MANUFACTURING AND HUMAN LABOR AS INFORMATION PROCESSES

Robert U. Ayres
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

and

International Institute for
Applied Systems Analysis
Laxenburg, Austria

RR-87-19
November 1987

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0089-5

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Cover design by Martin Schobel

Printed by Novographic, Vienna, Austria

Contents

Summary

1. Manufacturing as an Information Process

Several relatively unfamiliar concepts are presented together in this section. The first concept is that all manufacturing (indeed, production) can be thought of as the concentration of *information* in matter. The second key concept (elaborated in a previous paper) is that the economic system shares with living systems the characteristic that it is a "dissipative structure", capable of self-organization and growth. Like a living system, it uses free energy captured from the environment to drive its metabolic processes, but it *controls* these processes by means of information stored in structures (i.e., morphology), which in turn permit functional specialization and differentiation.

There are two different types of information in question. The first type is proportional to the free energy or "available useful work" of a system. Hence, the information content of materials *per se* is a function of their chemical composition and physical state. The second type is also embodied in materials via their shape and surface finish. Methods of computation of both thermodynamic and shape information are outlined briefly.

Evidence is presented to show that the economic value (price) of morphological information is vastly greater than the value of metabolic information. This fact, in turn, suggests some elements of a generalized minimum information principle for optimizing both design and manufacturing. The proposed principle states that the optimal strategy is to minimize the amount of costly morphological information needed to achieve a given functional purpose, while the optimal manufacturing process for the given design is one that minimizes information loss (i.e., information *not* embodied).

2. Human Labor as an Information Process

This section has three major objectives. The first is to summarize relevant ergonomic literature on the human worker as an information processor. The second goal is to present a rationale for regarding motion as equivalent to processed information and (consistent with the time-motion literature since F.W. Taylor) reexpress Taylor's time-minimization approach to task optimization in terms of minimizing the useful information output required from a worker to achieve a given task (i.e., to embody information in a product). The third purpose of the section deals with the error-defect problem and the implications of the modern ergonomic view of errors as consequences of mental overload resulting in a sharp (nonlinear) increase in the proportion of "garbage" output to "useful" output.

A simple labor rate optimization model is developed, which shows that Taylor's optimization principles are valid only in the unrealistic case where errors (garbage) impose no costs. In the real world where errors and defects impose heavy costs in terms of detection, elimination, or correction, the Taylor principle must be modified.

Foreword

Two related papers are presented here as a package. Both papers can be regarded as part of an attempt to develop new and more productive approaches to quantitative measurement in the social and economic sciences. This is an oft-neglected but important aspect of IIASA's on-going Technology–Economy–Society (TES) Program.

The first paper deals with applications of information theory to economics, especially the analysis of manufacturing. Its explicit objective is to exhibit a practical methodology for computing information "stocks" and "flows", with particular application to information embodied in form and structure. It also discusses possible optimization principles both for production processes and product design, making use of the information-theoretic approach.

The second paper deals with a parallel topic, the application of information theory in the analysis of human labor. Here the objective is to reexamine F.W. Taylor's approach to task optimization from the modern ergonomic perspective, again using the language of information theory. The error–defect problem in manufacturing is addressed, and an interesting labor rate optimization model is developed, showing Taylor's approach to be an unrealistic special case.

THOMAS H. LEE
Program Leader
Technology–Economy–Society

Preface and acknowledgments

The two papers bound together here as a package originally grew out of a course of lectures given by the author, while at Carnegie Mellon University in spring 1985 and fall 1986, on the application of information theory in manufacturing. The material has since been considerably expanded, and some of the central ideas articulated here are finding direct application in IIASA's Computer Integrated Manufacturing project. For this reason, it seems appropriate to present the background material to a wider audience as an IIASA Research Report.

I am particularly grateful to Kathryn Jackson and Paul Zahray for having had the patience to hear my ideas in their early stages, and to Richard Tredgold and an anonymous reviewer for forcing me to face some fundamental difficulties. I am also grateful to Charles Berg and Harvey Brooks for encouragement and helpful suggestions.

ROBERT U. AYRES
Deputy Program Leader
Technology–Economy–Society
Project Leader
Computer Integrated Manufacturing

1. Manufacturing as an Information Process

1.1. Introduction

It is argued here that man-made things (artifacts) can be regarded as materials that have been imprinted with useful information. The term "useful" here explicitly implies economic value. It must be emphasized at the outset that information need not have economic value and that economic value is not simply proportional to information content. These points will be reiterated later in context. Here, too, the term "information" is used in the strict technical (Shannonian) sense, as a rough measure of the inherent likelihood or distinguishability of an object (or message) with respect to a reference environment. The relevant aspects of information theory have been reviewed in some detail in a previous paper (Ayres, 1987).

Valued added by manufacturing depends on functional capability added to crude materials. Functional capability arises, in turn, from physical properties associated with specified physical and chemical composition, shape, and finish (an aspect of shape). Each of these must be held within given limits or tolerances. The required precision of compositional and dimensional specifications defines the minimum amount of information embodied in each component part. Further information is added when parts are combined and assembled into subassemblies, machines, process equipment, structures, and systems.

It is suggested, hereafter, that the "factor services" of capital and labor comprise the *information input* to the production process while the output or product is analogous to the *information output*, or "message". Evidently, the information ultimately embodied in the output (message) by labor and capital is far less than the information embodied in the inputs, including labor. Most of the input information is, in fact, wasted. An efficient production process – or an efficient economic system – must surely be one that wastes as little information as possible. A good design, on the other hand, should require as little information as possible to achieve a given functional purpose, other factors remaining equal. The idea that is elaborated below is that the function of the factor services is to *concentrate* information (or reduce entropy) in certain specified materials or products, even though the total amount of information in the environment, as a whole, is actually reduced (i.e., global entropy increases).

Georgescu-Roegen's characterization (quoted in Allred, 1977) of the economic process as a "transformation of states of low entropy to states of high entropy" is, so far as it goes, merely an implication of the second law of thermodynamics. As such, it is neither more nor less true of the economic system than of other natural processes, although Georgescu-Roegen does well to remind economists that the economic system is inherently dissipative.

What is perhaps more relevant is that economic systems appear to share a fundamental characteristic of living systems. Both are exemplars of self-organization and structure maintained by a continuous flow of free energy, or "available useful work" (Ayres 1987). The first term is more familiar, but the second is more precise. The terms "essergy" and "exergy" have also been suggested, although neither is widely accepted. Despite the imprecise language, free or available energy can be thought of as energy not yet converted to entropy, just as entropy can be regarded as unavailable energy. Simple self-organizing systems of several kinds have been studied in detail by Prigogine and his colleagues (Prigogine *et al.*, 1972), and more recently by others. So-called "dissipative structures" are characterized by local entropy minima (Prigogine, 1955), far from the static equilibrium state of maximum entropy (and maximum disorder).

In the biological case, self-organization is also characterized by specialization of *structure* (morphology) and *function* within cells, among specialized cells in larger organisms, and even among members of a community. This morphological structure allows more efficient accumulation of genetic information among lower organisms, and learned brain information among higher ones. Only humans, among all species on the evolutionary ladder, have learned to use external structures to enhance the storage and processing of information.

In the economic case, there is also an orderly specialization of functions (i.e., division of labor and diversity of products and services) together with rapid accumulation of useful information that can be stored in the form of labor skills, capital, and "pure" technology. In both cases, the enhanced rate of information accumulation is evidently associated with an enhanced rate of free energy dissipation or entropy creation.

The foregoing remarks admittedly verge on the philosophical. They are certainly too broad to elaborate – still less substantiate – in detail in the scope of a single paper. The focus hereafter is much narrower, on the physical production processes within economic systems. The major points to be emphasized in this paper are as follows:

(1) Information is closely related to free energy and, hence, to entropy. Information may be "embodied" in an open system by reducing its entropy and increasing its free energy.

(2) Information can be divided into at least three types with respect to its economic value (or usefulness):

Type I – Information embodied as free energy in fossil fuels or solar radiation.

Type II – Thermodynamic information associated with metabolic (materials processing) functions in the economic system.

Type III – Morphological (form, shape, finish) information associated with manufacturing, assembly, and construction functions in the economic system.

(3) Value added by economic activities is associated with the conversion of Type I information into Type II and Type III information, respectively.

(4) *Metabolic* information is currently cheaper by many orders of magnitude than *morphological* (shape) information. The latter is, in some sense, more useful. This statement can be reformulated more precisely in terms of the marginal price (value) of a "bit" of metabolic information (P_{metab}) vis-à-vis the marginal price (value) of a "bit" of morphological information, (P_{morph}). In general: $P_{\text{morph}} >> P_{\text{metab}}$.

(5) From this inequality an important design and manufacturing rule follows: the optimum strategy must be to reduce the amount of morphological (forming and shaping) information required to achieve a functional performance level. This may be accomplished by modifying the composition to simplify the shaping–forming–assembly requirements. On the other hand, any attempt to improve performance strictly by modifying or manipulating shapes (i.e., by embodying more information in shapes) soon becomes excessively costly to implement.

(6) It also follows from the basic inequality that fundamental research on improving the currently low efficiency of shaping–forming–assembly might be well worth the investment. The principle of "minimum information loss" is suggested as a guideline for future research.

The next two sections of this paper deal with surface information and shape information, respectively. The fourth section derives the relative magnitudes of P_{morph} and P_{metab}. The fifth section discusses some economic implications; and the sixth, the implications for design and manufacturing.

A key point to be emphasized here is that information is gained or lost by any change in the thermodynamic state of a material subsystem with respect to its environment. In fact, it is shown in *Appendix 1.A* that the information embodied in such a subsystem is proportional to the free energy or available useful work in that subsystem. Of course, free energy can only be concentrated in one material subsystem – for instance, a metal – at the expense of others, such as fuels. There is always a reduction in the total amount of free energy in the combined subsystems, considered as a whole. It is this loss that is the thermodynamic "cost" of materials extraction, separation and refining processes. The information on negentropy added to or concentrated in the refined materials is therefore also roughly proportional to the overall expenditure of free energy (mainly from fossil fuels) to drive the metabolic functions of the economic system. This "used up" free energy is converted to entropy. The economic implications of this are discussed later in Sections 4 and 5.

1.2. Surface information; general considerations

The natural starting point for a discussion of morphological information is to consider the thermodynamic information associated with a shape. To be as precise as possible, let us first assume the existence of a regular three-dimensional lattice of geometrical points fixed in Cartesian space. If every one of the points were occupied by molecules, the result would be a perfect crystal.

The next logical step is to define a plane surface through the lattice. Since all crystals have a number of natural planar surfaces, it is reasonable to begin with a plane that exactly coincides with one layer of the lattice. That is to say, a subset of lattice points are in the plane. The plane defines a "surface" if all lattice sites on one side of the plane are empty and at least some of the sites in the plane itself are occupied. Because the surface is associated with a solid, it can also be assumed that most of the sites on the nonempty side of the plane are occupied.

An information value associated with the occupancy of this surface can now be determined. Assume the plane surface has a finite area enclosing M lattice points. There are just two possible states: "occupied" and "empty". Assuming independence, the total number of possible occupancy-complexions W_{occ} for the surface is exactly 2^M. If all complexions are equally likely, the information gained by any observation that determines exactly which lattice sites in the plane are occupied is

$$H_{\text{occ}} = \mathrm{k} \ln W = \mathrm{k}\, M \ln 2 \tag{1.1}$$

or kln2 (1 bit) per lattice site. It is tempting to note that this is inherently a small number compared to any entropic change of thermodynamic origin associated with the entire mass of the solid because the number of particles (or sites) on any surface M is very small compared to the number N in the volume. In fact, since the surface area (in any unit) is roughly the 2/3 power of the volume, it follows that

$$M \simeq N^{2/3}$$

If the solid has a mass of the order of mole, then N is of the order of Avogadro's number ($\mathrm{A} = 6 \times 10^{23}$), whence M is somewhat less than 10^{16}. In other words, for a one-mole solid, the surface entropy appears to be smaller than the volume entropy by a factor of at least 10^8. This discrepancy in magnitudes makes it appear, at first glance, that the information associated with a surface can be neglected in comparison with the information associated with a volume, which is always proportional to the number of molecules in the volume [see *Appendix 1.A*, equation (1.A.5)].

However, the foregoing argument is incomplete. On deeper reflection, it is clear that surface information is not completely defined by occupancy information. It is true that all three-dimensional shapes must have surfaces. It is also true that a given (plane) surface with M possible lattice sites must have an information "content" of Mkln2, as argued above. In fact, the assumption of a plane surface can be relaxed, since it was not required to derive the result. It is important to emphasize that the "surface" information derived above relates only to lattice site occupancy and not to geometry. From the lattice occupancy perspective, all geometries (or shapes) are the same. It follows that the information gained by a choice of one particular surface geometry from among all possible *nonoverlapping* surface geometries with a common perimeter has not yet been taken into account. (Surfaces that overlap, in the sense of differing only at a small subset of points, must be excluded, since they have different possible occupancy states *only* to the extent that they include different lattice sites.)

The next question is: how many different nonoverlapping surfaces with the same perimeter must be considered? The number of distinguishable surfaces with a common perimeter is extremely large if the nonoverlapping condition is relaxed. In fact, it can be shown that the number of common-perimeter surfaces containing M lattice sites from only two adjacent layers (with M sites each) is slightly less than 2^M due to edge effects. However, most of these surfaces have many lattice sites in common with others, and the total number of possible different occupancy complexions encompassed by all the surfaces limited to two layers is roughly 2^{2M}. Notice that the total number of possible sites on the two layers appears in the exponent.

By a logical extension of the above reasoning, virtually all lattice points in the volume can belong to at least one (distorted) geometrical surface passing through the specified perimeter, and the sum total of different points in all such surfaces - subtracting overlaps - is just the number of lattice points in the volume itself. Following this argument to its limit, it is logical to conjecture that the number of different occupancy "complexions" encompassed by all possible surfaces passing through the volume with a common perimeter is 2^N, where N is the total number of lattice sites in the volume as a whole. Obviously, N corresponds essentially to the number of molecules, since most sites are occupied. Hence, when shape uncertainty is added to surface occupancy uncertainty, the information gained by a particular choice is roughly

$$H_{\text{shape}} = \mathrm{k}\ N \ln 2 \tag{1.2}$$

or N bits, given the assumption of equal probability of all surfaces as well as all occupancy-states. It follows that information (entropy) associated with a particular choice of surface is roughly comparable in magnitude to the thermodynamic information associated with a particular composition and distribution of quantum states. In the economic context, the equal probability assumption clearly does not apply to shapes. Some shapes are far more probable than others, resulting in a vast reduction in the information value of a particular choice in the "real" world, as compared with the theoretical possibilities. As a consequence, the amount of morphological information "embodied" in selecting the shape of a real object (such as a washer, bolt, piston, or bearing) is comparatively small in magnitude, as will be seen in the next section.

1.3. Morphological information embodied in manufactured shapes

To estimate the magnitude of "useful" shape information – in contrast to the information associated with all possible shapes – one must abandon the combinatorial methods of statistical mechanics and approach the question in terms of geometry and dimensional precision. The two can be discussed together, since dimensional precision is a part of parametric specification. In practice, most part shapes are quite simple or are constructed from simple geometrical elements, such as straight lines, circles, and angles. It is convenient to divide part shapes into two basic groups, viz., two-dimensional planar (flat), and three-dimensional (prismatic). Each group can be further subdivided, based on symmetries and whether or not the shapes can be obtained from simpler symmetrical shapes by bending, winding, or by stretching/shrinking. Thus, nine major shape categories are shown in *Table 1.1.* It can be seen that simple shapes are generally constructed by sequences of geometrical operations, such as rotations, translations, and intersections. Many of these geometrical operations have counterparts in physical manufacturing processes, though physical processes do not always correspond exactly to geometrical ones.

The information embodied in a complex geometrical shape defined by the intersection of surfaces consists of two components:

(1) Dimensional specifications for each surface.
(2) Construction instructions for combining several surfaces of specified types (planes, conic sections, etc.).

These two components are combined if each surface is completely defined and oriented with respect to a single common Cartesian (or other) coordinate system. This orientation requirement applies to surfaces that are symmetric with respect to rotation around an axis (or three axes in the case of spheres). To "orient" such a surface, simply imagine that coordinates are fixed and imprinted on its surface (as a map of the world is imprinted on a globe). A plane or flat surface is completely defined by the four parameters of a first-order (linear) equation of the form:

$$ax + by + cz + d = 0 \tag{1.3}$$

Only three of these parameters are independent. A point in space or a vector from the origin are also defined by three parameters. A conic section (ellipsoid, paraboloid, or hyperboloid) is completely defined by the ten parameters (nine independent) of a second-order (quadratic) equation of the form:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + ky + l = 0 \tag{1.4}$$

By extension, more complex surfaces are defined by cubic, quartic and higher-order equations with 20 (19), 35 (34), or more parameters, respectively. Note that many of the simpler and more familiar surfaces are defined by specialized quadratic forms in which many of the nine parameters of a quadratic can be collapsed into a smaller number. For instance, a located sphere is fully defined by its radius and the distance and direction of its center from the origin (four parameters). A located and oriented cylinder of infinite length is defined by its axial vector (the intersection of planes) plus a radius, or seven parameters in all. Indeed, a plane is also a special (three-parameter) case of a generalized quadratic form, in which most of the parameters are set equal to zero. Similarly, a quadratic is a special case of a cubic, and so on, to higher and higher orders. By this logic, an arbitrary shape can be regarded as a locus of intersecting surfaces of Pth order.

