bar{e})}{\rho(\bar{e})}.$$

By construction of E_δ, this yields a probability measure on $\bar{e}$. It is straightforward to verify that since η is the rcpd of ρ with respect to π, the measure $\rho_{\bar{e}}$ also possesses an rcpd with respect to π, a version of which is given by the formula

$$\eta_{\bar{e}}(s, de) = \frac{\eta(s, de)}{\eta(s, \bar{e})} 1_{\bar{e}}(e)$$

(which is defined up to a set of $\pi_* \rho_{\bar{e}}$-measure zero in its first argument, and which we may take to be a markovian kernel off this zero-measure set). As usual, by composing $\eta_{\bar{e}}$ with the measure $\pi_* \rho_{\bar{e}}$ we can reconstruct $\rho_{\bar{e}}$. We shall, however, compose $\eta_{\bar{e}}$ with $\lambda_{\pi(\bar{e})}$ instead, defining a new measure $\nu_{\bar{e}}$ as

$$\nu_{\bar{e}}(C) = \int_{\pi(\bar{e})} \lambda_{\pi(\bar{e})}(ds) \;\cdot\; \eta_{\bar{e}}(s, C), \quad \bar{e} \in E_\delta,$$

where C is any measurable subset of $\bar{e}$. This is, by construction, a probability measure supported on $\bar{e}$, which gives the embedding of E_δ in X by the map α as follows:

$$\alpha(\bar{e}) = \int_{\bar{e}} e \; \nu_{\bar{e}}(de), \quad \bar{e} \in E_\delta. \tag{3.3}$$

As indicated above we will denote the image $\alpha(E_\delta)$ in X by E'_δ and the image $\beta(S_\delta)$ in Y by S'_δ.

We now show that these embeddings α and β respect the original map π in the sense that for any $\bar{e} \in E_\delta$, $\pi(\alpha(\bar{e})) = \beta(\pi_\delta(\bar{e}))$. This is satisfying, as it displays a consistency of the perspective maps at all scales, and expresses a connection between the discretizations at the various scales.

To see why $\pi(\alpha(\bar{e}))$ should equal $\beta(\pi_\delta(\bar{e}))$, note that since π is linear, we may take π inside the integral defining $\alpha(\bar{e})$, so that

$$\begin{aligned}\pi(\alpha(\bar{e})) &= \int_{\bar{e}} \pi(e)\nu_{\bar{e}}(de) \\ &= \int_{\pi(\bar{e})} \lambda_{\pi(\bar{e})}(ds) \int_{\bar{e}} \eta_{\bar{e}}(s, de)\pi(e).\end{aligned}$$

But $\eta_{\bar{e}}(s, \cdot)$ is supported on the fibre where $\pi(e) = s$, so that

$$\begin{aligned}\pi(\alpha(\bar{e})) &= \int_{\pi(\bar{e})} \lambda_{\pi(\bar{e})}(ds) \ \cdot \ s \int_{\bar{e}} \eta_{\bar{e}}(s, de) \\ &= \beta(\pi_\delta(\bar{e})).\end{aligned}$$

If we had used the measure $\rho_{\bar{e}}$ in defining the embedding α of E_δ, we would not have obtained this result.

Finally, we come to the definition of η_δ, which appears in $O_\delta = (X_\delta, Y_\delta, E_\delta, S_\delta, \pi_\delta, \eta_\delta)$, and $O'_\delta = (X, Y, E'_\delta, S'_\delta, \pi, \eta_\delta)$. η_δ is the discretization of η

$$\eta_\delta(\bar{s}, \{\bar{e}\}) = \int_{\bar{s}} \lambda_{\bar{s}}(dt)\eta(t, \bar{e}), \quad \bar{s} \in S_\delta, \bar{e} \in E_\delta. \tag{3.4}$$

This is by construction a markovian kernel on $S_\delta \times \mathcal{E}_\delta$. Here we are merely averaging the contributions from the various original fibres of π in the given partition subset $\bar{e}$. We can view η_δ as a kernel on $S'_\delta \times \mathcal{E}'_\delta$, simply by using the identifications α and β.

In general, E_δ and S_δ need not be recursively enumerable, and a fortiori the function f, defined as in 2.1 above using η_δ, need not be recursive. Thus a discretization O_δ of a non-Turing observer O may not have a Turing simulation.

4. Effective simulation: The algebraic case

There is at least one natural class of observers for which suitable discretizations sometimes have canonical Turing simulations. These are the “algebraic observers," such as the biological motion observer of chapter one or the structure-from-motion observer of chapter two. In the case of the structure-from-motion observer (section four of chapter two), E is the locus of points in $\mathbf{R}^{18}$ satisfying Equations **2**–4.2 through **2**–4.9, and S is the image of E in $\mathbf{R}^{12}$ by the projection π. The polynomial equations defining E have integer coefficients. Thus we can apply the following general result:

4.1. Suppose $Y = \mathbf{R}^n$, $X = \mathbf{R}^{n+m}$, and $\pi: X \to Y$ is projection onto a set of n of the coordinates of X. Suppose E is the locus of zeroes in X of a finite set of polynomial equations (in the $n+m$ variables of X) with integer coefficients. Let $S = \pi(E)$, and let X_δ, Y_δ, E_δ, S_δ, π_δ be the discretizations resulting from the partition δ of X and Y as described in the previous section, where δ can be any partition into subsets whose boundaries are defined by any integer coefficient algebraic equations.[4] Then

(i) S_δ is a recursive subset of Y_δ;

(ii) E_δ is a recursive subset of X_δ. For all $\bar{y} \in Y_\delta$, $\pi_\delta^{-1}(\bar{y}) \cap E_\delta$ is a recursive subset of $\pi_\delta^{-1}(\bar{y})$ (and therefore of X_δ).

