HARVARD | BUSINESS | SCHOOL

To: EO seminar participants

From: Carliss Baldwin and Kim Clark

Re: Readings for the seminar 4/4/02

We believe in the power of modularity, and thus we have organized the readings for the seminar in a modular fashion. This PDF file contains three short papers. They are bookmarked for your convenience, and you can read them in any sequence you like.

The papers are:

"The Value and Consequences of Modularity in Design." (13 pp.) This paper summarizes the main arguments of our book *Design Rules, Volume 1, The Power of Modularity*.

"The Fundamental Theorem of Design Economics." (6 pp.) This paper establishes the scope of our basic valuation approach.

"Institutional Forms, Part 1: The Technology of Design and its Problems." (28 pp.) This paper describes some of the issues of institutional design that we plan to explore in *Design Rules, Volume 2*. It is new work, and we think it is relevant to the economics of organizations.

We look forward to seeing you on April 4!

SOLDIERS FIELD | BOSTON, MA | PH 617.495.6673 | FX 617.496.4191 | 'CBALDWIN@HBS.EDU'| GEORGE F. BAKER FOUNDATION

Comments Welcome

The Value and Consequences of Modularity in Design

A Summary of the Argument in
Design Rules, Volume 1

Carliss Y. Baldwin
Kim B. Clark

Harvard Business School

This Draft:
March 14, 2002

Our thanks to Masahiko Aoki, Wayne Collier, Sonali Shah, David Sharman, Kevin Sullivan, Jason Woodard, and members of the Negotiations, Organizations and Markets group at Harvard Business School for sharing key insights. We thank especially Barbara Feinberg for the key role that she played in helping us to refine and strengthen our arguments. We alone are responsible for errors, oversights and faulty reasoning.

Direct correspondence to:
Carliss Y. Baldwin
Harvard Business School
cbaldwin@hbs.edu

1 An Overview of the Research

We are seeking to develop a technological theory of *modularity* and *design evolution* that can inform economic theories of industry evolution. Using the computer industry as a defining example, our book, *Design Rules*, explains:[1]

- what modularity is and how it can be attained;
- how modularity makes it possible for designs to evolve in a localized fashion (modular design evolution);
- how economic incentives operating at many points in a modular design eventually may lead to the emergence of a "modular cluster" of autonomous firms;
- how modular designs create economic conflicts whose resolution may require new institutions regulating finance, employment and intellectual property within the cluster.

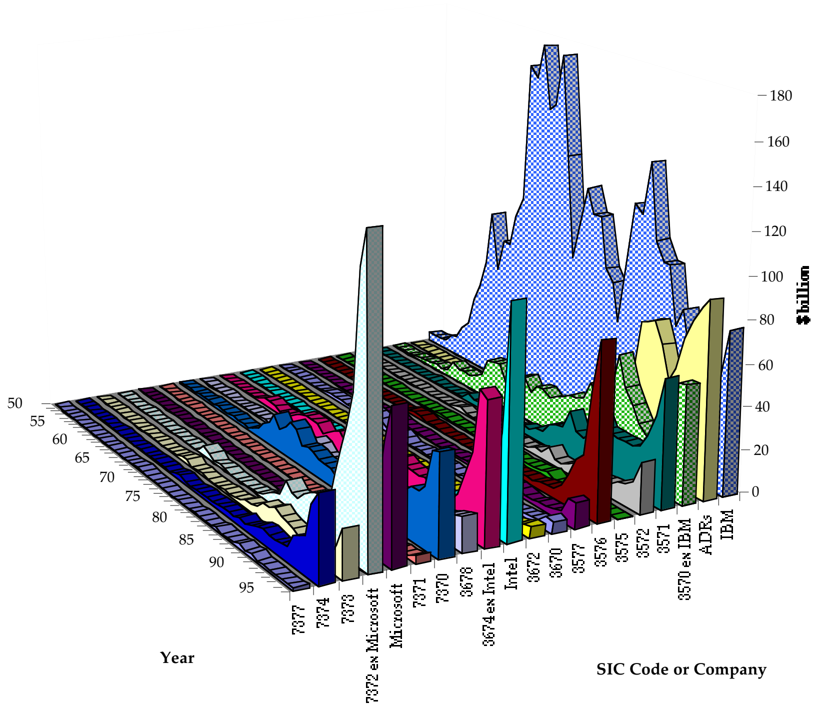
The best way to motivate the study of modularity is to show the effect it can have on the structure of an industry. We collected data on the market values of substantially all the public corporations in the computer industry from 1950 to 1996, broken out into sixteen subsectors (see Figure 1). The data tell a story of industry evolution that runs counter to conventional wisdom. The dominant theories of industry evolution describe a process of pre-emptive investment by large, well-capitalized firms, leading to stable market structures and high levels of concentration over long periods of time.[2] Figure 1 shows that there was indeed a period in which the computer industry was highly concentrated, with IBM playing the role of dominant firm. (IBM's market value is the blue "mountain range" that forms the backdrop of the chart.) But in the 1980s, the computer industry "got away" from IBM. In 1969, 71% of the market value of the computer industry was tied up in IBM stock; by 1996, no firm accounted for more than 15% of the total value of the industry.