Without some *a priori* limitation, the information embodied in a material shape would be of the order of the number of molecules in the volume of material, or N bits, as pointed out above. This reflects the fact that any given shape is, in a sense, selected from among an enormous variety of possible shapes. However, it seems likely that in practice any surface likely to be required can be adequately approximated "piecewise" by a finite number of plane or quadratic surfaces.

A rather deep question must be addressed at this point. Obviously, it makes a big difference in the information content *whether a surface is regarded as a special case of a general Pth-order equation or whether it is "prespecified" as a plane, i.e., selected from the category of plane surfaces.* In the former case, the surface is defined in terms of only three independent parameters, while for the Pth order it is defined by

$$N = \sum_{j=1}^{P} [\sum_{i=0}^{j} (i + 1)] - 1 \tag{1.5}$$

independent parameters (of which all but four are numerically set equal to zero). Assuming, for purposes of argument, that the second point of view is the appropriate one, i.e., that many surfaces can be prespecified as planes and hence defined by only three independent parameters, a similar problem arises where there is a second plane surface parallel to the first. Should this be regarded as requiring three more independent parameters or only a single parameter corresponding to the distance between them? *Again, the answer depends upon whether the parallelism is a condition of belonging to a particular shape category or not.* Note that strict parallelism is a condition of belonging to the category of flat parts (category 1 in *Table 1.1*).

The analogous situation arises for some curved surfaces. In the case of cylinders and spheres, with their common centers or axes of rotation, it can be termed concentrism, but it is effectively equivalent to local parallelism. Note that local (as opposed to strict) parallelism is a defining characteristic of the major surfaces of all the shapes in categories 2, 3, and 4 of *Table 1.1.*

Table 1.1. Parts shape taxonomy, based on symmetry.

Category	*Shape specification*	*Examples*
1. Flat shapes (x–y plane); made by cutting, shearing, or punching	– Thickness – Edge profile is in x–y plane	Fabric parts, can tops, washers, generator core laminations
2. Distorted x–y planar shapes (with extension in z direction), invariant under any rotation around z axis; made by drawing or stamping	As #1 plus – profile of intersection with x–y plane	Metal cups, cans, tire rims, spherical or parabolic reflectors, roller-bearing races, stamped wheel hubs
3. Flat shapes made by folding and/or bending around z axis without surface distortion	As #2	Paper or metal boxes or origami shapes, tubes wound from strip, metal gutters
4. Nonsymmetric; made by stamping	– Thickness – Specification of nonflat surface – Edge profiles on surface	Auto body shapes, etc.
5. Prismatic (3-D) shapes that are invariant under rotations around z axis; made by molding, rolling, grinding, turning, drilling, or boring	– Profile of intersection with x–z plane – Specification of end plane(s) or surfaces	Ball bearings, pins, nails, bushings, piston wheels, axles, tires
6. Prismatic (3-D) shapes that are invariant under pure translation along z axis	As #5	T-beams, H-beams, rails, piles, decorative moldings, windowframe extrusions, wheel spokes
7. Prismatic (3-D) shapes that are invariant under a set of finite rotations around or translations along z axis; made by rolling, turning, or milling	As #5 plus – profile of intersection with x–y plane	Threaded connectors, worm gears, helical gears, bevel gears, spur gears, sprockets
8. Prismatic (3-D) shapes made from translationally invariant shapes by winding or bending without distortion	As #5 plus – specification of curve in some plane or 3-D space	Electrical windings, wire springs, hangers, hooks, pipe systems, crank handles, horseshoes, chain links
9. Prismatic (3-D) shapes that are nonsymmetrical; made by casting, molding, hammer forging, and/or milling	Specification of intersecting planes or curved surfaces	Engine blocks, base parts for machines, turbine blades, crank-camshafts, connecting rods, cutlery, hand tools

Similarly, since rotational symmetry is a condition of belonging to category 5, all the rotational surfaces are (by definition) concentric cylinders. Once the central axis is located and oriented, only one additional parameter is required for each. Incidentally, the same argument holds for cylindrical holes, given that the shape category can be prespecified. The argument applies, with some modification, to rotational surfaces in category 7.

Finally, translational invariance is a condition of belonging to category 6. This implies that the entire external surface (except for the ends) is defined by a cross-section, i.e., a closed perimeter defined on a two-dimensional plane. The perimeter may be defined by intersections of straight lines, quadratics, or higher-order curves. (However, as noted previously, one can assume that virtually any higher-order curve can be adequate approximated piecewise by several quadratics.) The information required to specify any shape in category 6 is, therefore, the information required to specify the perimeter of the translationally invariant (long) segment plus the two bounding surfaces at the ends. The simplest case is, of course, a cylinder – also belonging to category 5 – where the cross-sectional perimeter is a circle.

If each shape category in *Table 1.1* were equally probable, the information equivalence of a choice of category *per se* would be $\log_2 9 \simeq 3.2$ bits. In practice, simpler shapes predominate, so the information embodied in a shape choice such as category 1 (flat), 5 (rotationally invariant), or 6 (translationally invariant) is even less – probably of the order of two bits or so. On the other hand, the truly nonsymmetrical shapes belonging to categories 4 or 9 are much less common – hence, less probable. For instance, the information equivalence of such an *a priori* choice is greater than 10 bits if the *a priori* probability is less than 10^{-3}.

Of course, the classification in *Table 1.1* is not unique. Nor is it sufficiently detailed for serious analysis. As it happens, a number of practical parts classification coding schemes have been introduced over the past 20 years under the rubric Group Technology (or GT) to facilitate the design of manufacturing systems. [See, for example, Burbridge (1975); Devries *et al.* (1976); Edwards (1971); Gallagher and Knight (1973); Ham and Ross (1977); Mitrafanov (1966); and Opitz (1970).] For example, Opitz's scheme is a five-digit (decimal) code. The Opitz code for a hexagonal nut is 30500. In this case the first digit (3) implies that the part is roughly rotational, with deviation from perfect symmetry, with a length/diameter ratio <2. The second digit (0) implies hexagonal shape. The third digit (5) implies a rotational *internal* shape with screw threads. The fourth digit (0) implies flat, unstructured, plane external surfaces; and the fifth digit (0) implies an absence of auxiliary holes in gear teeth (see *Figure 1.1*).

Again, it is evident that, although there are 10^5 different Opitz code possibilities, they are not equally likely to occur. The notion of a "random" part is itself not well-defined. However, as a "gedanken" experiment, one can suppose that all manufactured products produced on a typical day next year are collected in a large warehouse. Suppose that each product is then dismantled and reduced to its individual component parts. The pile might contain 109 individual parts or more. Finally, suppose that all the parts are coded and sorted out into 105 file drawers by Opitz code. Some code numbers (such as 30500) would be used a great many times, whereas some others would be used very seldom.

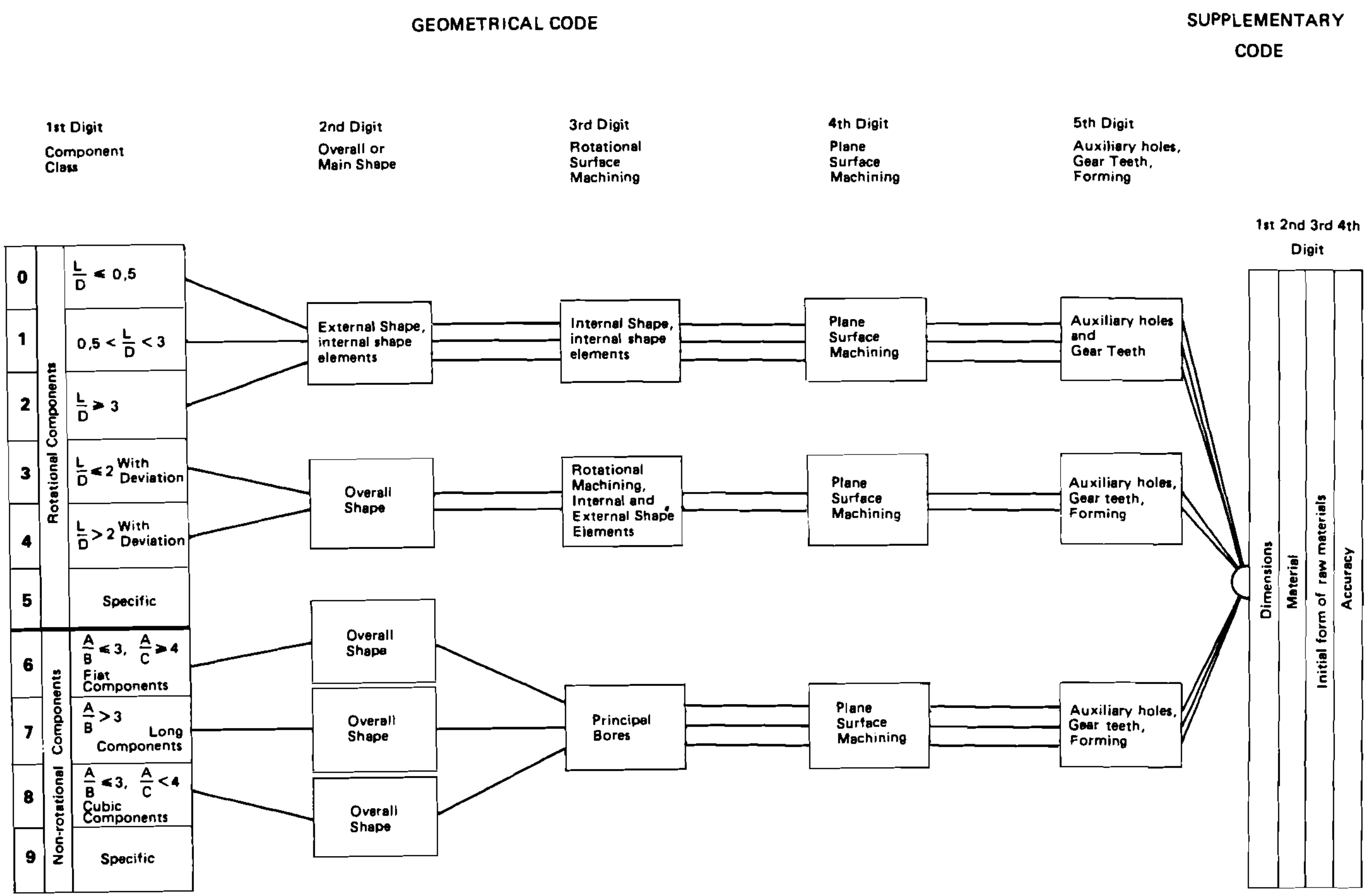

Figure 1.1. Flourey workpiece classification. Source: Opitz (1967).

The frequency distribution of parts across all possible code numbers should be roughly independent of the size of the pile of parts. This frequency distribution describes the current relative probability of each code number. ***The information "contained" in the ith code number is the negative logarithm (base 2) of its probability of occurrence:***

$$H_{\text{code}(i)} = -\log_2 P_i \tag{1.6}$$

For shapes that are frequently encountered, the probability is relatively high and the information value is relatively low (3–8 bits). Conversely, for very infrequently encountered or unique shapes, the information value can be arbitrarily high. [The problem of determining the information content associated with a unique or unusual shape can probably best be approached in another way. It has been suggested that the appropriate method of computation is to count the number of instructions required to program a numerical control (NC)-machine tool to cut the shape. For a turbine blade, for instance, the program might require several thousand instructions (= bits).]

Once the shape category is fixed, the range of possibilities for parameter specification is much reduced. One can suppose, for convenience, that each of these parameters is a number (in some system of measuring units) specified to a standard accuracy of 1 part in 1,000. This level of accuracy, or tolerance, is essentially equivalent to 10 bits of information per parameter, because $2^{10} = 1024 \simeq 1{,}000$. Hence $\log_2 1{,}000 \simeq 10$ bits. (In cases where lesser or greater accuracy is needed, a suitable correction term is subtracted or added.) It follows that a localized plane surface defined to standard accuracy "embodies" about 30 bits, while two parallel planes separated by a fixed distance embody 40 bits. An intersection of N localized planes ($N \geq 4$) therefore embodies $30N$ bits. An intersection of M pairs of parallel planes ($M \geq 2$) embodies $40M$ bits. The intersection of three localized pairs of parallel planes embodies $3 \times 40 = 120$ bits of information.

The above is actually an overspecification, since it also includes both location and orientation information that is not needed to define the shape itself. Location and orientation in Cartesian space requires six parameters (or 60 bits). A "pure" shape is (by definition) invariant under all translations of the center of mass and all rotations around its center. Location and orientation require six parameters to be specified, or 60 bits of information. This must be subtracted. The total amount of information embodied in simple shapes is therefore determined by (a) the shape category, (b) the number of surfaces, and (c) the form or order of the surface-defining equation (e.g., quadratic, cubic, etc.).

A tetrahedron defined by the common intersection of four nonparallel planes, independent of position or orientation, therefore embodies $(4 \times 30) - 60 = 60$ bits or 10 bits for each of six edges that can be independently fixed. In the case of the parallelepiped above, there are also 60 bits of embodied information, corresponding to 10 bits for each of three independent edges and three independent angles. For the set of shapes defined by the intersection of a plane surface and a conic section, the embodied information is, again, $30 + 90 - 60 = 60$ bits. (For an ellipsoid, such an intersection defines two or more shapes, as when an orange is cut into two parts. To specify which of the two parts is intended requires one additional bit of information.) By similar logic, a shape defined by a localized conic section (90 bits) intersected by two nonparallel planes (60 bits) embodies 60 + 90 - 60 = 90 bits. A generalized conic section intersected by two parallel planes (40 bits) embodies 40 + 90 - 60 = 70 bits. In both of these last two cases, there may be up to four distinct regions of intersection. To select one of the four possibilities requires an additional two bits ($\log_2 4$).

In practice, it would normally be most convenient to use a GT code, such as the Opitz code. Once a part is classified, the number of independent dimensional parameters is easily determined from the code. For example, in the case of the "hex nut", there are six dimensional specifications altogether: external diameter, thickness, internal radius, and depth, width, and pitch of the thread groove. Each of these six parameters corresponds to 10 bits of information in the case of "standard" precision. In addition, the screw thread may be right-handed or left-handed. Thus, in addition to the code itself, the complete specification involves exactly 61 bits of information.

To summarize: the amount of morphological information embodied in a simple manufactured shape can be computed as follows:

(1) Determine the required precision of parametric specification. In general, it is reasonable to assume 10 bits per parameter.
(2) Determine the categorical specification, using any general purpose GT code. (This could be a major research project, of course.)
(3) Determine the frequency distribution of part shapes among possible categories defined by the code. The code information is the logarithm of the (inverse) frequency of the code for the shape.
(4) Determine the number of parameters needed to select a specific shape *within* the designated code category, taking into account parallelism and concentrism, as appropriate, and multiply by 10 bits per parameter. Add this number to the code information.

More complex shapes can be constructed geometrically by combining or superimposing simpler shapes, either positive or negative (holes). The combination process is closely analogous to assembly, as will be seen. For instance, a rivet is a simple cylindrical shape with a hemisphere at one end. A cylindrical hole is a negative cylindrical shape superimposed on a positive shape. Complex shapes in the real world may be formed by a sequence of forming operations, starting with a simple shape. Alternatively, complex shapes can be *assembled* or *constructed* from simple shapes. This is discussed further in *Appendix 1.B.*

1.4. The relative prices of metabolic and morphological information

As suggested previously, manufacturing can be considered as consisting of two distinct information conversion processes:

(1) Metabolic (or materials processing) activities.
(2) Morphological (shaping or forming) activities.

It follows that manufacturing value added can be subdivided into two components, viz.,

$$V_{\text{mfg.}} = V_{\text{metab}} + V_{\text{morph}} = P_{\text{metab}}\, H_{\text{metab}} + P_{\text{morph}}\, H_{\text{morph}} \tag{1.7}$$

If the US manufacturing value added ($585 billion in 1977) is crudely divided into metabolic and morphological components, the former would seem, at first glance, to include most or all of the following processes:

- extraction or winning
- beneficiation (physical separation)
- digestion or leaching
- carbothermic or electrolytic reduction
- refining (including petroleum)
- alloying
- chemical synthesis
- food processing
- dehydration, calcination
- distillation and related separation processes

Industries using these processes are also the greatest users of energy, in relation to value added, as will be shown hereafter. The five most energy-intensive sectors are as shown in *Table 1.2*.

Table 1.2. Metabolic process activities.

Sector	*Purchased energy 1980* (10^{12} BTU)	*Process total energy*[a] *1980* (10^{12} BTU)	*Value added 1977* ($billions)
Chemicals	2,717	3,354	56,721
Primary metals	2,277	3,712	37,568
Petroleum refining	1,178	3,061	16,378
Pulp and paper	1,278	2,328	22,171
Stone, clay and glass	1,132	1,132	19,130

[a]Excluding feedstock energy ultimately *embodied* in product, but including "waste fuels" derived from feedstock. Energy data for 1980 were compiled by Doblin (1985) from the Census of Manufactures and various special surveys. Unfortunately, comparable data for more recent years have not been published. Value added figures are also difficult to obtain, not being published on a regular basis; 1977 was the nearest year I could find. Using figures for exactly corresponding years would not affect the conclusions significantly.