This result obtains by applying the Theorem of Tarski on the decidability of polynomial inequalities.[5] We omit the details here.

Condition (i) of 4.1 corresponds to the first requirement (given in section two above) for the Turing simulator of O_δ to exist. Condition (ii) is a necessary condition for the function f associated to the observer O_δ to be recursive, but it is certainly not sufficient for this purpose; this depends ultimately on the nature of η_δ.

Finally, we suggest that the real issue vis-à-vis the relationship between perception and computation is not so much the existence of a Turing simulation for a given discretization of the observer, but is rather the structure of the collection of all the Turing simulations (assuming they exist) for the discretizations of the observer at a collection of scales. Here we give only a brief sketch of these ideas.

Recall that with the introduction of the observer O'_δ we have a natural way to compare the discretizations of the original observer O for various partitions δ. Let us consider a set Δ of partitions, which we can view as partially ordered by "fineness": δ_1 is finer than δ_2 if every element of X_{δ_1} is a subset of some element of X_{δ_2}. Let us further assume that if δ_1 is finer than δ_2, and if moreover

[4] This includes the cases where the partitioning subsets are hyperrectangles or hypercubes. Recall that the cylinder $\pi_\delta^{-1}(\bar{y})$ is a union of elements of X_δ.

[5] See, e.g., Jacobson 1974.

O_{δ_1} and O_{δ_2} have canonical Turing simulations T_1 and T_2, then there is a natural way to compare T_2 with T_1 as Turing machines. Finally, assume that we have fixed an appropriate notion of equivalence of Turing machines. Granting all this, the following sample definition gives the flavor of what we have in mind:

Definition 4.2. O has a Δ-*effective simulation* if
1. Each O_δ for $\delta \in \Delta$ has a canonical Turing simulation.
2. The comparisons between the machines corresponding to sufficiently fine δ's is an equivalence.
3. As δ gets fine, the limit of the η_δ is η.[6]

O has a Δ-effective simulation if the family of discretized observers obtained from O using the partitions in Δ has a certain stability. Intuitively, what is significant is not the particular family of partitions Δ, but rather that there exists even one Δ for which the definition is satisfied, provided that this Δ contains arbitrarily fine partitions. The definition then asserts that the original O, although it may be given as a non-discrete object, has a Turing machine representation which is stably scale-independent. This motivates the following sequel to Definition 4.2:

Definition 4.3. With the notation and assumptions as above, suppose that there exists a family Δ which contains arbitrarily fine partitions for which O has a Δ-effective simulation. Then we will say simply that O has an *effective simulation.*

Here is a sample conjecture to accompany our sample definition:

4.4. Suppose $O = (X, Y, E, S, \pi, \eta)$ satisfies the hypotheses of 4.1. Suppose that for some integer k, exactly k points of E lie over each point s of S via π, and that $\eta(s, \cdot)$ assigns probability $1/k$ to each of these k points. Then O has an effective simulation.

We cannot give a detailed analysis of 4.4 here since the notion of Turing equivalence used in Definition 4.2 has not been precisely specified. We mention, however, that the key idea is to find a family Δ of partitions so that, for all sufficiently fine $\delta \in \Delta$, the following property holds: For each $s \in S$ and $\bar{e} \in E_\delta$, $\bar{e} \cap E$ contains at most one point from the original fibre $\pi^{-1}(s) \cap E$.

[6] Here we mean that the transports of the η_δ to E'_δ converge to η as the E'_δ converge to E.

For this purpose the δ's cannot, in general, be hypercube partitions; in fact the simplest δ's that work are hyperrectangle partitions, where the proportions of the rectangles depend on the "slope" of E relative to S. In particular these proportions will need to vary within the same partition δ, depending on the location of the hyperrectangle in X. In any case, once the δ's have this property, all the maps π_δ are k-to-one, and one can check (using Definition 3.4 of η_δ) that given $\bar{s} \in S_\delta$, $\eta_\delta(\bar{s}, \cdot)$ is simply the constant function $1/k$ on $\pi^{-1}(\bar{s}) \cap E_\delta$ (and is identically 0 on $X_\delta - (\pi^{-1}(\bar{s}) \cap E_\delta))$. Since the set $\pi^{-1}(\bar{s}) \cap E_\delta$ is RE by 4.1, it then follows that $\eta_\delta(\bar{s}, \cdot)$ is Turing computable.

We summarize the main ideas: For a Δ-effective simulation, as $\delta \in \Delta$ gets finer, both the combinatorial geometry of the maps π_δ and some essential computational character of the η_δ must stabilize. Moreover, the O_δ must converge to O. What we have, then, is a system of successively finer discretizations O_δ, converging to O, whose stable structural properties (i.e., properties which hold for all sufficiently small δ) reflect the perceptually relevant properties of the original O. Thus, the fundamental structure of O is accessible at finite stages of discretization, in a manner which is independent of scale, at least for sufficiently small scales. It seems clear that, in the absence of this kind of stability, the existence of Turing simulations for the individual O_δ's alone is an insufficient hypothesis to justify a "perception as computation" viewpoint. We propose, rather, that the analysis of effective simulation is an appropriate context in which to investigate the relationship between perception and computation.