---

[1] The arguments and all figures in this paper are taken from C.Y. Baldwin and K.B. Clark, *Design Rules, Volume 1: The Power of Modularity*, © MIT Press, 2000, reprinted by permission. Volume 2 is in progress.

[2] The original theory of pre-emptive investment leading to industry concentration, with supporting historical evidence, was put forward by Alfred Chandler (1962, 1977). A complementary theory of concentration following the emergence of a "dominant design" was put forward by William Abernathy and James Utterback (1978). Modern formulations of these theories and some large-scale empirical tests have been developed by John Sutton(1992) and Steven Klepper (1996). Oliver Williamson (1985, Ch. 11) has interpreted the structures of modern corporations (unified and multi-divisional) as responses to potential opportunism (the hazards of market contracting). It is our position that the basic "task structures" and the economic incentives of modular design (and production) systems are different from the task structures and incentives of classic large-volume, high-flow-through production and distribution systems. Therefore the organizational forms that arise to coordinate modular design (and production) may not ressemble the classic structures of the modern corporation.

Figure 1
The Market Value of the Computer Industry
By sector, 1950-1996 in constant 1996 US dollars



Source: Baldwin and Clark, 2000, Plate 1-1.

By 1996, the computer industry consisted of a large *modular cluster* of over 1000 firms, no one of which was very large relative to the whole. The total market value of the industry, which increased dramatically through the 1980s and 1990s, was dispersed across the sixteen sub-industries. Finally, the connections among products in the subindustries were (and are) quite complicated. Most computer firms did not design and make whole computer systems. Instead they designed or made *modules* that were parts of larger systems.

In *Design Rules*, *Volume 1: The Power of Modularity*, we argue that a fundamental *modularity* in computer designs caused the industry to evolve from its initial concentrated structure to a highly dispersed structure. Modularity allows design tasks to be divided among groups that can work independently, and do not have to be parts of the same firm. Compatibility among modules is ensured by "design rules", which govern the architecture and interfaces of the system. The design rules must be adhered to by all, and hence can be a source of economic power to the firms that control them.

Our theory of modular design and design evolution can be summarized as follows:

- Modularity creates options;
- Modular designs evolve as the options are pursued and exercised.

We explain and amplify these points below.

2 Modularity creates options.

When the design of an artifact is "modularized," the elements of the design are split up and assigned to modules according to a formal architecture or plan. Some of the modules are "hidden," meaning that design decisions in those modules do not affect decisions in other modules; some of the modules are "visible," meaning that they embody "design rules" that hidden-module designers must obey if the modules are to work together.

In general, modularizations serve three purposes, any of which may justify an investment in modularity:

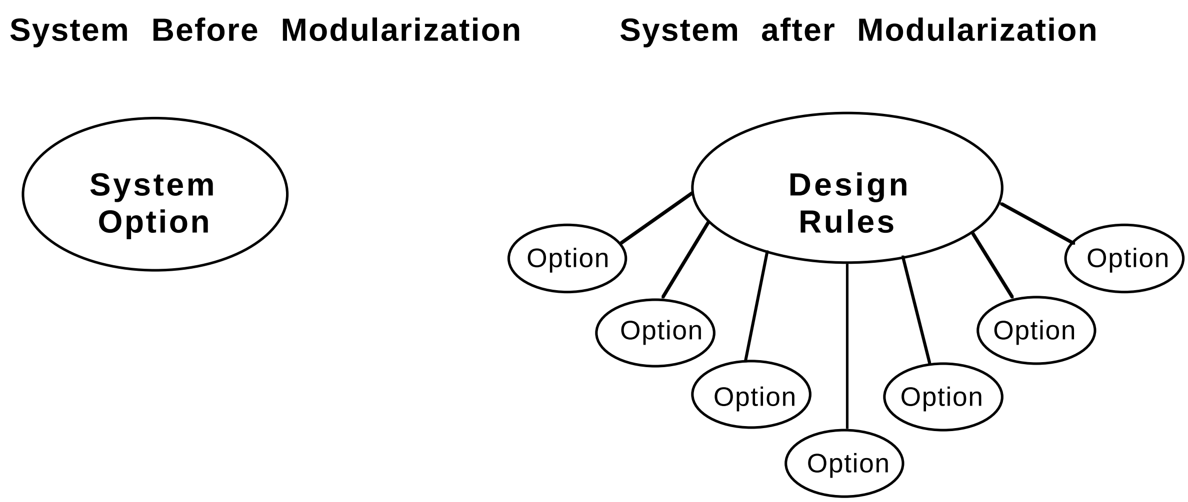
- Modularity makes complexity manageable;
- Modularity enables parallel work; and
- Modularity is tolerant of uncertainty.