These five sectors together accounted for about $152 billion in value added, and 8.55×10^{15} BTU (quads) in *purchased* energy consumption and about 13.6 quads in *total* energy consumption. The difference is accounted for by energy derived from waste materials, such as wood. Part of the energy is used to transform crude, petroleum, and chemical feedstocks into more useful forms, and part of it is used to transform fossil fuels into electricity. (Electricity is counted at its thermal equivalent value: 3,412 BTU per kWh.) Detailed process analysis (e.g., Battelle, 1975) shows that most of the remainder is used to separate metals from ores and to increase the free energy or available useful work in structural materials prior to subsequent shaping, forming, and assembly processes. Of course, the free energy in the fuels and electricity used in the manufacturing processes is simultaneously dissipated and lost.

In the case of the chemicals industry, the biggest user of process energy, simple mineral and hydrocarbon feedstocks [mainly sulfur, sodium chloride, nitrogen, oxygen, methane, propane, butane, and benzene, toluene, and xylene (B-T-X)] are first converted to more reactive chemicals, such as sulfuric acid, chlorine, caustic soda, hydrochloric acid, ammonia, acetylene, ethylene, propylene, methanol, ethanol, and so on. Except for the production of sulfur dioxide from sulfur, most of these first stage reactions are endothermic, which means they require substantial amounts of process energy from an external source. This where the "purchased energy" in the chemical industry is largely used. In most cases, the chains of subsequent reactions to produce more complex chemicals are actually exothermic or self-energized. (There are, obviously, many exceptions. For instance, several important polymerization reactions are endothermic and cannot proceed spontaneously.) Here the energy is extracted from the intermediates, whence the free energy of the products is less than the free energy of the intermediate inputs. One estimate (Burwell, 1983) puts the "nonpurchased" fraction of total energy at 15%, implying that about 25%–30% of feedstock energy is lost in conversion.

In the case of the petroleum industry, process energy is used both for separation (distillation) and for cracking, reforming, alkylation, and other processes to increase the gasoline yield per barrel of crude oil and to purify the products, especially by removing sulfur. The industry both consumes "feedstocks" – mainly liquid propane gas (LPG) from natural gas liquefaction plants – and produces them – mainly ethylene, propylene, butylene, and B-T-X. Roughly 10% of the free energy originally in the crude oil is lost in these various conversion processes.

In the case of pulp and paper, process energy is needed mainly to get rid of excess water and recycle the various leaching chemicals. About 40% of the total process energy used in the industry is now derived from the burning of waste lignin and cellulose. In principle, this figure could be much larger, but more efficient use of the biomass energy is inhibited by the large amounts of water used in all the pulping and digestion processes. An idealized papermaking process would produce net free energy, not consume it.

The primary metals industry has four distinct branches: ferrous and nonferrous, primary and secondary. The primary ferrous branch extracts iron ore (Fe_2O_3, Fe_3O_4), smelts the ore with coke in a blast furnace, and then refines the impure pig iron by reacting it with oxygen in a basic oxygen furnace (BOF). The final product – pure iron or carbon steel – has a much higher free energy than the ore from which it was extracted. (It could actually be "burned" again as fuel.) However, the overall process involves a loss of all the free energy in the various fuels, especially coking coal. Overall efficiency in these terms is currently around 30% of "ideal" efficiency (Gyftopoulos *et al.*, 1974). So called "electric" steel and ferro-alloys are the secondary branch of the ferrous metals sector. It is based on remelting and repurifying scrap. In this case, there is essentially no gain or loss in free energy of the steel, although the electric energy (for melting) is lost.

Primary nonferrous metals can be subdivided into those with oxide ores (Al) and sulfide ores (Cu, Pb, Zn). In both cases the ore beneficiation process is very energy-intensive. Aluminum ore (bauxite) is first converted to nearly pure alumina (Al_2O_3) by a chemical leaching–dehydration process. The dry alumina is then reduced to metal in an electrolytic cell. The process is highly endothermic. The free energy of the product aluminum is, of course, much greater than that of the ore. (In principle, it, too, could be burned as a fuel.) However, much more process energy is lost.

Copper lead and zinc are chalcophile (sulfur-loving) metals. Their ores tend to be rather low in grade and must usually be finely ground and beneficiated by a physical process, such as flotation–filtration. This, again, is very energy-intensive. Subsequently, the concentrate is "roasted" to drive off sulfur and arsenic – subsequently recovered in modern plants – and the concentrate is then smelted in a furnace. A final electrolytic purification stage is needed for copper. The first (roasting) stage is theoretically exothermic, although fuel is used to speed it up, but the second stage is endothermic. In principle, the combined roasting–smelting process with sulfur recovery ought to produce net free energy; in practice it never will. The final purification steps to eliminate or recover minor impurities, such as gold and silver, cadmium, tellurium, and selenium, are quite energy-intensive in the aggregate. In fact, pure copper requires nearly as much energy to produce, in practice, as aluminum.

The stone, clay, and glass sector consumes energy mainly in the manufacture of quicklime (CaO), hydraulic cement, and plaster of paris, and in the melting of glass. The first and second of these involve calcining – the use of heat from fuel to drive CO_2 and H_2O away from hydrated calcium carbonate (limestone). The third process (to make plaster) is a simple dehydration – the use of heat to drive H_2O away from a mineral calcium sulfate (gypsum). The resulting materials are eager to recombine with water, yielding an inert mineral solid and releasing heat in the process. This actually occurs when these building materials are used by the construction industry. Thus, the free energy in both the initial and final materials is equally zero. The process energy used in this materials sector has only one practical function, namely, to enable the materials to be "fluidized" for purposes of forming and shaping. The same thing is also true for glass. Brickmaking, also in this sector, is essentially a forming–shaping activity.

Evidently, part (perhaps 20%) of the energy used in the primary metals sector for melting and casting and all of the energy used in stone, clay, and glass sector are really attributable to forming and shaping, not extraction or refining.

Table 1.3. Morphological (forming–shaping) activities.

Sector	*Purchased energy 1980* (10^{12} BTU)	*Value added 1977* ($billions)
Fabricated metal products	362	45,512
Nonelectrical machinery	337	67,223
Electrical machinery and electronics	240	50,366
Transportation equipment	344	64,291
Instruments and related products	80	18,762
Total	1,363	246,200

By comparison, the five sectors covered in *Table 1.2* accounted for $246.2 billion in value added and 1.4 quads of energy consumed, mostly as electricity (see *Table 1.3*). The products of these sectors are metal components or machines and instruments of varying degrees of sophistication. Many processes are used in these sectors, but most of the energy is used for the following:

- casting (foundry)
- forging, pressing, or rolling
- stamping, bending
- cutting (drilling, boring, machining)
- grinding
- welding and soldering
- assembly

Much less energy is consumed per dollar of valued added, and the free energy content of the final products is invariably *less* than the free energy content of the purchased materials from which they are made. If the energy used in the stone, clay, and glass sector and 20% of the energy used in the primary metals sector (and 30% of the valued added) are attributed to forming and shaping, then we have roughly the following summary comparison (considering only 10 sectors):

(1) Metabolic activities: separation, reduction, refining, purification, and synthesis of materials:

  (a) Value added (1977) – $141 billion
  (b) Energy (1980) – 12.1×10^{15} BTU

(2) Morphological activities: melting or liquefaction of materials for purposes of forming–shaping, forging, bending, pressing, cutting, grinding, joining, and assembly:

  (a) Value added (1977) – $257.5 billion
  (b) Energy (1980) – 2.87×10^{15} BTU

For completeness, it may be noted that the remaining 10 manufacturing sectors normally included in manufacturing had a total value added of $186.5 billion and a total energy consumption of less than 2×10^{15} BTU. Nearly half of this energy was used in the food processing sector, which is more nearly metabolic than morphological.

Summarizing, it is clear that in the manufacturing sectors

$$V_{\mathrm{metab}} < V_{\mathrm{morph}} \tag{1.8}$$

whence

$$P_{\mathrm{metab}}\, H_{\mathrm{metab}} < P_{\mathrm{morph}}\, H_{\mathrm{morph}} \tag{1.9}$$

It was established in the preceding sections that, if H is always measured in bits, the useful shape information for a standard machine part is of the order of magnitude of 10 bits per parameter (for precision of 10^{-3}) plus a few more bits for the code specification. The total would usually be less than 100 bits and seldom more than 1,000 bits. In the assembly process, information is lost, not gained, so the morphological information embodied in a machine with 1,000 parts would be of the order of 10^5 to 10^6 bits, even allowing for a few specialized parts with moderately complex shapes.

A highly sophisticated machine, such as a helicopter, might have 10^5 or possibly 10^6 parts, many of them individually complex and requiring high precision. Yet the total morphological information content could scarcely exceed 10^8 or 10^9 bits. By comparison, the thermodynamic information embodied in any metal alloy or synthetic chemical is likely to be of the order of 10–100 kT per mole or 10^{24} to 10^{25} bits/mole. (A helicopter or jet engine would require 10^3 or 10^4 moles of mass.)

An obvious implication of these facts is that the ratio of metabolic to morphological information in the economic system is currently in the neighborhood of 10^{20}. Hence, for the manufacturing sectors of the economy, it is certainly accurate to say that

$$H_{\mathrm{morph}} \ll H_{\mathrm{metab}} \tag{1.10}$$

It follows from (1.9) and (1.10) that

$$P_{\mathrm{morph}} \gg P_{\mathrm{metab}} \tag{1.11}$$

The foregoing analysis can also be used to estimate, at least very roughly, the actual value of P_{metab}, assuming H_{metab} is proportional to the free energy or available useful work B consumed or dissipated. This, in turn, is essentially equal to the free energy stored in the fuels used up or, using equation (A.2) in *Appendix 1.A*

$$\Delta H_{\mathrm{metab}}\,(\mathrm{bits}) = \Delta B / T_0 \tag{1.12}$$

Here T_0 is the temperature of the ambient environment (i.e., the earth's surface), and ΔB is the change in available free energy. Thus,

$$P_{\text{metab}}(\$) = \frac{V_{\text{metab}}(\$)}{H_{\text{metab}}(\text{bits})} = \frac{V_{\text{metab}} \times T_0}{B} \tag{1.13}$$

where P_{metab} is given in \$/bit.

Substituting $V_{\text{metab}} \simeq 140 \times 10^9$, $T_0 = 300^\circ\text{K}$ and $B \simeq 12 \times 10^{15}\,\text{BTU} = 1.266 \times 10^{19}$ joules, one finds approximately

$$P_{\text{metab}} \simeq 3.3 \times 10^{-6} \quad (\$/\text{joule}^\circ\text{K})$$

Each joule/°K is equivalent to 10^{23} bits, so the price per bit is, very roughly,

$$P_{\text{metab}} \simeq 3.3 \times 10^{-29} (\$/\text{bit})$$

By the above arguments, we see that P_{morph} is of the order of 10^{20} larger or, roughly,

$$P_{\text{morph}} = 3 \times 10^{-9} (\$/\text{bit})$$

In words, it costs about 10^{20} times as much to embody a bit of *morphological* information in a manufactured product as it does to use a bit of information (as free energy) in thermodynamic or metabolic processes. As noted earlier, these results are highly insensitive to the exact years for which the data were taken.

1.5. Information and value

Theories attempting to relate economic value to a single factor (such as labor or energy) have a long and somewhat disreputable history in economics. It must be emphasized at the outset that no such notion is contemplated here. To be sure, I do argue that labor skills, capital, available energy and technology are all more or less embodied – or "condensed" – forms of information. It does *not* follow that the market price of a given product or service is (or should be) simply or directly related to its numerical information content. In particular, there is no justification for confusing thermodynamic and morphological information in this regard.

A far more plausible possibility is that condensed or embodied information of a given kind has a relatively well-defined cost per unit. Two examples will be presented briefly in this section. The first example pertains to the cost of physical separation of pure substances from mixtures or solutions. In this case, the process is metabolic, and the major cost element is energy. I also show that the cost tends to be a linear (or near-linear) function of the equivalent information added by the separation process. The second example pertains to the cost of accuracy in manual machining, a very labor-intensive process. Evidence is set forth suggesting that the cost of increasing precision is a highly nonlinear function of the equivalent morphological information embodied.

Example 1: Cost of separation–concentration

As stated in *Appendix 1.A*, the information embodied per mole by concentration, or lost by diffusion, can generally be approximated by Boltzmann's ideal gas approximation:

$$H_c = \mathrm{R} \ln(X_c/X_0) = \mathrm{R} \ln c \qquad (1.14)$$

where X_c is the mole fraction in the concentrated state, X_0 is the mole fraction in the diffused state, and R is the ideal gas constant (˜2 cal/mole). The ratio of mole fractions is equal to the concentration ratio c.

It is important to note that the incremental information added by concentration depends on the starting point. Not much is gained by starting from a highly concentrated source. However, if we are interested in comparing the information embodied in different materials in absolute terms, the best way is to calculate the information that would be lost if the material were completely dispersed (diffused) into the environment. (The appropriate definition of "environment" would be the earth's crust for most solids, oceans for water-soluble salts or liquids, or the atmosphere for gases.) The difference between information lost by diffusion, and information added by concentration from high-quality natural sources, is, of course, a gift from nature.

It has been suggested, e.g., by Sherwood (1978), that costs (or prices) of many materials are inversely proportional to their original concentrations in their "ore" or original form, and therefore proportional to the concentration factor needed to purify them. This relationship implies a linear relationship between the logarithms of price (cents/lb) and concentration factors. Such a relationship is indeed observed for many materials, particularly where several different initial concentrations. This applies to various minerals, such as vermiculite, diatomite, graphite, asbestos, sheet mica, and gold. It also applies to atmospheric gases (oxygen, nitrogen, argon) and to various chemicals found in brine. See *Figure 1.2*.

It must be noted that a linear relationship between cost–price and concentration factor is *not* a linear relationship between cost–price and embodied information. In fact, it is the logarithm of cost that is proportional to embodied information. In other words, the cost–price seems to be, on average, an exponential function of embodied information associated with physical concentration:

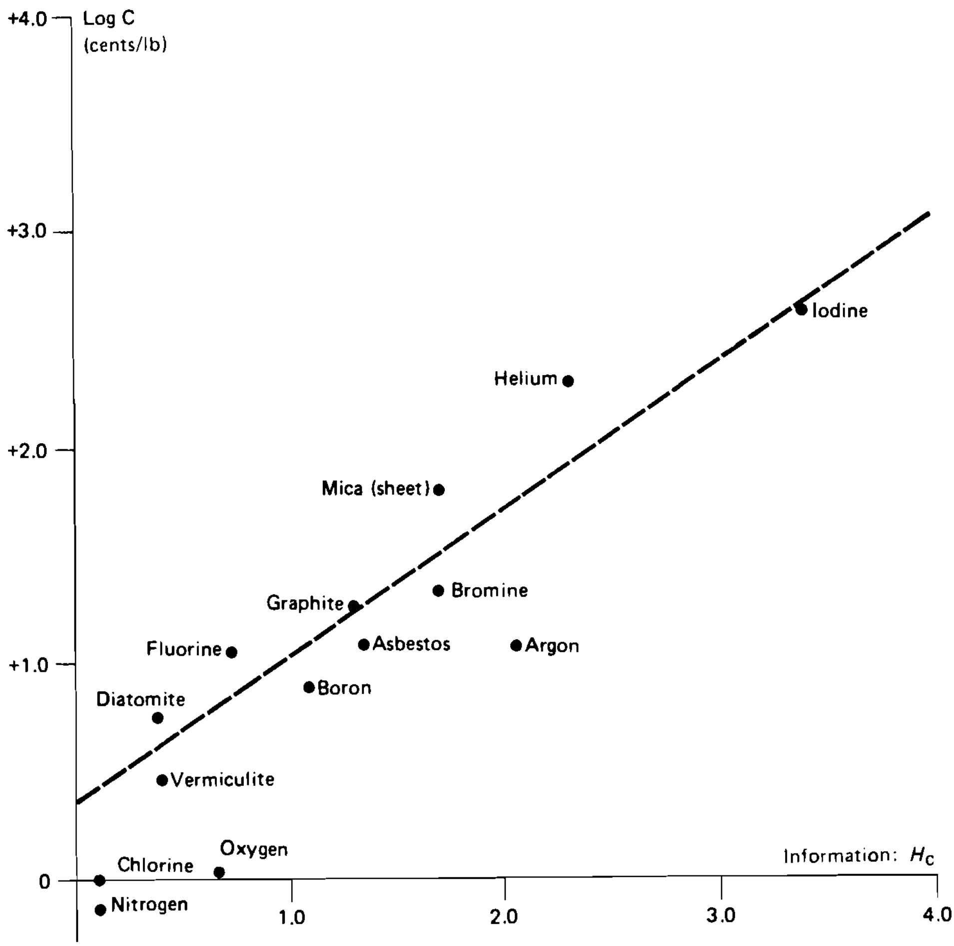

Figure 1.2. Cost of separation versus information added.

$$C \simeq \exp H_c \tag{1.15}$$

In *Figure 1.2*, I have compared only materials requiring *physical* separation, which eliminates one of the complicating factors. Consider for instance, the extremely complex multi-stage process for refining platinum group metals from their ores versus the extraordinarily simple process for refining mercury from its ore (simple low-temperature retorting). Consider also the vast differences between by-products, such as arsenic, and primary products in this regard. Still it is more than a little surprising to observe an apparent clear relationship between cost–price and concentration, in view of the enormous differences, for instance, in scale of production or use among different substances. Further, there are differences in inherent utility from one material to another. As an example, consider the great inherent utility of platinum as a catalyst compared to osmium, an equally rare metal of the platinum group with no known uses whatever.

Example 2: Cost of increasing precision

In machining operations, the information $H(t)$ required to achieve a tolerance t can be written as

$$H(t) = \mathrm{K}\log_2(1/t) = -\ \mathrm{K}\log_2 t \tag{1.16}$$

where K is a constant determined by the unit of information (e.g., bits) and t is usually defined for convenience as the maximum allowable machining error per unit (inch) of linear tool travel on the workpiece.

Table 1.4. Cost-tolerance relationship.

Tolerance t	$H(t)$ (bits)	*Relative cost* (Boltz, 1976)
$.064 \sim 2^{-4}$	4	0.75
$.048 \sim .05$		1.0
.040		1.2
$.032 \simeq 2^{-5}$	5	1.5
.024		2.0
.020		2.4
.016	6	3.0
.012		4.0
.008	7	6.0
.006		
.004	8	12.0
.003		
.002	9	24.0
.0015		
$.001 \simeq 2^{-10}$	10	48.0

A cost-tolerance relationship taken from a standard engineering handbook (Boltz, 1976) is given in *Table 1.4*. While the relative cost figures given are only approximate (taken from a graph – see *Figure 1.3*), it is clear that the relative cost is not a simple linear function of information content. In fact, in the normal range of tolerances from 2^{-5} to 2^{-10} where precision increases by a factor of 2^5 = 32, the information content only doubles. This implies that relative cost increases as the fifth power of relative information content or precision, viz.,

$$C_1/C_0 \simeq (H_1/H_0)^5 \tag{1.17}$$

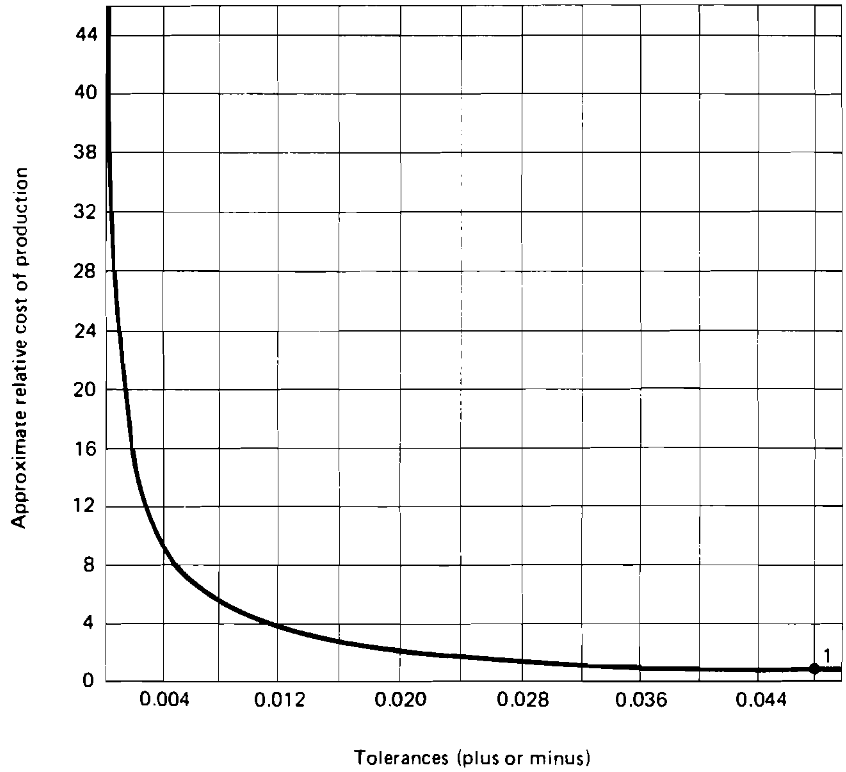

Figure 1.3. Cost versus tolerance relationship. Source: Boltz (1976).

This relationship presumably reflects a body of experience with manually operated machine tools. Therefore, it can be interpreted as an average relationship between skilled machinists' time inputs as a function of information actually embodied in the workpiece. Another way to look at it is to note that embodied information is a very small fractional power of cost, viz.,

$$(H_1/H_0) \sim (C_1/C_0)^{1/5} \tag{1.18}$$

The time-information relationship for human labor is discussed in the second part of this Research Report. One is roughly proportional to the other. That is, the amount of visual or tactile information required for motion control determines the time required for an elementary motion. Why, then, should the amount of information *embodied* in the product not be simply proportional to labor time? No definitive answer can be given here. However, it would seem likely that the nonlinearity of the cost-precision function results from the onset of mental overload as a human machinist approaches the limits of his natural sensory discrimination ability. As is discussed in more detail in the second part of this Report, each successive manual correction becomes more and more difficult as the discrimination limit is approached. Therefore, it takes longer.

A numerically controlled machine tool with internal feedback control might also be ultimately limited by the inherent discrimination ability of its detectors or built-in "senses", but specialized electronic devices can be designed to be much more sensitive in a particular domain than the general purpose human sensory system. This is why NC or CNC (computer numerical control) machines can perform high-precision machining operations far faster than humans, although they offer little benefit in the case of simpler operations or lower precision. One would therefore expect the relationship between control information input and information embodied in output to be much more nearly linear for NC or CNC machines.

1.6. The principle of minimum morphological information

Four conclusions follow immediately from the observation (in Section 1.4) that morphological information is extremely expensive compared to thermodynamic or metabolic information.

First, insofar as there is a clear tradeoff between materials composition and design (form and shape), the former should always be preferred. In practice, this means materials should be selected, wherever possible, to *simplify* design and, in particular, to minimize forming–shaping and assembly operations.

Second, it appears to follow that materials research and development should focus on two objectives:

(1) Making materials that are easier to form and shape.
(2) Giving materials new physical properties, especially by deliberately introducing inhomogeneities (such as alternate layers of conductors and insulators) or properties that can be altered reversibly by external influences, thus permitting "monolithic" materials to substitute for assemblies of many parts. [It hardly needs to be pointed out that integrated circuitry and silicon chips exemplify the monolithic approach. However there are a number of other less familiar examples. See Ayres (1986).]

Third, manufacturing research should focus on ways to increase the overall efficiency of forming, shaping, and assembly operations. Given that the design specifies shape, dimensions, tolerances, and the choice of material, it still remains to choose an optimal method of shaping and dimensioning to the required tolerances. In practice, this involves the selection of a sequence of unit operations that will convert a "formless" piece of the desired material into its final shape. To achieve this conversion, there are three generalized strategies, which may be termed (1) material addition or build-up, (2) material subtraction, and (3) "pure" deformation.

In practice, all three of the above strategies – and combinations of them – are used in different situations. One cannot *a priori* eliminate any of them from consideration, because the functional requirements of the product may dictate constraints that make it impossible to select one strategy for all purposes. For example, a build-up or assembly approach is, to some extent, unavoidable if the product necessarily involves combinations of different materials, such as conductors and insulators. It is also unavoidable if there are moving parts in juxtaposition with fixed ones, especially if – as is often the case – the moving parts are largely contained within a fixed shell, as with a motor, compressor, pump, engine, or transmission.

Fourth, design research should seek to identify designs with *minimum morphological information*. This is consistent with some of the familiar rules of thumb, such as minimizing the number of distinct parts, especially connectors, and the number of high-precision interfaces. It is also essentially identical to the design axioms formulated by Suh *et al.* (1978; Nakazawa and Suh, 1984).

Nevertheless, until recently, designers have given relatively little consideration to seeking opportunities to minimize lost information in manufacturing. Boothroyd and Dewhurst (1983) may have been the first to see the potential. They have been able to demonstrate impressive savings at the design stage by reducing the number of distinct parts, such as connectors. The principle of parts minimization is clearly implied by the larger principle of minimum information in design. At the same time, it is also compatible with the principle of minimum information loss in manufacturing. Fewer parts mean fewer manufacturing operations, fewer assembly operations, and less cost.

To determine the minimum (theoretical) number of parts, Boothroyd and Dewhurst suggest the following criteria:

(1) Does the part move relative to all other parts already assembled?
(2) Must the part be of a different material than, or be isolated from, all other parts already assembled? Only fundamental reasons concerned with material properties are acceptable.
(3) Must the part be separate from all other parts already assembled because, otherwise, necessary assembly or disassembly of other parts would be impossible?

The number of distinct parts can only be reduced, in some cases, by making some individual parts more complex and costly. This may require more sophisticated methods in another manufacturing domain. Each case certainly requires detailed analysis. However, the foregoing analysis strongly suggests that, on balance, the extra cost of manufacturing some components will be more than justified by savings at the assembly stage.

The principle of minimizing embodied design information goes beyond the formulation of Boothroyd and Dewhurst. It also implies that the designer should seek to minimize the number of dimensional parameters needed to describe the product, as well as the precision with which those parameters need to be specified.

By a very similar line of reasoning, it can also be argued that any process involving the creation of a complex shape by subtraction (as a sculptor creates a shape from a block of stone) is also likely to be inherently inefficient. This is so, particularly, where the starting point is a precisely dimensioned metal block or cylindrical rod, much of which is subsequently cut away. Again, there is a loss of morphological information.

The ideal forming process, by the above criteria, would require neither the addition nor the subtraction of materials from the net final shape. Either addition or subtraction involves waste. Examples of "pure" deformation processes include injection molding (for plastics), die casting, and investment casting; a variant would substitute metal or ceramic powder for molten materials. Solid deformation processes, such as rolling, extrusion, bending, and forging, are also inherently efficient to the extent that they can be precisely controlled.

The loss of position–orientation information in assembly is a major cause of inefficiency. The problem is most acute in a traditional multipurpose machine shop, where individual machines are used successively to process batches of components and each batch of parts is piled in a bin prior to the next operation. As the part comes off the machine, its position and orientation is precisely determined. Yet the information is lost when the parts are dropped into the bin and must be reestablished later when a worker picks up the part, orients it, and positions it for the next operation. The benefits of mechanical transfer systems are easily explained in terms of *not losing* this information. An automatic transfer system, such as a conveyor belt, delivers the parts to the next station with reduced positional and orientational indeterminacy and, hence, less information is needed to reorient the part subsequently. A synchronous transfer machine can deliver parts in *exactly* the required position and orientation, with no loss of information.

Of course, such machines (like all forms of physical capital) embody a great deal of morphological information in their design and construction, which is gradually lost as the machine wears out. Unlike materials *per se*, morphological information is not conserved. However, the major drawback to more widespread use of such machines is their inherent inflexibility. Technological advances may ameliorate this drawback in the near future.

Thus, the trend toward hard automation in materials handling can be seen as a response to the imperative of reducing morphological information losses. Much the same can be said of recent trends to improve product quality by substituting computers for human workers in certain operations. Clearly, flaws and defects are a very costly form of lost information; and it is increasingly easier to reduce the error rate by introducing machines than to detect and correct errors by reworking once they are embodied in a faulty component or assembly. In summary, the principle of minimum morphological information loss seems to explain much of what is going on in industry today.

APPENDIX 1.A

Computation of Thermodynamic Information H_{thermo}

The relationships between entropy and information have been discussed extensively, e.g., by Shannon (1948), Brillouin (1951, 1953, 1962), and Tribus (1961a, 1961b). In a thermodynamic context, Evans (1969) has defined information as a measure of ***distinguishability of a system from its surroundings.*** The more distinguishable the system, the more information one has about it. Information can also be regarded as reduction of uncertainty, hence, as an entropy difference.

Thermodynamic information H_{thermo} lost in a process is defined in terms of the entropy S of an initially distinguishable system as follows:

$$H_{\text{thermo}} = S_0 - S = (S_0 - S_f) + (S_f - S) \tag{A.1}$$

where the system starts with energy E, volume V, and molar composition N_i of the ith chemical species. Note that S_f is the entropy of the same material system after it has reached thermal equilibrium with the environment, but without any material exchange with the environment (i.e., no diffusion). Of course, S_0 is the entropy of the final diffused state.

It is conceptually helpful to divide each process into two phases: one that goes to thermal equilibrium without material diffusion and a second consisting of diffusion alone. This is convenient since many industrial processes of interest leave the system in thermal equilibrium without any material mixing with the environment. On the other hand, some processes (such as combustion) do result in initially concentrated materials (e.g., fuels) being indistinguishably diffused or mixed with the atmosphere, as noted later.

Following the general procedure described by various authors, e.g., Evans (1969), Tribus and McIrvine (1971), Gyftopoulos *et al.* (1974), Ross *et al.* (1975), assume the atmosphere (environment) has temperature T_0, pressure P_0 and chemical potentials μ_{i0}. Consider a distinguishable subsystem with chemical species present in molecular concentrations n_i. Let

$$H_{\text{thermo}} = \left\{ \Delta U + P_0 \Delta V - T_0 \Delta S - \Sigma \mu_{i0} n_i \right\} / T_0 = B / T_0 \tag{A.2}$$

$$= H_{\text{chem}} + H_{\text{conc}}$$

The numerator B of (A.2) is, in fact, the available useful work, sometimes called essergy or exergy. Define

$$H_{\text{chem}} = \left\{ \Delta U + P_0 \Delta V - T_0 \Delta S \right\} / T_0 \tag{A.3}$$

and

$$H_{\text{conc}} = -(1/T_0) \Sigma n_i \mu_{i0} \tag{A.4}$$

The first component H_{chem}, defined by (A.3), can be regarded as the information lost (or gained) as a result of thermal and/or chemical processes alone. Similarly, H_{conc} is the information gained by concentration (or lost by diffusion) alone.

As it happens, the concentration–diffusion term (A.4) is the easiest to compute, at least for ideal gases. It is also the only term of interest in a few interesting cases. For instance, gold mining and diamond mining are examples of pure physical concentration activities with no chemical transformations involved. This is also true of most metal ore beneficiation processes (e.g., by flotation), although beneficiation is generally followed by a chemical transformation (reduction).

The entropy term H_{conc} associated with increased concentration (or unmixing) can be computed by making use of Boltzmann's statistical definition of entropy of diffusion. For an ideal gas, this has been shown to be

$$H_{\text{conc}} = \mathrm{R} \Sigma n_i \ln(X_{ci}/X_{0i}) \tag{A.5}$$

where R is the ideal constant, n_i is the number of moles of the ith molecular species being diffused, X_{ci} is the mole fraction of the ith species in the concentrated state before diffusion and X_{0i} is the mole fraction of the ith species in the diffused state. [Note that R = kA, where A = 6.02×10^{23} (Avogadro's number) and k = 3.298×10^{-24} cal/°K (Boltzmann's constant). Hence, R = 1.985 cal/°K per mole or 8.31 joules/°K.] Evidently (A.5) represents the amount of information lost when a concentrated material, otherwise in equilibrium, is diffused into an environmental sink, such as the oceans or atmosphere. Obviously, the same expression can be interpreted as the information initially embodied in the concentrated material. See *Table A.1* for examples of information embodied by physical concentration and potentially lost by diffusion.

Table A.1. Information added by physical separation (or lost by diffusion), using the ideal gas approximation.

Element or compound	*Concentration or diffusion process*	X_c - *Concentrated mole fraction*	X_0 - *Dispersed mole fraction*	X_c/X_0	$H_{cal}/^\circ K$ *per mole*
Oxygen	separation from air	1	0.21	4.75	+3.09
Nitrogen	separation from air	1	0.79	1.266	+0.47
Helium	separation from air	1	5×10^{-6}	2×10^{-5}	+24.24
Helium	separation from natural gas (0.3% by vol.)	1	3×10^{-3}	3.3×10^{-2}	+6.95
CO_2	separation from coke combustion products	1	0.21	4.75	+3.09
CO_2	diffusion to air from coke combustion products	0.21	3×10^{-4}	700	-13.00
NaCl	separation from seawater	1	10^{-2}	10^{2}	+9.13
MgCl	separation from seawater	1	10^{-3}	10^{3}	+13.70
Gold	separation from seawater (.011 ppb by wt)	1	10^{-12}	10^{12}	+54.79
Gold	separation from average ore (7 gm/ton)	1	10^{-6}	10^{6}	+27.39
Gold	dispersion to earth's crust	1	5×10^{-10}	2×10^{9}	-53.59

Source: Calculated by author based on data from US Geological Survey and US Bureau of Mines.

The most important chemical process to consider is combustion. Electric power generation, process heat, and space heat all involve combustion of hydrocarbon fuels in air. So-called carbothermic reduction of metal oxides can also be considered, for purposes of this analysis, as a combustion process, where the oxidant is an oxide ore rather than pure oxygen. [As it happens, heat of combustion is approximately given by $-\Delta H_c = a\ n\ R(T_c - T_0)$ where a is a constant and T_c is the combustion temperature.]