In this context, "tolerant of uncertainty" means that particular elements of a modular design may be changed *after the fact* and *in unforeseen ways* as long as the design rules are obeyed.

Thus, modular designs offer alternatives that non-modular ("interdependent') designs do not provide. Specifically, in the hidden modules, designers may replace early, inferior solutions with later, superior solutions. Such alternatives can be modeled as "real options." Figure 2 portrays how the option structure of a system changes as it goes from an interdependent to a modular design structure.[3]

[3] A "modular design structure" is a particular structure of interdependencies among design or process parameters or, equivalently, tasks. The actual structure of any design or process or any set of tasks can be determined using the "Design Structure Matrix" mapping tools developed by Donald Steward (1981) and Steven Eppinger (1991). For numerous applications of this methodology, see http://web.mit.edu/dsm/publications_name.htm.

Figure 2
Modularity Creates Options



Source: Baldwin and Clark, 2000, p. 237.

The real options in a modular design are valuable. In *Design Rules* we sought to categorize the major options implicit in a modular design, and to explain how each type can be valued in accordance with modern finance theory. The key drivers of the "net option value" of a particular module are (1) the "technical potential" of the module (labeled σ, because it operates like volatility in financial option theory); (2) the cost of mounting independent design experiments; and (3) the "visibility" of the module in question. The option value of a system of modules in turn can be approximated by adding up the net option values inherent in each module and subtracting the cost of creating the modular architecture. A positive value in this calculation justifies investment in a new modular architecture.

2.1 An Example: The Value of Splitting

As an example, let V_1 denote the value of a one-module design with N tasks. Assume that the final outcome of the design process, denoted X, is a random variable that is normally distributed with mean zero and variance $\sigma^2 N$. Consistent with this being a one-module design, once the designers finish their task, they can only take the new design or leave it. The value of the new design *if it is superior to the old one* (whose value is normalized to zero), is:

$$V_1 = S_0 + E(X_N^+) \quad ; \tag{1}$$

where E( ) indicates "expected value" and X has a normal distribution with mean zero and variance $\sigma^2 N$. The "+" superscript on X means that the expectation applies only to outcomes above zero.[4]

Now let us suppose a design process is partitioned into j independent modules, while the number of tasks, N, remains the same. The expected value of the modular design, denoted V_j , is then:

$$V_j = S_0 + E(X_1^+) + E(X_2^+) + ... + E(X_j^+) \quad ; \tag{2}$$

where X_i is the contribution to overall system value of the ith module. Equation (2) indicates that each module's value can be compared to the benchmark established for that module. If the new module design has value greater than zero, meaning that its performance in the eyes of consumers is superior to the existing one's, the new design will be incorporated into the system. Otherwise the existing design will continue to be utilized.

As long as the distribution of the sum of module values remains the same as before the split, the modular approach is bound to yield a higher total value than the unmodularized approach. For one thing, the modular designers could tie their own hands and commit to take all or none of the new designs. If they did so, they could expect their design to perform as well as a corresponding one-module design. However, the modular designers can also consider module-

---

[4]More formally,

$$E(X^+) = \int_0^\infty X f(X)\, dX \quad ;$$

where f(X) is the density of a normal distribution. We use the simpler notation throughout the text. In the interest of simplicity and clarity, we also suppress adjustments for time and risk.

level improvements. These options only add to the value of the whole. Mathematically, the option values are reflected in the fact that each expectation in equation (2) ranges over the positive half of a probability distribution; realizations that are negative (i.e., fall short of the existing design) will be culled out.