The heat of combustion is normally denoted as ΔH_c. It is tabulated in standard references, such as the *Chemical Engineers Handbook* (Perry and Chilton, 1973). It is defined as the sum

$$-\Delta H_c = \Delta U + P_0 \Delta V \tag{A.6}$$

which takes care of the first two terms on the right-hand side of equation (A.3).

Note that $-\Delta H_c$ must *not* be confused with H, which is our measure of information. Most combustion and reduction processes (except in internal combustion engines) occur at constant (atmospheric) pressure, so P_0 is fixed. In cases when fuel and oxidant are both initially in vapor form there is no volume change ($\Delta V = 0$), since n moles of inputs are converted into exactly n moles of outputs (assuming H_2O in the combustion products to be in vapor form). It may be noted that, in general, if the volume of reaction products (at standard temperature and pressure) is smaller than the volume of inputs (fuel plus oxidant), the atmosphere does work on the system. This can be "bootlegged", in principle, and becomes part of the available work or essergy. On the other hand, if the volume of combustion products exceeds the volume of inputs (as when either fuel or oxidant is a liquid or solid), the system does work on the atmosphere that *cannot* be recovered, and which must therefore be subtracted from the available work. In practice, these corrections are fairly small.

In any case, the first two terms on the right-hand side of (A.3) are now taken care of. It remains to compute the entropy change of the system ΔS due to an irreversible heating process. This can be broken down into two components, as follows:

$$\Delta S = \Delta S_1 + \Delta S_2 \tag{A.7}$$

where ΔS_1 results from the change in composition from reactants to products (mixing and virtual dispersion of reactants; virtual concentration of reaction products) while ΔS_2 results from isobaric heating of the reaction products in the flame. The first term ΔS_1 is, in general, positive but quite small (~0.5% assuming H_2O as vapor), arising mainly from the initial mixing of fuel and air. The basic formula for entropy of mixing already has been given in (A.5) above and only the numerical values remain to be determined. The second term in (A.7) is quantitatively more important in almost all cases. To calculate the entropy of isobaric heating (ΔS_2), one may recall the standard textbook definition of entropy change in terms of heat flow:

$$\Delta S_2 = \int dQ/T \tag{A.8}$$

where the integration is carried out from T_0 to the combustion temperature T_c.

But, subject to very general conditions one can express Q in terms of T, P, viz.

$$\begin{aligned} dQ &= (\delta Q/\delta T)_{\mathrm{p}} dT + (\delta Q/\delta \mathrm{P})_T d\mathrm{P} \\ &= C_{\mathrm{p}} dT + \Lambda_{\mathrm{p}} d\mathrm{P} \end{aligned} \tag{A.9}$$

where C_{p} is defined as the heat capacity at constant pressure (a function of temperature only) and Λ_p is the so-called latent heat of change of pressure. However for combustion processes in the atmosphere, pressure is constant. Hence $dP = 0$ and the second term of (A.9) can be dropped. Also, as a reasonable first approximation, one can assume a constant, average value C_p over the temperature range in question. Thus

$$\Delta S_2 = nC_p \int_{T_0}^{T_c} dT/T = nC_p \ln(T_c/T_0) \tag{A.10}$$

where T_c is the combustion temperature, n is the number of moles of combustion products, and

$$C_p = \mathrm{aR} \tag{A.11}$$

where R is the ideal gas constant.

One can derive an approximate expression for T_c, viz.

$$T_c = T_0(1+\mathrm{h}/n\mathrm{a}) \tag{A.12}$$

where h and a are measured constants.

As a grand general approximation for hydrocarbon fuels: h = 19.5M, a = 4.8, n = 0.6M, and h/na $\simeq$ 6.77, where M is the molecular weight.

A useful general approximation for the loss of available useful work (or essergy) by combustion, not including any contribution from the diffusion of combustion products into the atmosphere, follows directly from (A.10), (A.11), and (A.12), viz.

$$B_c = B[1 - (n\mathrm{a}/\mathrm{h})\ln(1+\mathrm{h}/n\mathrm{a})] = 0.7B \tag{A.13}$$

In words, even if we make no allowance for the diffusion of combustion products into the atmosphere, about 30% of the initially available chemical energy B becomes unavailable in any combustion process simply as a result of the entropic loss due to isobaric heating of combustion products. The diffusion process could theoretically yield some available work, but it is not recoverable in practice and can be therefore ignored.

It follows from (A.2) that the thermodynamic information *embodied* in metals and products, such as refined petroleum products, paper, aluminum, and cement, is proportional to B_{min}, the minimum amount of available work needed in theory to produce them. For metals normally found in oxides, the analysis is fairly straightforward. For metals found in nature as sulfides, the situation is complicated by the fact that the sulfide form is first converted to an oxide and the sulfur is driven off as SO_2. In principle, the sulfur could be burned as fuel to recover available work, but this is not done. The "purified" metals actually embody less information than the sulfide minerals. Data for other industrial processes are shown in *Table A.2*. In the case of petroleum and paper, the figures given in the last column do *not* include the heat of combustion of the organic materials *per se*, which is also ultimately recoverable. However, the figures for steel and aluminum do roughly correspond to the heats of formation of the pure metallic ores from the metal and oxygen. For example, the heat of formation of ferric oxide Fe_2O_3 from pure iron and pure oxygen is –198.5 kcal/gm-mole at 25°C. This translates to 4.46×10^6 BTU/ton of Fe_2O_3 or 6.4×10^6 BTU/ton of iron. It is the heat that would be released (as heat of combustion) if the pure iron were burned – as a fuel – in air.

Table A.2. Comparison of specific fuel consumption (BTU/ton $\times 10^6$) of known processes with theoretical minimum for selected US industries.

Process	*Actual 1986 specific fuel consumption*	*Potential fuel consumption using 1973 technology*	*Theoretical minimum specific fuel consumption B_{min} based on thermodynamic availability analysis*
Iron and steel	26.5	17.2	6.0
Petroleum refining	4.4	3.3	0.4
Paper	39.0[a]	23.8[a]	> –0.2[b] < +0.1
Primary aluminium production[c]	190	152	25.2
Cement	7.9	4.7	8.8

[a]Includes 14.5×10^6 BTU/ton of paper produced from waste products consumed as fuel by paper industry.
[b]Negative value means that no fuel is required.
[c]Does not include effect of scrap recycling.
Source: Gyftopoulos *et al.* (1974); see also Hall *et al.* (1975).

APPENDIX 1.B

Information of Orientation and Assembly

The composition of complex shapes from simpler shape components (including holes) was discussed briefly at the end of Section 1.3. This notion requires further elucidation. In the assembly (composition) process, each component must be correctly positioned (that is, oriented *with respect to all others*). For convenience, one may assume that one shape is fixed – the core shape – and another is brought to it, just as a component is added to an assembly. The information embodied in a specified combination of the two shapes is the sum of the information embodied in each shape separately plus a relative locational and an orientational component that depends on the symmetry of each component. There may also be a loss of information if an interior surface is permanently eliminated by joining.

Recalling that complex geometrical shapes can always be constructed by superposing simpler ones (both positive and negative), a straightforward decomposition strategy can be adopted:

(1) Identify and classify the core shape, as distinct from any minor exterior projections or interior spaces, holes, slots, or grooves. In a few cases, there can be some ambiguity. For instance, is a gear tooth an exterior projection or is the space between two gear teeth a slot? However, most of these special cases are included under category 7 in *Table 1.1* and can be treated separately.
(2) Identify the core shapes of exterior projections, such as knobs, heads, etc.
(3) Identify the core shapes of interior spaces within the main core shape.
(4) Identify the core shapes of exterior projections within interior spaces, if any.
(5) Identify the core shapes of interior spaces within external projections, if any. Etc., etc.

This decomposition scheme is applicable to either 2-D or 3-D shapes. It can be continued indefinitely, although three or four steps will suffice for almost any part used industrially; and if more than five steps are required, the design is probably unnecessarily complex. The core shapes, in turn, can be decomposed into surfaces, as discussed above. When the decomposition is complete, the vast majority of the surfaces will be either flat, cylindrical, conic, or spherical. A few (but important) surfaces will have more complex profiles, such as involute curves, parabolas, ellipses, etc.

In the most general case, one may suppose that the major (core) piece is fixed in space and an asymmetric projection or hole must be located and oriented exactly with respect to a coordinate system embodied in it. This requires six degrees of freedom and corresponds to 60 bits of information, as noted above. Symmetries can reduce the orientation information requirement, however. This is clear from the observation that a nonsymmetric shape looks different, depending on its orientation, whereas a symmetrical shape looks exactly the same in two or more different orientations. This is important when a symmetrical shape is added to (or subtracted from) an oriented core shape: *the more symmetrical the second shape is, the more equivalent ways there are to bring them together and the less information is needed to specify the superposition operation.* It is convenient to define a symmetry correction term H_{sym}, which has to be *subtracted* from the total information embodied in the combined shape. A few illustrative values of H_{sym} are listed in *Table B.1.*

Table B.1. Symmetry information embodied in simple shapes.

Shape	*Number of equivalent directional orientations*	H_{sym} (bits)
Rectangle (2-D)	2	$\log_2 2 = 1$
Equilateral triangle (2-D)	3	$\log_2 3 = 1.585$
Square (2-D)	4	$\log_2 4 = 2$
Equilateral pentagon (2-D)	5	$\log_2 5 = 2.322$
Equilateral tetrahedron (3-D)	12	$\log_2 12 = 3.585$
Rectangular parallelopiped (3-D)	12	$\log_2 12 = 3.585$
Cube (3-D)	24	$\log_2 24 = 4.585$
Right cone (3-D)[a]	1000	$\log_2 1000 = 10$
Ellipsoid (3-D)	1000	$\log_2 1000 = 10$
Right cylinder	2×1000	$10 + \log_2 2 = 11$
Sphere (3-D)	$(1000)^3$	$3 \log_2 1000 \simeq 30$

[a]By our convention on standard precision (see text), we assume 1000 distinguishable angular orientations around one axis. If a greater precision is assumed, the number of bits associated with orientation of a cone, ellipsoid, cylinder, or sphere is correspondingly greater.

If the piece to be added or subtracted is rotationally symmetric about one axis, for instance, there is a reduction of 10 bits in the amount of additional orientation information required. In practice, most holes are round, of course.

It is also important to think about the case where the core shape also has symmetry. On consideration, it is clear that two situations are conceptually equivalent: either piece could be designated as the core. Again, the symmetry information must be subtracted. If both pieces are symmetric, however, it does not follow that there must be a subtraction for each. On the contrary, whichever piece is selected as the core can be oriented arbitrarily, by assumption, and its symmetry (if any) does not reduce the information required for superposition.

When physical shapes are permanently joined, the attached surfaces disappear and become indistinguishable. This results in lost information. To take a simple example, suppose we join two right cylinders of the same diameter together, end to end, to form a longer cylinder. The two joined plane surfaces simply disappear as far as the final shape is concerned. The information embodied in the two adjoining surfaces is lost.

Moreover, the two cylindrical surfaces also merge and become one. Suppose the two original right cylinders each embodied $40 + 30 - 50 = 20$ bits of information in their shapes, the final merged cylinder also embodies 20 bits, even though an additional $50 - 2 = 48$ bits information is subtracted from the orientational information. [Since each cylinder is symmetric end to end, there are actually four possible ways of joining them. Selection of any one of these four possibilities creates $\log_2 4 = 2$ bits of information, which (in principle) are required to position and orient the two original cylinders exactly end to end.] Thus, $20 + 50 - 2 = 68$ bits of information are actually lost in this particular assembly process. (Notice that the more symmetric the parts being joined, the less information is lost in superposition or assembly.)

A different situation arises when one disk has greater diameter than the other. In this case, only one of the plane surfaces – the side of the smaller disk – completely disappears. The surface with the larger area does not disappear, although it is partly covered up, and only the information associated with flat end of the smaller disk is actually lost.

The shape in *Figure B.1* is useful to make several observations. Given that it belongs to category 5 in *Table 1.1* it can be defined as the intersection of four parallel planes at right angles with two concentric infinite cylinders. The first plane is specified by three parameters (30 bits) and the other three planes are specified by one parameter each or 30 bits in all, making a total of 60 bits. The first infinite cylinder orthogonal to the plane embodies 40 bits and the second (concentric) cylinder adds 10 bits of information. Thus, the complete shape embodies $60 + 50 - 60 = 50$ bits of information (there are two radii and three lengths).

The shape can be decomposed in two ways: (1) as a superposition of three cylinders, merged end to end or (2) as a superposition of two cylinders, one of which fits inside an axial hole in the other. In the first case, the three simple right cylinders each embody $40 + 40 - 60 = 20$ bits of information or 60 in all. After the end-to-end assembly [requiring $(2 \times 50) - 3$ bits of positional and orientational information], one is left with a final shape embodying only 50 bits, – as noted above, which means $60 + 97 - 50 = 107$ bits of information are lost in the assembly process. Now consider the other starting point: two cylinders, one hollow. The hollow cylinder embodies $40 + 50 - 60 = 30$ bits, while the solid cylinder embodies 20 bits, or 50 bits altogether. After combination, the resulting shape still embodies 50 bits as noted and $50 - 1 = 49$ bits of positional and orientational information are required to bring the two cylinders together. In this case, the information loss is 49 bits, less than half of loss in the first case.

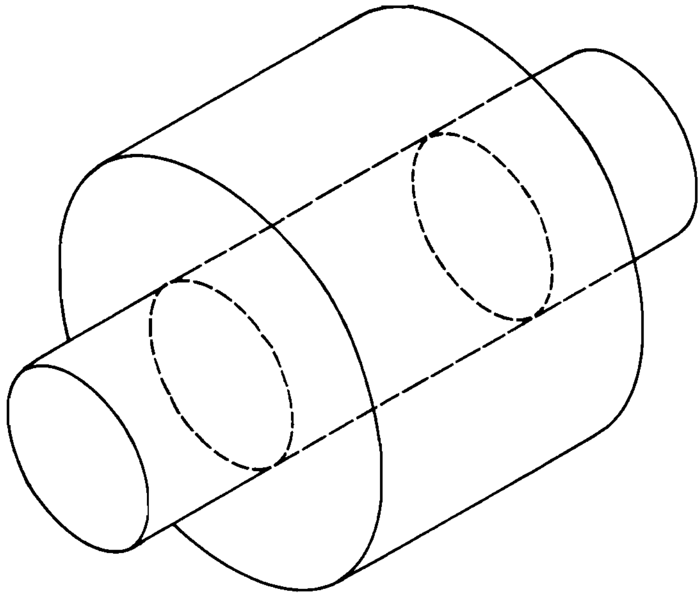

Figure B.1. Superposition of cylinders.

The above example illustrates two key points:

(1) There may be several ways to decompose a complex shape into simpler ones and, conversely, several ways of composing (i.e., assembling) a complex shape from simpler ones. An ideal decomposition is one such that the information embodied in the components is equal to the information embodied in the final shape.

(2) The information lost in the composition (i.e., assembly) process depends strongly on the number of components. The fewer distinct components required, the less information will be lost.

The preceding discussion suggests a fundamental efficiency measure: the ratio of information embodied in the final shape *per se* to information embodied in the component shapes plus information required for positioning and orientation. The information efficiency of constructing the shape in *Figure B.1* from three cylinders is evidently $50/157 \simeq 0.32$, while the information efficiency of constructing it from two cylinders (one hollow) is $50/99 \simeq 0.50$.

2. Human Labor as an Information Process

2.1. Embodied versus disembodied information

In the first half of this report, I focused attention on quantitative computations of the information associated with the composition and microstructure of materials and information added to materials by manufacturing processes. The primary purpose of that section was to draw attention to possible strategies for evaluating and comparing alternative production processes and product designs.

The use of the word "information" hereafter is not intended as a synonym for knowledge, but as the term is used by Shannon and others in information theory (Shannon and Weaver, 1949). A more complete discussion is presented in an earlier paper (Ayres 1987). The Shannonian interpretation of information as uncertainty reduction obviously applies to communication of messages (Shannon's original concern) and can easily be extended to many other kinds of disembodied information.

This section deals primarily with information inputs to, and outputs of, human labor. Labor outputs are, in the first instance, *motions*, but this paper is concerned specifically with motions involved in manufacturing processes. A simple workplace optimization model is then presented, based on ergonomic concepts.

2.2. Ergonomic background: The worker as information processor

For our purpose here, it is appropriate to consider workers as information processors who convert information inputs (mainly sensory data) to some kind of output form (generally, decisions). Output rates of human workers differ widely in various situations. The information *output* of a worker in a manufacturing environment is, by definition, his/her information *input* to the process itself, but it is not necessarily the amount of morphological information embodied in the workpiece being processed. The latter would have to be considered independently, as discussed in Part 1, above. In general, output is much smaller than input. In effect, there is information wasted in every process, as schematically indicated in *Figure 2.1*.

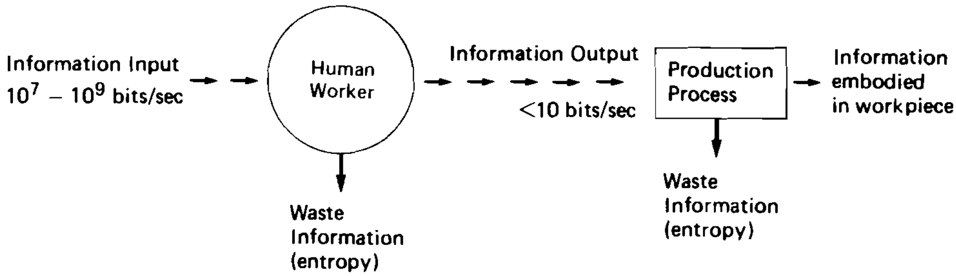

Figure 2.1. Inputs are variously estimated. The information *input* rate to human eyes has been estimated by, e.g., McCormick (1970).

Information inputs to humans in all situations – not only the workplace – are initially conveyed by the five senses: vision (eyes), hearing (ears), touch, taste, and smell. It is known from a variety of evidence, including the amount of brain tissue devoted to each type of sensory signal processing, that, for humans, vision accounts for as much as 90% of total information inputs, while touch and hearing account for most of the remainder (see *Table 2.1*). This does not imply that vision is the only sense that must be considered, however. On the contrary, there are a number specialized workplace situations where audio or tactile information is at least as important as vision – for instance, inserting a nut on a bolt, or tightening it – and there are some industrial tasks where other senses are critical. Many assembly operations are difficult or impossible to accomplish without tactile feedback. However, the very fact that humans cannot "turn off" their senses (except vision) means that significant amounts of waste input information are inevitable in most cases.

Estimates of the (input) transmission capacity of the human eye vary considerably, depending on the measurement technique used. For instance, combining data on monocular visual acuity and so-called "flicker fusion" frequency leads to an estimate of 4.3×10^6 bits/sec (Jacobson, 1951). This implies an estimate of 4.3×10^6 bits/sec for each optic nerve fiber. However, other data have suggested greater input capacities, e.g., 5×10^7 bits/sec (Marko, 1967). Visual input capacity is evidently a function of luminance (light intensity) and may be as high as 109 bits/sec at high luminance levels (Kelley, 1968). The quantity of information that can be extracted from visual or audio inputs (i.e., channel capacity) can be estimated in terms of human ability to discriminate among stimuli and respond suitably. Response normally involves a physical action. The discrimination ability varies according to the nature of the sensory input, as shown in *Table 2.2*, which assumes no time limitations.

The speed with which a single cell can detect and respond to a stimulus is limited by the rate at which certain chemical reactions take place and chemical signal carriers diffuse physically through the cell. This translates into a maximum rate at which a cell can process information. For most single cells, the maximum "output rate" is about 4,000 bits/sec.

Table 2.1. Sensory inputs (bits/sec).

Sense	*Number of receptors*	*Number of nerve fibers*
Vision	2×10^8 (retina)	2×10^6 (optic nerve)
Hearing	3×10^4	2×10^4
Touch	5×10^5	1×10^4
Smell	1×10^7	2×10^4
Taste	1×10^7	2×10^3
Pain	3×10^6	
Heat	1×10^4	1×10^6
Cold	1×10^5	

Source: Marko (1967).

Table 2.2. Information in sensory stimuli.

Sensory modality and stimulus	*W = no. of levels that can be discriminated on absolute basis*	*$\ln W$ = no. of information bits transmitted*[a] *= H*
Vision, single dimensions:		
Pointer position on linear scale	9	3.1
Short exposure	10	3.2
Long exposure	15	3.9
Visual size	5–7	2.3–2.8
Hue	9	3.1
Vision, combined dimensions:		
Size, brightness, and hue	17	4.1
Hue and saturation	11–15	3.5–3.9
Audition, single dimensions:		
Pure tones	5	2.3
Loudness	4–5	1.7–2.3
Audition, combined dimensions:		
Six variables	150	7.2
Odor, single dimension	4	2.0
Odor, combined dimensions:		
Kind, intensity, and number	16	4.0
Taste:		
Saltiness	4	1.9
Sweetness	3	1.7

[a]Amount of information in absolute evaluations on different stimulus dimensions (McCormick, 1970).

In the case of an organ, consisting of a function structure with several kinds of different cells, stimuli are initially received by one type of cell (and not, in general, simultaneously). As the cells individually respond, by producing a chemical signal carrier, there is a gradual buildup of the concentration of that chemical until some threshold level is reached at which point a "response" can be defined for the organ as a whole. Animal organs differ in sensitivity and response rate. Eyes and ears are obviously specialized for large-capacity and rapid response; but the effective range seems to be 340–460 bits/sec, depending on the nature of the input stimulus and the output response. Complete organisms (e.g., animals) can respond much more slowly (10–15 bits/sec), while the maximum response rate for human groups or organizations is still lower (3–4.5 bits/sec). Whether the information processor is a cell, an organ, or an organism, there is a tendency toward overload if the input rate is too high. Successive stimuli begin to interfere with each other and create "noise". [For a detailed theoretical treatment see Sheridan (1982) and Sheridan and Ferrell (1974)]. The difficulty is, in brief, that the detector, whether it is a single molecule or an organ, necessarily changes its state in the process of detection. After responding to one impulse, it is not ready to receive another until returning to the original "ready" state. In the case of humans, the latency period ranges from 0.1 to 0.2 seconds (Miller, 1978).

Within a single cell, the processes of detection and response are essentially indistinguishable, because the internal processes are all chemical reactions; but for an organism, there is a sequence of internal stimulus–response stages culminating in a physical motion. Again, there is an inherent latency period characteristic of each step in the sequence. However, the longest latency period is probably associated with initiating a muscular contraction. Each muscle involved in a motion must relax and return to its original condition and position before it can repeat that motion. The latency period for hand motions is typically about 0.3 seconds. Taking into account relaxation or latency times, the characteristic relationship between information outputs (responses) and inputs is linear for low input rates, reaching a maximum (as noted) and declining thereafter as interference and overload problems occur.

2.3. Examples: Theory and experiment

Typing offers a practical case in point, where the information output computations are comparatively straightforward. A very good typist can sustain a rate of more than 100 words per minute or around 8 characters per second for short periods, or 5–6 characters/second for extended periods. [Typing in a realistic environment is a far more complex activity than this section might seem to imply. See, for example, Salthouse (1984).] If the typist were selecting each character from N equally probable possibilities, the information content of each character in a 26-letter alphabet (plus a space) would be $1n_2 27$ or 4.75 bits. However, taking into account unequal probabilities of occurrence of various letters alone, the information content drops to 4.03 bits. Taking into account preferential combinations of 2 letters, 3 letters, and n letters brings the figure down to less than 3.3 bits. Further taking into account the fact that language consists of words and sentences with unequal word frequencies and subject to grammatical rules adds significantly to the redundancy. Shannon (1951) has calculated some of these factors exactly from published frequency tables and estimated others. Shannon concludes that the information content of English is about 1.5 bits/character. [See also Cherry (1978).] This implies that a very expert typist (100 wpm) can achieve a maximum short-period output rate of 12 bits/sec and a sustained output of 7.5–9 bits/sec.

Direct measurements (Quastler and Wulff, 1955), based on various experimental arrangements, imply maximum channel output capacities, or output rates, of subjects doing various activities for short periods as follows:

Typing random sequences of letters	9.6–14.5 bits/sec
Piano playing	22 bits/sec
Reading	24 bits/sec
Mental arithmetic	24 bits/sec

These numbers are in reasonable agreement with the previous calculation. An important point emerging from the experimental research on keyboard operation is that error rates tend to be constant – but not zero – at speeds up to about 3.2 keys/sec, at which point channel capacity reaches 9.6 bits/sec for 8 keys, 12.8 bits/sec for 16 keys, and 14.5 bits/sec for 32 keys. At slightly faster speeds (up to 4 keys/sec or so), errors increase in proportion to speed, keeping output constant. Beyond this point, errors increase nonlinearly and effective output drops quite sharply, as shown in *Figure 2.2.* Some implications of this data are discussed later in connection with the error–defect problem.

Experiments by others have led to comparable results. For instance, experiments requiring subjects to respond verbally to visual stimuli gave output rates up to 30 bits/sec (Licklider *et al.* 1954). When the required response was to point a finger at a target, the maximum rate dropped to 15 bits/sec. But when verbal and manual responses were permitted simultaneously, the maximum output rate actually increased to 45 bits/sec – the sum of the individual rates. This (and other evidence) suggests that several parallel independent processing channels exist in the human brain.

2.4. Output of a worker: Time and motion

The question now arises: how can the information output of a worker be estimated for typical tasks involving arm (or other body) movements? F.W. Taylor (1911) argued that each particular industrial task can be optimized from a time–motion point of view, so that it is accomplished with minimum effort. In Taylor's words:

> For each job there is the quickest time in which it can be done by a first-class man. This time may be called the quickest time or the *standard* time for the job.
>
> The best way . . . to determine how much work a first-class man can do in a day . . . is to divide the man's work into its elements and time each element separately.

Harold and Lillian Gilbreth subsequently pioneered the development of micromotion study and postulated that elementary motions could be timed as multiples of a fundamental time unit – the "therblig".

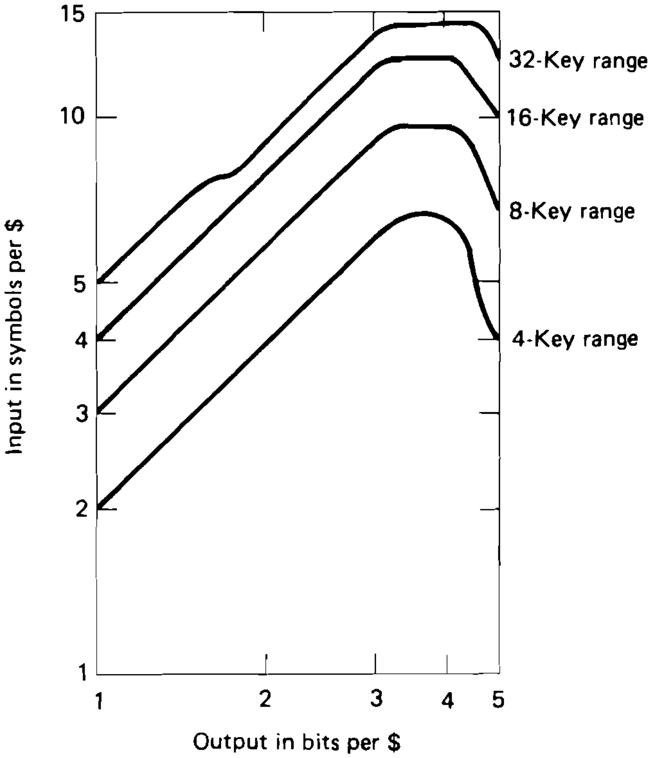

Figure 2.2. Typing performance curves. Source: Quastler and Wulff (1955).

The basic empirical "law of fundamental times", attributed to Segur (1946), states that "within reasonable limits, the time required by experts to perform a fundamental motion is a constant" (Steffy, 1971). If each task can be broken down into fundamental motions, and each motion requires a certain minimum amount of time, then it follows that each well-defined task also requires a minimum amount of time that can be determined from the motions themselves.

A number of methods of subdividing tasks into work elements have been proposed:

MTA:	Methods–Time Analysis	Segur (1925)
WF:	Work Factor	Quick *et al.* (1938)
	Engstrom System	Engstrom and Engstrom (1940)
	400 System	Western Electric (1944)
MTM:	Methods–Time Measurement	Maynard *et al.* (1948)
MTS:	Methods–Time Standards	General Electric (1950)
BMT:	Basic Motion Time Study	J.D. Woods and Gordon Ltd. (1951)
DMT:	Dimension Motion Time	General Electric (1954)

Industrial ergonomists have identified at least 10 elementary arm, hand, and eye motions, each with several subcases depending on situation variables. One well-known list, known as the Methods–Time Measurement or MTM system [the standard references for which are Antis *et al.* (1979) and Maynard *et al.* (1948)], is as follows:

Motion	*Abbreviation*	*Cases*
Reach (empty hand), gross arm	R	A through E, depending on exactness of location & size of object. Function of distance
Move (loaded hand), gross arm	M	A, B, C, depending on exactness of location. Function of distance
Turn (wrist)	T	Functions of angle
Apply pressure (thumb or wrist)	AP	A, B, depending on whether pressure is released
Grasp	G	A, B, C, depending on size of object
Position (orient)	P	Function of symmetry, fit
Release	RL	
Disengage	D	Function of fit, recoil, etc.
Eye travel	ET	
Eye focus	FF	

Any standard manufacturing task can be decomposed into a series of such elementary motions, given information on the location and orientation of workpieces, tools, parts, etc. An example of such a decomposition is shown in *Figure 2.3*. Each elementary motion requires a characteristic length of time for humans, tabulated in standard manuals (Antis *et al.*, 1979). For example, *Table 2.3* displays the average time required for various cases of "reach" as a function of distance moved. (All MTM measurements use a standard time unit or TMU corresponding to 10^{-5} hours or .036 sec). Given a task decomposition, such as that shown in *Figure 2.3* (assumed to be optimal), and a set of elementary motion timetables, such as *Table 2.3*, the "ideal" time for the task can be computed easily.

2.5. Output of a worker: Time and information

One can subdivide each physical (i.e., manipulative) task into two kinds of mental activity. Each requires a characteristic increment of time to:

(1) Decide on a course of action by evaluating information inputs, e.g., scanning and interpreting instrument dials, etc.
(2) Control arm and hand motions.

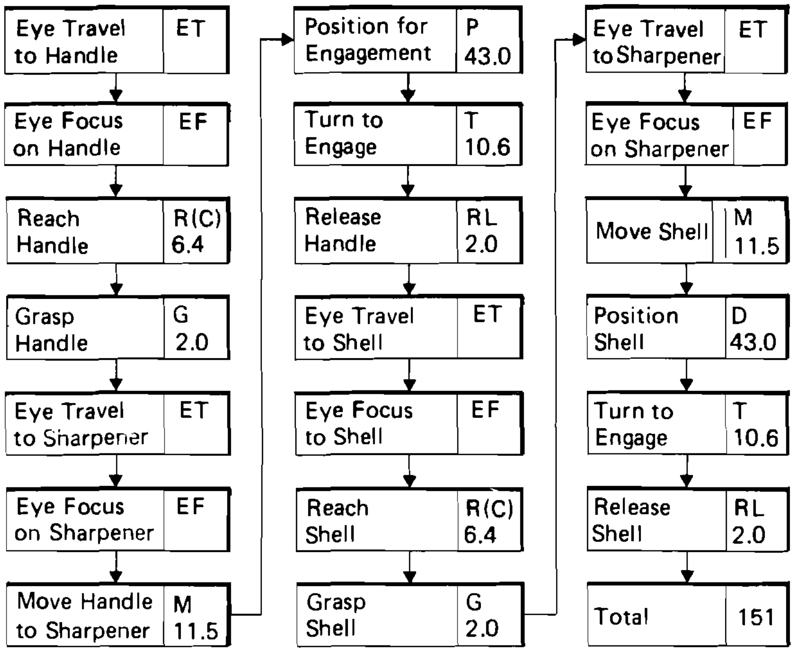

Figure 2.3. Pencil sharpener assembly (1 hand, full sight) MTM. Numbers in boxes refer to TMU. Eye motions occur simultaneously with other motions, hence taking no real time. Source: Ayres *et al.* (1984).

Table 2.3. MTM values for various types of "reach".

Distance moved (inches)	*Time in TMU* (1 TMU = .00001 hr = .036 sec)					
	RA[a]	*RB*[b] *or RC*[c]	*RD*[d]	*RE*[e]	*RA*	*RB*
1	2.5	2.5	3.6	2.4	2.3	2.3
4	6.1	6.4	8.4	6.8	4.9	4.3
12	9.6	12.9	14.2	11.8	8.1	10.1
30	17.5	25.8	26.7	22.9	15.3	23.2

[a]Reach to object in fixed location, or to object in other hand or on which other hand rests.
[b]Reach to single object in location that may vary slightly from cycle to cycle.
[c]Reach to object jumbled with other objects in a group so that search and select occur.
[d]Reach to a very small object, or where accurate grasp is required.
[e]Reach to indefinite location, to get hand in position for body balance, or for next motion, or out of the way.
Source: Antis *et al.* (1979).

For the first type of mental activity (decision), the time required is given by Hick's law, viz.

$$T_d = K_p + H_d / C_d \tag{2.1}$$

where K_p is the minimum delay time associated with sensory perception, C_d is the effective perceptual response channel capacity and H_d is the amount of information output (decisions) in bits (Hick, 1952). Depending on the sensory mode: K_p = 0.15–.225 sec for visually presented information, 0.12–0.18 sec for auditory information, and 0.115–0.19 sec for tactile information (Salvendy and Knight, 1982). Hick estimated C_d at 4.5 bits/sec, which is well below the maximum channel capacity figure suggested by others.

The time required for physical motion *per se* is analogously given by

$$T_m = K_m + H_m / C_m \tag{2.2}$$

where K_m is the minimum delay time associated with initiating a motion (0.177 sec for hand motion), C_m is the channel capacity for motion control, and the amount of control information required is

$$H_m = \log_2(2A / W) \tag{2.3}$$

Here A is the amplitude of motion, and W is the target width (Fitts, 1954).