Thus, holding the distribution of aggregate value fixed, higher degrees of modularity increase the value of a complex design. This result is a special case of a well-known theorem, first stated by Robert Merton.[5] For general probability distributions, assuming aggregate value is conserved, Merton showed that a "portfolio of options" is more valuable than an "option on a portfolio."

More generally, let X_α be the performance of a module of size αN. By assumption, X_α is a random variable, normally distributed with mean zero and variance $\sigma^2(\alpha N)$. We define z_α as follows:

$$z_\alpha = \frac{X_\alpha}{\sigma\,(\alpha N)^{1/2}} \;;$$

z_α is normally distributed with mean zero and variance one.

Substituting standard normal variates in equation (2), dropping S_0, and collecting terms, we have:

$$V_\alpha = \sigma N^{1/2} \left(\alpha_1^{1/2} + \alpha_2^{1/2} + \ldots + \alpha_j^{1/2} \right) E(z^+) \;; \qquad (3)$$

where $E(z^+)$ is the expectation of the right tail of a standard normal distribution and equals .3989.

In this fashion, we can compare the value of a modular design to the value of a corresponding unmodularized effort. We summarize the relationship in the following proposition:

Proposition 1. Under the assumptions of our model, let a design problem of complexity N be partitioned into j independent modules of complexity ($\alpha_1 N$, $\alpha_2 N$, ... , $\alpha_j N$). The modular design has value:

$$V_\alpha = \left(\alpha_1^{1/2} + \alpha_2^{1/2} + \ldots + \alpha_j^{1/2} \right) V_1 \;; \qquad (4)$$

relative to V the corresponding unmodularized design effort.

[5] Merton, 1973.

Proof.

By definition, a one-module design has both j and α_j equal to one. Thus $V_1 = \sigma \; N^{1/2} E(z^+)$. Collecting terms and substituting in equation (3) yields the result.

From the fact that $< \alpha_1, \alpha_2, \ldots, \alpha_j >$ are fractions that sum to one, it follows that the sum of their square roots is greater than one. Thus, as expected, if we ignore any change in system value and the costs of achieving modularity, a modular design is always more valuable than the corresponding non-modular design. Moreover, additional modularization increases value: if a module of size α is split into sub-modules of size β and γ, such that $\beta + \gamma = \alpha$, then the subdivided module's contribution to overall value will rise because $\beta^{1/2} + \gamma^{1/2} > \alpha^{1/2}$.

2.2 The Value of Parallel Experiments (Substitution)

In the previous section, we assumed that designers would create only *one* new design per module. However, they could as well decide to run several parallel experiments on each module and select the best of these outcomes. How much better is the first-best from the second- or third-best experimental design? The answer to this question is the value of the modular operator we call “substitution.” Intuitively, a modular design decouples experiments, and allows designers to substitute a superior design for an inferior design of any module.

To quantify the value of substitution, let us suppose that for a module comprising n tasks, the designers initiate k parallel, independent design efforts. They then have the option to select the best of k outcomes for the final design.[6] Let Q(X; k) denote the "value of the best of k designs" as long as it is better than zero, for a random payoff function X.[7] The value of a design process with j modules and k_i experiments in the ith module is then:

$$V(X_1 \ldots X_j ; k_1 \ldots k_j) = S_0 + Q(X_1; k_1) + Q(X_2; k_2) + \ldots + Q(X_j; k_j) \qquad (5)$$

Like equation (2), equation (5) is a “portfolio of options” result that applies to any set of probability distributions, as long as the total system value is conserved, and module values are

[6] Stulz, 1982, provides a general analysis of parallel options in his valuation of the option to select the maximum of two risky assets.

[7] This distribution of the best of k designs is well known in statistics: it is the distribution of the "maximum order statistic of a sample of size k." Lindgren, 1968.

additive. But, again, it is too general to be of much use. We can gain additional insight by focusing on normal distibutions and symmetric modules.

In general, the appropriate number of experiments to run on a module depends on that module's technical potential, complexity, and visibility to other modules. If, however, we assume that modules are symmetric, then it will be optimal to run the same number of experiments on each module. The 2j arguments in equation (5) then collapse to two, and the value of the design process as a whole, denoted V(j,k), is:

$$V(j,k) = S_0 + \sigma (Nj)^{1/2} Q(k) \qquad (6)$$

2.3 The Costs of Modularity

Modularity creates value but it is not free. There are, first of all, the costs of making an interdependent system modular: the cost of creating so-called "hidden modules" and disseminating design rules. The process of of modularizing a complex system is generally a lengthy, pain-staking process for e*very important design dependency must be understood and addressed via a design rule.*

Obviously the density of the dependencies matters here: some systems are naturally more "loosely-coupled" than others. Circuits, the physical system on which computers are based, are one-dimensional; whereas mechanical solids are three-dimensional. Clearly it is harder to split up complex, curved, 3-dimensional designs, and to create flexible interfaces for them: there are more dependencies to manage, and the tolerances are much tighter. Thus modularizing an automobile's design is a tougher problem than modularizing a circuit design: the cost of creating a modular architecture and related interfaces for an automobile will be higher than for a VLSI circuit.[8]

It is also costly to run the experiments needed to realize the potential value of a modular system. Finally, it is costly to design the tests that are needed to determine whether specific modules are compatible with a given system, and which modules perform best. The costs of

[8] This has led some scholars, like Daniel Whitney at MIT, to predict that autos and airplanes will achieve only limited modularity in practice. If that is so, the option values of such systems will be limited in relation to systems that *can* be modularized. See Whitney, 1996; Sharman, 2002; Sharman, Yassine and Carlile, 2002.

architecture, experiments, and tests are all inherent in the modular design process itself. The interaction of option value and these costs causes each module in a large system to have a unique value profile.

Additionally, if the module experiments are distributed over a cluster of independent firms, there will be transaction costs: what Ronald Coase labelled "the costs of using the market."[9] As designs and/or physical artifacts are transferred from enterprises working on modules to enterprises working on systems integration and testing, there will be costs of search and price determination. Hence, the division of modular design efforts across a decentralized cluster of firms will multiply transaction costs in the same proportion that it multiplies modules and experiments.

Last but not least, agency costs and costs of opportunistic behavior are implicit in the modular system and its institutional surroundings. There may be opportunities for the architects of the modular system and the systems integration and testing groups to "hold up" other parts of this system by threatening to withdraw their services or their intellectual property. In other words, there are potential "points of control" in the modular system. This in turn means that the distribution of value across agents and the dynamic evolution of the system itself will depend on the allocation of key property rights: for example, who owns the architecture? Who owns the interfaces? Who has rights of access to design information? Who has rights of exclusion?

3 Modular designs evolve as the options are pursued and exercised.

Modular designs create value in the form of valuable real options. But how will that value be realized? In *Design Rules*, we argue that the value of a modular system will be realized over time via *modular design evolution* (MDE).

The promise implicit in a modular design is that parts of the system — the modules— can be modified after the fact at low cost. Foresighted actors seeking financial rewards will thus be motivated to pursue these options, and they will exercise the ones that are "in the money" at some future point in time (the actual date may be uncertain). Exercising an "in the money" option in this case means introducing a new, superior version of a particular module and reaping the economic rewards. The rewards may take the form of higher product revenue, or lower process cost, or both.

The valuable options in a modular design thus motivate economic actors to pursue innovation, and the exercise of the options constitutes innovation. It follows that a modular design defines a set of evolutionary paths or trajectories in the sense originally defined by Nelson and Winter (1977), Sahal (1983), and Dosi (1988), and others.[10] There will be at least one trajectory per hidden module, and there may be more if the full potential of the actions we call "modular operators" is realized.[11]

As the history of a modular design unfolds, if the promise of the options is realized, we will "see" design evolution. The economically motivated actors in the system will pursue and then exercise design options on the basis of their inherent economic value. Their innovations will cause the individual hidden module designs to change over time in ways that create economic value. Architectures and interfaces will sometimes change, too, but less frequently.

This, we argue, is how innovation works in the microcosm of a modular system. Most changes will not be big sweeping disruptions of the whole, although those are not ruled out. Most changes instead will involve replacing one small modular element with another correspondingly small element that will do the same job in the system, only better. The overall picture is one of ordered, *but not wholly predictable*, progress towards higher economic value over time.

---

[9] Coase, 1937.

[10] See, for example, David, 1987; Langlois, 1992; Langlois and Robertson, 1992; Tushman and Murmann, 1998.

[11] Operators are "units of action" in a formal model of a complex adaptive system. The concept is due to John Holland, 1992.

References

Abernathy, William and James Utterback (1978) "Patterns of Industrial Innovation," *Technology Review* 80:41-47.

Alexander, Christopher (1964) *Notes on the Synthesis of Form*, Harvard University Press, Cambridge, MA.

Aoki, Mashiko (1999) "Information and Governance in the Silicon Valley Model," Stanford University, http://wwwecon.stanford.edu/faculty/workp/swp99028.html, viewed Jan 12, 2001.