The motion control channel capacity C_m is assumed by Salvendy and Knight (1982) to be 5 bits/sec, even though (as noted previously) expert typists can reach twice this rate. However, an alternative estimate is derived as follows. It is experimentally observed that most arm movements occur in two stages:

(1) An open-loop gross ballistic motion, which moves the arm into the vicinity of the target, with about 93% accuracy (7% error in amplitude), based on information stored in memory.
(2) A series of closed-loop, visually controlled correction movements, if higher precision is required. Each such correction requires about 0.3 sec and reduces location error by 93%.

Thus, for a single-stage (open-loop) gross ballistic motion, setting $2A / W = (.07)^{-1}$ = whence

$$H_m = \log_2 14.3 = 3.84 \text{ bits} \tag{2.4}$$

Assuming the parametric values given above (C_m = 5 bits/sec)

$$T_m = 0.177 + 0.768 = 0.945 \text{ sec} \tag{2.5}$$

However, for each successive (closed-loop) correction

$$H_i = 3.84 \text{ bits and } T_i = 0.3 \text{ sec} \tag{2.6}$$

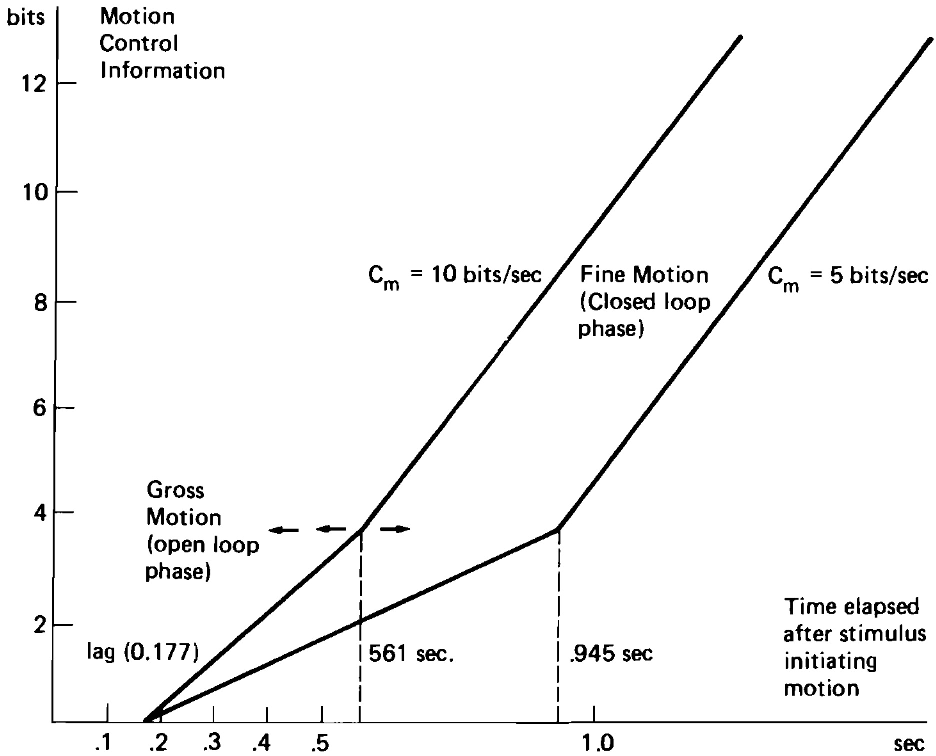

Figure 2.4. Information for motion control versus time.

Plotting the above relationship for two different values of C_m yields the results shown in *Figure 2.4*. It can be seen that the control information channel capacity required for motion - hence, the mental load - increases sharply for the closed-loop search-correction phase, viz.

$$1.84/0.3 = 12.4 \text{ bits/sec} \tag{2.7}$$

This requires intense concentration, at a level that is almost certainly not sustainable over long periods. It is unclear whether open-loop (gross) motion control channel capacity is really less than closed-loop channel capacity or, if it is, why. The reverse would seem more plausible. The experimental data are ambiguous. There are still many unresolved puzzles in this field.

There is evidence from other experiments, too, that the time required for certain tasks is a function of the information input rate. Such tasks are inherently sensory-intensive, if not sense-limited. A particularly direct example comes from Murrel (1965), who shows a relationship between illumination, accuracy, and time (*Figure 2.5*). The figure shows that for fixed levels of illumination (25-ft L, 10-ft L, 5-ft L), accuracy increases with time. If the available time is fixed, on the other hand, accuracy is an increasing function of illumination. Illumination level can be regarded as a rough measure of visual information received by the worker. Interesting results of a similar nature were obtained by McCormick and Niven (1952) in experiments with students. They found that examination scores (accuracy) were significantly improved by increased illumination, although the effect of illumination saturated beyond a certain point. Comparable results have been obtained by Kaswan and Young (1965).

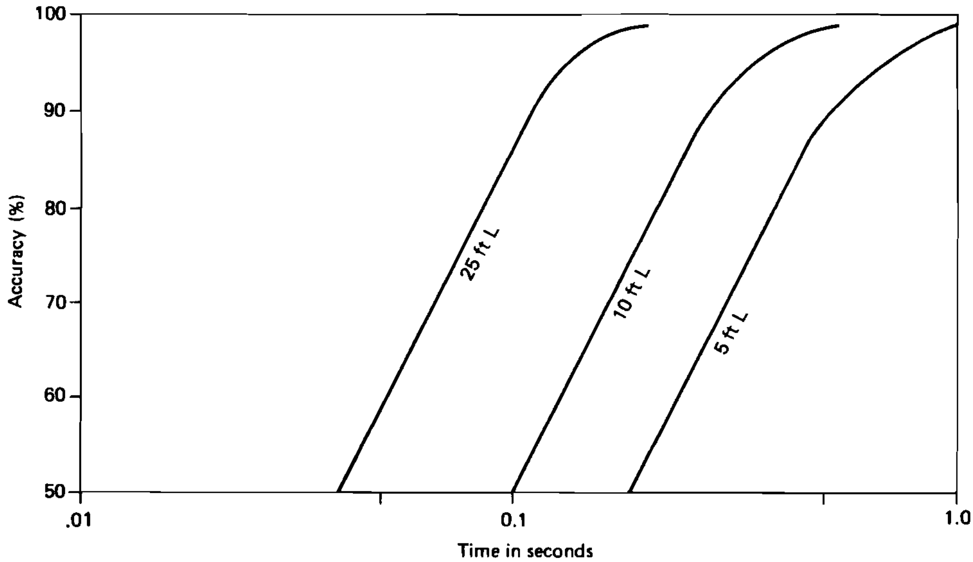

Figure 2.5. A relationship between illumination, accuracy, and time of viewing.

Recent experiments by Ayres *et al.* (1984; also reported in Ayres, 1986) obtained similar results for a series of manual tasks requiring both visual and tactile sensing. Experiments such as these imply that various tasks can be characterized and ranked in terms of their relative sensory-dependence. It is quite plausible that the feasibility of substituting a "smart sensor" for a human worker is essentially inversely proportional to the sensory-dependence of the task, i.e., the degree to which task performance is degraded by sensory deprivation.

2.6. Output of a worker: Motion and information

Two things related to the same thing must be related to each other. In the preceding two sections it was argued, first, that any set of human motions requires a minimum amount of time and, second, that information required for motion control purposes can be estimated on the basis of the time required. From this it follows that a given sequence of motions requires (at least) a given amount of information output in the form of control signals. On the average, the information output of a manual worker performing a manipulative task is essentially proportional to the time required for the motions.

Not all motions are productive, of course, just as not all information is useful. The original purpose of Taylor's time-motion analysis (1911) was to compare alternative *methods* (i.e., sequences of elementary motions) of performing a given task, so as to select the best one. When the optimal method has been found, workplace layout can be optimized in its turn. Finally, operators can then be trained systematically. It is estimated that 80% of the benefits of time-motion study are attributable to task sequence optimization and worker training, while 20% are attributable to better workplace layout (Steffy, 1982).

The cumulative effect of training can be significant. This class of phenomenon is generally known as the learning curve or the "experience curve". It is often used to predict unit cost reductions as a function of cumulative production; see Cunningham (1980). Experience in operating cigar-making machines, for instance, showed that a speed-skill improvement factor of 2.5 was possible, but only after 2 years and 3 million machine cycles of experience (Crossman, 1956, 1961). The speed-skill improvement presumably is due to an increase in the output of *useful* information, rather than in the *total* amount of information output. The gain in useful information can be interpreted as a decrease in the amount of waste motion, including errors.

It was pointed out in Part 1 of this paper that the amount of morphological information actually embodied in a product, such as an assembly, can be computed directly, at least in principle. It can be assumed that if such a task is carefully optimized from both time-motion and workplace layout perspectives, and if the worker is adequately trained, the motion information output of the worker will be roughly proportional to the morphological information ultimately embodied in the product (assembly).

The ergonomic evidence presented up to this point clearly suggests that Taylor's "quickest time" principle of task optimization can be reexpressed in different language. In effect, Taylor implies that the optimal workrate is that which maximizes the output of (specified) motions, where each job is described as a sequence of motions. Relating motion to information processed (as above) further implies that *information output* is thereby maximized by Taylor's principle.

However, this interpretation is misleading, as will be seen. It fails to distinguish clearly between "useful" information outputs and "garbage" (or error) outputs. The experiments cited earlier (e.g., *Figure 2.2* and many similar experiments) show maximum *useful* information output as a function of input rate, not maximum *total* information output (including wrong letters, wrong notes, etc.). Motion *per se* cannot be equated with *useful* information output, only with *total* information output.

As pointed out earlier, the ergonomic evidence clearly implies that there exists a maximum rate of *useful* information output for human workers. This, by itself, is not inconsistent with Taylor's point of departure. However, Taylorism neglects the fact that useful information is often combined with non-useful information (errors or garbage). Seeking to maximize the former makes sense if (and only if) the associated errors or garbage output have no economic importance and can be neglected. [Curiously, Taylor's two most famous examples – the optimization of shoveling (of ore or coal) and wheelbarrow loading – are among the few and rare cases where mistakes (resulting in spillage) do not create a significant new problem with associated costs. In these cases, also, the limits of human performance are more physical than mental.]

2.7. The error–defect problem

Defects in products arise from one of two sources: design flaws and operational errors by workers. The latter predominate, in practice, but both are ultimately attributable to human error. Industrial engineers and ergonomists attack the problem by seeking to understand and eliminate the environmental factors that increase human error propensity or HEP.

Factors that tend to increase the error rate include:

(1) Emotional stress.
(2) Physical strain and discomfort.
(3) Poor illumination.
(4) Information load (overload).

The influence of these factors on human performance and error rate is discussed in a number of ergonomics and human factors studies and textbooks, such as Meister (1971, 1976); Grandjean (1980); and Swain and Guttmann (1983). Most of today's "quality control" methods are predicated on the reduction or elimination of adverse factors.

Statistical quality control, now practiced with outstanding success by the Japanese, was based in large measure on the pioneering work of Shewhart (1980; originally published in 1931), of Bell Laboratories, and systematized by the Western Electric Company. Statistical methods were transferred to the Japanese after World War II by Bell System engineers at the request of US Occupation authorities, who were anxious to get the Japanese telephone system functioning again (Trevor, 1986). The concept of total quality control (TQC) was first introduced in 1967 by A.V. Feigenbaum (1983), of the General Electric

Company. But there is no doubt that these methods have been implemented most successfully by Japanese manufacturers in recent years, and have become a major competitive weapon. There many examples illustrating how Japanese methods have been able to reduce defect rates and reject rates by factors of 10 to 100 from levels considered normal and acceptable in US plants only a decade or so ago (see, e.g., Garvin, 1983).

Nevertheless, organizational and "human factors" approaches to quality must ultimately approach natural limits of a fundamental sort, as pointed out above. Defect rates cannot in practice be reduced to absolute zero in any production system operated by human workers. Occasional mistakes are inevitable because of human variability. Not all mistakes can be detected and, of those discovered, not all can be corrected or adjusted for. Beyond some point, the cost of additional inspection, testing, and rework must exceed the benefits.

When organizational and ergonomic strategies for error–defect control are exhausted, the only option left is to eliminate the remaining source of error from the production system. This source is the human workers themselves, especially those having direct contact with workpieces and production machines. Recognition of this dilemma provides the motivation for introducing computers in place of humans as machine controllers and operatives.

Table 2.4. Imputed error rates for certain manufacturing operations.

Error	*Number of observations*	*Observed number of defects*	*Calculated error probability HEP_i*
Two wires transposed	13,083	22	0.0006
Component omitted	103,880	213	0.00003
Solder joint omitted	47,075	596	0.00005
Operation, such as applying staking, omitted	59,435	11	0.00003
Component wired backward (diodes, capacitors, etc.)	2,610	27	0.001
Wrong value component used	103,880	213	0.0002
Lead left unclipped	33,000	551	0.00003
Component damaged by burn from soldering iron	103,880	213	0.001
Solder splash	47,075	596	0.001
Excess solder	47,075	596	0.0005
Insufficient solder	47,075	596	0.002
Hole in solder	47,075	596	0.07

Source: Rook (1962).

A 1961 study of the sources of 23,000 product defects in electrical assembly plants revealed that 82% were caused directly by worker errors while the remainder could be attributed to "design flaws". The results are shown in *Table 2.4* (Rook, 1962). It must be emphasized that these are *net* error rates computed by examining finished products, and therefore after ordinary inspection, error detection, and rework. The intrinsic human error probability is clearly higher than the net figures shown. More recent work by Swain (1977) and Swain and Gutmann (1983) suggests that intrinsic human error probabilities for various repetitive operations may be around an order of magnitude *higher* than Rook's number, or around 10^{-3} per opportunity. Error rates in some activities, such as computation and number transcription, approach one error in 30 opportunities (McKenney and McFarlan, 1982). In this particular case, at least, electronic computers are now at least 100,000 times more reliable than humans.

It is true that these error rates (HEPs) are not absolute. They are subject to environmental conditions, for instance. But stress, discomfort, and distraction tend to increase the HEP above its normal level. In fact, under life-threatening stress conditions, the HEP can approach 25% (Swain and Gutmann, 1983).

There is also a learning effect, as mentioned previously. Practice does tend to make perfect, although even with practice the error rate does not approach zero in any production situation. Training can increase accuracy as well as efficiency; but, as noted previously, the number of repetitions (cycles) needed to reach maximum accuracy tends to be rather large.

2.8. Errors and information overload

There is compelling empirical evidence that human errors beyond the minimum level are due in large part to "information overload" – inability of the organism to respond to input signals as fast as they are received. Many studies have found the following pattern: As information output (i.e., human responses or actions) is plotted as a function of the rate of information input, the curve rises linearly at low input rates, reaches a maximum at some input rate (depending on the type of input and the type of output), and then drops off as input rates rise further (*Figure 2.6*). The literature has been summarized by Miller (1978), who notes that the same pattern applies to cells (e.g., neurons), sensory organs, organisms, groups, and organizations. In general, the maximum information output level occurs at lower and lower input levels, as one proceeds along this sequence. For human organisms, the maximum output level is of the order of 10^{-15} bits/sec.

As information output rate falls below a simple proportion of input, the information loss (i.e., error rate) rises sharply. This implies a *strongly nonlinear* relationship between output rate and error rate (*Figure 2.7*). It also implies that, to maximize the output rate of any information processing activity, it is necessary to accept a significant nonzero error rate. To reduce the error rate to near zero also necessitates sharply reducing the output rate.

The economic optimum obviously depends on the cost of discovering and correcting errors and defects vis-à-vis the implicit cost of reducing errors by cutting output rates or cycle times. There is good reason to suppose, however, that work rates adopted in most modern factories are based on largely on job descriptions and standards set in the early years of method–time–motion studies during the heyday of Taylor's scientific management (1900–1925). At that time there was very little understanding of fundamental ergonomic relationships. It is likely

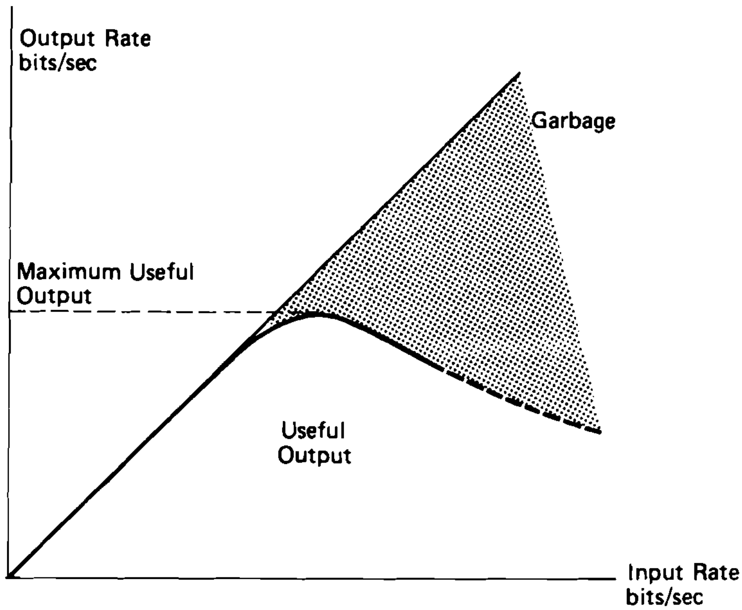

Figure 2.6. Information overload phenomenon.

that, in most cases, rates were adjusted to *maximum* output levels rather than to levels that would maximize the net useful output of the entire production process (i.e., after error detection and correction). Given the increasing mechanical complexity of most products and the increasing cost to manufacturers of uncorrected defects, it would appear that the optimum work rate (if human labor is used) should actually be adjusted downward from levels set early in the twentieth century. The amount of the downward adjustment depends on the complexity of the product because the cost of eliminating defects (or tolerating them) increases sharply with product complexity. A simple economic model of the optimum output rate is given in the next section.