_______ (2001) *Towards a Comparative Institutional Analysis*, MIT Press, Cambridge, MA.

Baldwin, Carliss Y. and Kim B. Clark (2000) *Design Rules, Volume 1, The Power of Modularity*, MIT Press, Cambridge MA.

Chandler, Alfred D. (1962) *Strategy and Structure*, MIT Press, Cambridge, MA.

_______ (1977), *The Visible Hand*: *The Managerial Revolution in American Business*, Harvard University Press, Cambridge, MA.

Coase, Ronald H. (1937) "The Nature of the Firm," *Economica* 4:386-405.

David, Paul A. (1987). "Some New Standards for the Economics of Standardization in the Information Age," in Partha Dasgupta and Paul Stoneman, eds., *Economic Policy and Technological Performance*. Cambridge: Cambridge University Press.

Dosi, Giovanni (1988) "Sources, Procedures, and Microeconomic Effects of Innovation," *Journal of Economic Literature*, 26(September):1120-1171.

Eppinger, Steven D. (1991) "Model-based Approaches to Managing Concurrent Engineering" *Journal of Engineering Design*, 2: 283-290.

Holland, John, H. (1992) *Adaptation in Natural and Adaptive Systems*, MIT Press, Cambridge, MA.

Klepper, Steven (1996) "Entry, Exit, Growth and Innovation over the Product Life Cycle, *American Economic Review*, 86(30):562-583.

Langlois, Richard N. (1992) "External Economies and Economic Progress: The Case of the Microcomputer Industry," *Business History Review*, 66(1):1-51.

________ and Paul L. Robertson (1992). "Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries," *Research Policy*, 21: 297-313.

Lindgren, B.W. (1968), *Statistical Theory*, MacMillan, New York, Chapter 8.

Merton, Robert C. (1973) "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4(Spring): 141-183; reprinted in *Continuous Time Finance*, Basil Blackwell, Oxford, UK, 1990.

Nelson, Richard R. and Sidney G. Winter (1977) "In Search of a Useful Theory of Innovation," *Research Policy*, 6(1): 36-76.

Parnas, David L. (1972a) "A Technique for Software Module Specification with Examples," *Communications of the ACM* 15(May): 330-36.

_________ (1972b) "On the Criteria to Be Used in Decomposing Systems into Modules," *Communications of the ACM* 15(December): 1053-58.

_________, P.C. Clements, and D.M. Weiss (1985) "The Modular Structure of Complex Systems," *IEEE Transactions on Software Engineering*, SE-11(March): 259-66.

Sahal, Devendra (1985) "Technical Guideposts and Innovation," *Research Policy*, 14:61-82.

Sanchez, R. and Mahoney, J. T. (1996) "Modularity, flexibility and knowledge management in product and organizational design". *Strategic Management Journal*, 17: 63-76.

Sharman, David M. (2002) *Valuing Architecture For Strategic Purposes*, unpublished M.Sc. thesis, MIT, Cambridge, MA.

________, Ali A. Yassine, and Paul Carlile (2002) "Characterizing Modular Architectures," Proceedings of DETC '02, Design Theory & Methodology Conference, Montreal, Canada, September.

Simon, Herbert A. (1962) "The Architecture of Complexity," *Proceedings of the American Philosophical Society* 106: 467-482, repinted in *idem*, *The Sciences of the Artificial*, 2nd ed. Cambridge: MIT Press, 1981.

Spear, Steven J. (1999) *The Toyota Production System: An Example of Managing Complex Social/Technical Systems*, unpublished DBA dissertation, Harvard University, Graduate School of Business Administration, Boston, MA.

Steward, Donald V. (1981) "The Design Structure System: A Method for Managing the Design of Complex Systems," *IEEE Transactions on Engineering Management* EM-28(3): 71-74 (August).

R.M. Stulz, (1982) "Options on the Minimum or Maximum of Two Risky Assets," *Journal of Financial Economics*, 10: 161-185.

Sutton, John (1992) *Sunk Costs and Market Structure*, MIT Press, Cambridge, MA.

Tushman, M.L. and Murmann, J.P. (1998) "Dominant designs, technological cycles and organizational outcomes" in ***Research in Organizational Behavior***, (Staw, B. and Cummings, L.L. eds.) JAI Press, Vol. 20.

D.E. Whitney, 1996, "Why Mechanical Design Cannot Be Like VLSI Design," http://web.mit.edu/ctpid/www/Whitney/morepapers/design.pdf, viewed April 9, 2001.

Williamson, Oliver E. (1985), "The Economic Institutions of Capitalism," Free Press, New York, NY.

Comments Welcome

The Fundamental Theorem of Design Economics

Carliss Y. Baldwin
Kim B. Clark

Harvard Business School

This Draft:
March 14, 2002

Our thanks to Masahiko Aoki, Wayne Collier, Sonali Shah, David Sharman, Kevin Sullivan, Jason Woodard, and members of the Negotiations, Organizations and Markets group at Harvard Business School for sharing key insights. We alone are responsible for errors, oversights and faulty reasoning.

Direct correspondence to:
Carliss Y. Baldwin
Harvard Business School
cbaldwin@hbs.edu

1 The Fundamental Theorem

We are often asked, how general is the theory set forth in *Design Rules, Volume 1*? Does it apply to artifacts other than computers? Here we establish the scope of the design valuation methodology based on real options, which lies at the core of our theory.

To begin with, we need two definitions:

Define an *economic process* as a method that converts inputs, tangible and intangible, into outputs that have economic value. In other words, the outputs of the process can be sold for money. Define a *value representation* as a mathematical formula that sums up the financial costs and benefits of an economic process, using market prices for all inputs and outputs, and asset prices for all future cash flows.

Lemma. If an economic process is indivisible, then k, the number of processes to run in parallel will be an *ex ante* decision variable.

Proof.

An economic process is a method, and thus it can be enacted more than once. How many times to enact the process is an *ex ante* decision variable. But if the process is divisible, then the decision variable is a vector: the decision maker must decide how many times to enact each subprocess.

Proposition 1. (The Fundamental Theorem.) If an economic process is:

- indivisible;
- *ex ante* uncertain;
- *ex post* rankable by outcome;
- *ex post* contingent;
- costly; and
- has non-exclusive outputs;

and if better outcomes have higher financial value (are worth more money), then:

a. the value representation of the k processes will have the form:

$$V(k) = Q(k) - C(k)$$

where Q(k) is the present value of the expectation of the maximum of the outcomes of the k parallel processes:

$$Q(k) \equiv V_0[E(X^* \mid X^* = \max(X_1 \dots X_k)] \quad ;$$

and C(k) is the cost of the inputs to the k parallel processes. Note that Q(k) is both an order statistic expression and a real option expression.[1]

b. *Ex ante* optimization of these processes takes the form of finding and selecting k* such that:

$$V(k^*) = \max_k \; [Q(k) - C(k)] \quad .$$

c. Optimal k may be greater than one.

Proof:

1a. Assume the decision maker runs k processes. After the fact, the processes will have k outcomes, $<X_1, \dots X_k>$.

By the fact that the outcomes are rankable, there exists a highest X.

By the fact that the processes are contingent, the outputs which have the highest X can be used or supplied or applied to the purpose needed.

By the fact that the outputs are non-exclusive, those with the highest X can be used as many times as they are needed: the outputs of the other processes may be discarded.

The expected outcome obtained from running the k processes and selecting the best will be:

$$E[X^* \mid X^* = \max(X_1 \dots X_k)] \quad .$$

By the definition of economic process (see above), this outcome has a set of financial values in the future, that is a set of cash flows. If the cash flows are uncertain, they have expectations, and the expectations are well-defined functions of k. The present value of a series of expected cash flows in the future is a well-defined asset price. It, too, is a function of k, and we denote it Q(k):

---

[1] Cf. Lindgren, 1968; Merton, 1973.

$$Q(k) \equiv V_0\{E[X^* \mid X^* = \max (X_1 \ldots X_k)]\} \quad ;$$

Here the valuation operator, V_0, denotes both the mapping of outcome X^* to cash flows, and the conversion of future cash flows into a present value (an asset price).

Finally, by value additivity, we can subtract the present value of costs, denoted C(k), and represent the total value as:

$$V(k) = Q(k) - C(k) \quad .$$

1b. We have shown that V(k) is a well-defined function of the integer variable k. Therefore,

$$\max_k V(k)$$

is well-defined. The highest economic value in the present is obtained by selecting k^* = argmax V(k). (Note that while there is only one max V(k), there may be several argmaxes of the function. Selecting any argmax suffices for optimization.)

1c. If an economic process is *ex ante* uncertain, rankable, contingent, non-exclusive and *not costly*, then there is no *ex ante* upper bound to the number of times it should be run. The expectation of the maximum of k trials of the process is then strictly increasing in k.