2.9. Optimum work pace

We can assume, for purposes of analysis, that the worker's output response curve for a set of given conditions is a known function of input rate, all other factors remaining constant (see *Figure 2.6*). Define y as useful output rate (in bits/sec) and x as the input rate, also in bits/sec.

For very low rates of input (small x) we can also safely assume that

$$y = kx \quad 0 \leq k \leq 1 \tag{2.8}$$

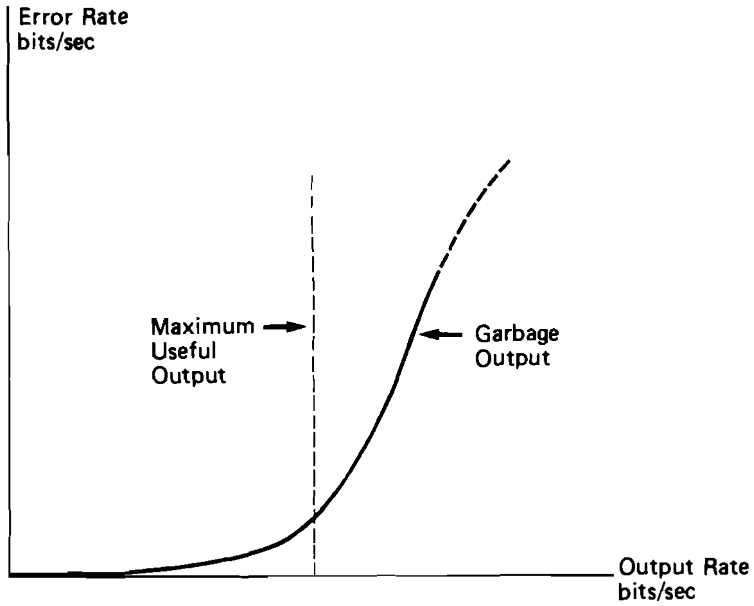

Figure 2.7. Error rate versus output rate.

where the constant k reflects only the technical nature of the task, i.e., the degree of information reduction involved in converting sensory stimuli into useful output (motions).

For higher rates of input $y(x)$ falls below the simple linear relationship. In fact, the conditions are

$$\frac{dy}{dx} > 0 \qquad \frac{d^2y}{dx^2} < 0$$

This characteristic behavior is usually interpreted as an overload phenomenon of some sort. The point of maximum output $\hat{y}$ corresponds to an input rate x_m, which can be found by the condition

$$\frac{dy}{dx} = 0 \tag{2.9}$$

assuming the function $y(x)$ is known.

The difference between potential useful output (if there were no overload or fatigue factor) and realizable output is, in effect, lost information or garbage output. For convenience, define garbage output $g(x)$ as follows

$$g(x) = kx - y(x) \tag{2.10}$$

By previous assumption, at very low input rates, $g(x)$ approaches zero, viz.

$$\lim_{x \to 0} g(x) = 0 \tag{2.11}$$

Rewriting (2.10), we can reinterpret the potential output as *total* output of useful information plus garbage, viz.

$$\underset{\text{total output}}{kx} = \underset{\text{useful output}}{y(x)} + \underset{\text{garbage output}}{g(x)} \tag{2.12}$$

Now suppose the worker is employed by a firm and his/her useful output has an economic value or shadow price P_y per unit. For generality, we can also assign a price P_g per unit to the garbage. Of course P_g is nonpositive, but it may or may not be zero. It can be interpreted as the (shadow) penalty cost of garbage to the firm employing the worker. Garbage, in this context, can be thought of as defective or spoiled output that must (at the very least) be separated and disposed of or, more likely, repaired or reworked.

If the garbage output is simply waste motion of some sort, then P_g can be neglected and simply set equal to zero; but this is not likely or important in a realistic situation. If the object of the work is to add value to a material workpiece, an error on the part of the worker has a finite probability, depending on the case, of introducing a defect to the workpiece. If the defective piece is immediately detected and discarded, the minimum loss is either (1) the value of all work done on the material in earlier stages of production, plus the value of the purchased material, less any salvage value; or (2) the cost of repairing the defect – whichever is less. On top of this must be added the *pro rata* cost of the inspection (defect detection), since inspection would not be required if there were no worker errors and/or resulting defects.

It follows immediately from the above analysis that, in general, P_g is negative, and nonzero:

$$P_g < 0 \tag{2.13}$$

Moreover, the absolute value of P_g is clearly an increasing function of the complexity of the entire production process and of the stage of the process where the defect occurs. This reflects the well-known point that repair and rework are far more expensive than original production because of the high labor intensity involved.

The value added per unit time by the worker can now be written

$$V(x) = P_y y(x) + P_g g(x) \tag{2.14}$$

The optimum workpace x is determined by maximizing $V(x)$, i.e., by satisfying the condition:

$$\frac{dV(x)}{dx} = 0 \tag{2.15}$$

This condition yields

$$\begin{aligned} 0 &= P_y \frac{dy}{dx} + P_g \frac{dg}{dx} \\ &= P_y \frac{dy}{dx} + P_g \left[\mathrm{k} - \frac{dy}{dx} \right] \\ &= (P_y - P_g) \frac{dy}{dx} + \mathrm{k} P_g \end{aligned} \tag{2.16}$$

whence

$$\frac{dy}{dx} = \frac{-k\ P_g}{(P_y - P_g)} \tag{2.17}$$

2.10. Concluding remarks

Clearly, the economic optimum output rate ($dy/dx = 0$) only coincides with the *maximum* output (Taylor) condition in the exceptional case where $P_g = 0$. Moreover, the larger the (negative) values of P_g, the greater the deviation and the lower the optimum value of x. Much of the criticism of crude Taylorism by industrial psychologists would seem to be amply justified in view of the above results. A more interesting and less obvious implication arises from the consideration that P_g is likely to be a function of the precision and performance (or complexity) of the end-product. Thus, for a simple end-product such as a paper clip or a poker chip, the unit value of a "bad" unit is not greater (in absolute value) than that of a "good" one. If 3 units out of a batch of 100 must be discarded, the value of the batch is essentially 97/100 of the potential maximum. Not so if the faulty part is built into a subassembly or a large machine. A faulty ball-bearing installed in a large turbine can cause damage and losses far in excess of the nominal value of a "good" unit. "For want of a nail the shoe was lost, for want of a shoe the horse was lost, for want of a horse the rider was lost, . . ." If the manufacturing task is one step in a sequence leading to a very complex product (such as a space shuttle), the value of P_g can be almost arbitrarily large and negative. In the limit as $P_g \to -\infty$ the optimum solution approaches $dy/dx = \mathrm{k}$, which yields $y(x) = \mathrm{k}x$. From (2.10) this condition corresponds to $g(x) = 0$. From (2.11) this requires $x = 0$, i.e., a work pace of zero!

In simple language, as errors and defects become increasingly costly, the optimum work pace becomes slower. If errors and defects are intolerable, the optimum pace (for human workers) is zero.

The obvious way out of this dilemma is to replace error-prone human workers by (more) reliable computer-controlled machines. To be sure, the foregoing arguments are simplified, but the underlying implication seems to be quite robust.

REFERENCES

Allred, J.C. (1977), Application of entropy concepts to national energy problems, unpublished background paper prepared for a symposium on Entropy and Economics, Houston, TX, October–November.

Antis, W., Honeycutt, J.M., Koch, E.N. (1979), *The Basic Motions of MTM,* Maynard Foundation, Pittsburgh, PA.

Ayres, R.U. (1986), *Complexity, Reliability and Design: The Coming Monolithic Revolution in Manufacturing,* Working Paper WP-86-48, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ayres, U. (1987), Self-organization in biology and economics, in M. Alonso (ed.), *Organization and Complexity*, Paragon House, New York. [NOTE: Forthcoming as an IIASA Research Report, Laxenburg, Austria.]

Ayres, R.U., Cappell, N., and Miller, S. (1984), Potential for substituting robots for humans, *Manufacturing Systems* **13** (1): 26–51.

Boltz, R.W. (1976), Production processes, in *Productivity Handbook,* 5th edition, Industrial Press, New York.

Boothroyd, G. and Dewhurst, P. (1983), *Design for Assembly*, Department of Mechanical Engineering, University of Massachusetts, Amherst, MA, 2nd edition.

Brillouin, L. (1951), Physical entropy and information, *J. of Applied Physics* **22** (3): 338–343.

Brillouin, L. (1953), Negentropy principle of information, *J. of Applied Physics* **24** (9): 1152–1163.

Brillouin, L. (1962), *Science and Information Theory,* 2nd edition, Academic Press, New York.

Burbridge, J.L. (1975), *The Introduction of Group Technology,* John Wiley, New York.

Burwell, C.C. (1983), *Industrial Electrification: Current Trends,* ORAU/IEA(83-4), Institute for Energy Analysis, Oak Ridge, TN, February.

Cherry, C. (1978), *On Human Communications,* 3rd edition, MIT Press, Cambridge MA.

Crossman, E.R.E.W. (1956), The measurement of perceptual load in manual operations (unpublished).

Crossman, E.R.E.W. (1961), A theory of the acquisition of speed–skill, *Ergonomics.*

Cunningham, J.A. (1980), Using the learning curve as a management tool, *IEEE Spectrum,* June.

Devries, M.F., Harvey, S.M., and Tipnis, V.A. (1976), *Group Technology: An Overview and Bibliography,* MDC 76-601, Machinability Data Center, Cincinatti, OH.

Doblin, C. (1985), The impact on energy consumption of changes in the structure of US manufacturing production. Unpublished manuscript. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Edwards, G.A.B. (1971), *Readings in Group Technology,* Machinery Publishing Company, London.

Evans, R.B. (1969), A proof that essergy is the only consistent measure of potential work for chemical systems, Ph.D. thesis, Dartmouth College, Hanover, NH.

Feigenbaum, A.V. (1983), *Total Quality Control,* 3rd edition, American Society of Quality Control. [NOTE: First published in 1967].

Fitts, P.M. (1954), The informational capacity of the human motor system in controlling the amplitude of movements, *Journal of Experimental Psychology* **47**: 381–391.

Gallagher, C.C. and Knight, W.A. (1973), *Group Technology,* Butterworths, London.

Garvin, D.A. (1983), Hard new evidence on American product quality underscores the task ahead for management, *Quality on the Line,* September–October: 65–75.

Grandjean, E. (1980), *Fitting the Task to the Man,* Taylor & Francis, London.

Gyftopoulos, E.P., Lazaridis, L.J., and Widmer, T.F. (1974), *Potential Fuel Effectiveness in Industry* [Series: Ford Foundation Energy Policy Project], Ballinger, Cambridge, MA.

Hall, E.H., Hanna, W.T., *et al.* (1975), *Evaluation of the Theoretical Potential for Energy Conservation in Seven Basic Industries.* PB 244772, Battelle Laboratories, Columbus, OH.

Ham, I. and Ross, D.T. (1977), *Integrated Computer-Aided Technology: Final Report,* U.S. Air Force Technical Report (AFML-TR-77-218), Wright Patterson Air Force Base, Dayton, OH, December.

Hick, W.E. (1952), On the rate of gain of information, *Quarterly Journal of Experimental Psychology* **44**: 11–26.

Jacobson, H. (1951), The informational capacity of the human eye, *Science* **113**: 292.

Kaswan, J. and Young, S. (1965), Effects of illumination exposure duration and task complexity on reaction time, in R. Haber (ed.), *Information Process Approaches to Visual Perception,* Holt, Rinehart & Winston, New York.

Kelley, C.R. (1968), *Manual and Automatic Control,* John Wiley, New York.

Licklider, J.C.R. *et al.* (1954), *Studies in Speech, Hearing and Communication,* Contract Final Report W191-222014, MIT Acoustic Laboratory, Cambridge, MA.

Marko, H. (1967), Information theory and cybernetics, *IEEE Spectrum* **4** (Nov): 75.

Maynard, H.B., Stegmerten, G.J., and Schwab, J.L. (1948), *Methods-Time Measurement,* McGraw-Hill, New York.

McCormick, E.J., (1970), *Human Factors in Engineering,* 3rd edition, McGraw-Hill, New York.

McCormick, E.J. and Niven, J.R. (1952), The effects of varying intensities of illumination upon performance on a motor task, *Applied Psychology* **36** (193).

McKenney, J.L. and McFarlan, F.W. (1982), The information archipelago – maps and bridges, *Harvard Business Review,* September–October.

Meister, D. (1971), *Human Factors: Theory and Practice,* John Wiley, New York.

Meister, D. (1976), *Behavioral Foundation of System Development,* John Wiley, New York.

Miller, J.G. (1978), *Living Systems,* McGraw-Hill, New York.

Mitrafanov, S.P. (1966), *Scientific Principles of Group Technology,* National Lending Library for Science and Technology, United Kingdom (English translation).

Murrel, K.F.H. (1965), *Human Performance in Industry,* Reinhold, New York.

Nakasawa, H. and Suh, N. (1984), Process planning based on information concept, *Robotics & Computer-Integrated Manufacturing* **1**(1): 115–123.

Opitz, H. (1967), *A Classification System to Describe Workpieces,* Pergamon Press, Oxford, UK. [NOTE: Undated, year approx.]

Opitz, H. (1970), *A Classification System to Describe Workpieces* [Series: Parts 1 and 2], Pergamon, London and New York.

Perry, R.H. and Chilton, C.H. (1973), *Chemical Engineer's Handbook,* 5th edition, McGraw-Hill, New York.

Prigogine, I. (1955), *Thermodynamics of Irreversible Processes,* Charles C. Thomas, Springfield, IL.

Prigogine, I., Nicolis, G., and Babloyantz, A. (1972), Thermodynamics of evolution, *Physics Today*, November/December.

Quastler, H. and Wulff, V.J. (1955), *Human Performance in Information,* R-62, Control Systems Laboratory, University of Illinois, Urbana, IL.

Rook, L.W. (1962), *Reduction of Human Error in Industrial Production,* Technical Memo SCTM 93-62(14), Sandia Corporation, Albuquerque, NM.

Salthouse, T. (1984), The skill of typing, *Scientific American* **250**(2), Feb: 128–135.

Salvendy, G. and Knight, J.L. (1982), Psychomotor work capabilities, in: Salvendy, G. (ed.), *Handbook of Industrial Engineering,* Wiley–Interscience, New York.

Segur, A.B. (1946), *Motion Time Analysis Instruction Manual,* A.B. Segur & Co., Oak Park, IL.

Shannon, C.E. (1948), A mathematical theory of communication, *Bell System Technical J.* **27**.

Shannon, C. (1951), Prediction and entropy of printed English, *Bell System Technical J.* **30**: 50–64.

Shannon, E. and Weaver, W. (1949), *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, IL.

Sheridan, T.B. (1982), *Measuring, Modeling and Augmenting Reliability of Man–Machine Systems,* MIT Press, Cambridge, MA.

Sheridan, T.B. and Ferrell, W.R. (1974), *Man–Machine Systems,* MIT Press, Cambridge, MA.

Sherwood, (1978), *New Worlds in Chemistry,* rev. ed., Basic Books, New York.

Shewhart, W.A. (1980), *Economic Control of Quality in Manufactured Products*, American Society of Quality Control. [NOTE: originally published in 1931.]

Steffy, W. (1971), Use of predetermined time standards, in H.B. Maynard (ed.), *Handbook of Industrial Engineering*, 3rd edition, McGraw-Hill, New York.

Suh, N.P., Bell, A.C., Wilson, D.R. and Rinderle, J.R. (1978), On an axiomatic approach to manufacturing systems, *J. of the Engineering Industry (Transactions ASME)* **100**: 127–130.

Swain, A.D. (1977), *Design Techniques for Improving Human Performance in Production,* published by the author, 712 Sundown Place, SE, Albuquerque, NM 87107.

Swain, A.D. and Guttman, H.E. (1983), *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Application,* Technical Report NUREG/CR-1278-F SAND-0200, Sandia Corporation, Albuquerque, NM.

Taylor, F.W. (1911), *Shop Management,* Harper & Row, New York.

Trevor, M. (1986), Quality control – learning from the Japanese, *Long-Range Planning* **19**(5): 46–53.

Tribus, M. (1961a), Information theory as the basis for thermostatics and thermodynamics, *J. of Applied Mechanics* **28** (March): 1–8.

Tribus, M. (1961b), *Thermostatics and Thermodynamics,* Van Nostrand, New York.

Western Electric (1949), *Elemental Time Standards for Basic Manual Work*, Western Electric, New York.