But if the process is costly, then there may be an upper bound to the number of times the process should be run. Indeed, there exist cost structures that can serve to make any number of trials, from 0 to infinity, the optimal number of trials. One such cost structure is:

$$C(k) = k\varepsilon \qquad \text{where } \varepsilon < Q(k^*+1) - Q(k^*) \quad , \qquad 0 \le k \le k^*;$$

$$C(k) > kQ(1) \quad , \qquad k > k^* \quad .$$

Under this cost structure, the optimal number of trials is k^*, which can be 0 or any positive integer. The cost structure works because $Q(k+1) - Q(k)$ is strictly decreasing in k. Thus a per-trial cost less than $Q(k^*+1) - Q(k^*)$ justifies investment in trials 1 through k^*; while a per-trial cost greater than Q(1) makes all trials unprofitable. **QED.**

2 The Fundamental Theorem in Modular Systems

By definition, modules are indivisible units of design activity within a larger, divisible and hierarchical system. Design processes focused on modules thus generally conform to the premises of the fundamental theorem.

However, moving up to the next level of aggregation, *modular systems* offer many complex and interesting ways of combining and recombining modules. In *Design Rules, Volume 1*, we attempted to capture those opportunities via six modular operators:

- splitting;
- substituting;
- excluding;
- augmenting;
- inverting; and
- porting.

The operators are logical actions that are applicable to individual modules and/or subsets of modules in a system of designs.

We went on to derive a generic value representation for each of the operators. Not surprisingly, one or more Q(k)-type functions appeared in the value representation of each operator.

The value of a system of modular design processes can thus be represented as a complex aggregation of Q(k)-type real options. However, the combinatorial properties of the operators quickly outpace the computational capacity of any known information-processing entity. As a result, even in small modular systems it is impossible to calculate, much less optimize, the value of the system over all possible alternative paths. In effect, sheer complexity mandates both decentralized decision-making and less-than-perfect optimization of an evolving modular system of designs.

References

Baldwin, Carliss Y. and Kim B. Clark (2000) *Design Rules, Volume 1, The Power of Modularity*, MIT Press, Cambridge MA.

Lindgren, B.W. (1968), *Statistical Theory*, MacMillan, New York, Chapter 8.

Merton, Robert C. (1973) "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4(Spring): 141-183; reprinted in *Continuous Time Finance*, Basil Blackwell, Oxford, UK, 1990.

Comments Welcome

Institutional Forms, Part 1:
The Technology of Design and its Problems

Carliss Y. Baldwin
Kim B. Clark

Harvard Business School

This Draft:
March 22, 2002

Our thanks to Masahiko Aoki, Wayne Collier, Sonali Shah, David Sharman, Kevin Sullivan, Jason Woodard, and members of the Negotiations, Organizations and Markets group at Harvard Business School for sharing key insights. We alone are responsible for errors, oversights and faulty reasoning.

Direct correspondence to:
Carliss Y. Baldwin
Harvard Business School
cbaldwin@hbs.edu

1 Introduction

We are setting out to write a series of working papers whose purpose is to explore a variety of institutional forms that can be used to exploit a "technology of design." We are looking for forms that are institutions in the sense defined by Aoki (2001). Such institutions can be conceptualized and analyzed as equilibria of socially constructed, repeated games.[1]

The premise of this investigation is that institutional forms are built on "technological substrates." Substrates define and delimit what is physically possible in nature. Therefore, the technological substrate of a set of institutional forms defines both the opportunities open to and natural constraints imposed on the forms. On top of the technological substrate, human beings may construct institutional forms, either consciously by design, or in a decentralized, "evolutionary" manner. The institutional forms are social mechanisms and structures that make it possible to realize the opportunities of the substrate without violating its physical or logical constraints.

Thus, we stipulate, *the purpose of institutional forms is to "solve" problems of action posed by a technological substrate*. For any given technological substrate, however, several institutional forms may be effective.[2] What are these forms? How do they work? How well do they perform? And how they are related to one another, ie, can they coexist? These are the main questions that we will address in this investigation.

The particular technological substrate investigated here is a "technology of design." We posit a population of qualified individuals called "designers," who repeatedly encounter and solve design problems. Their problems have a particular structure in which designers try to find the best of many possible solutions created by independent trials or experiments. We call this a "Q(k) structure," which we formally define below. In *Design Rules, Volume 1,* we offered evidence to the effect that Q(k) structures correspond to the way real design problems present themselves to designers in the computer industry.[3] Here we are building on that prior work.

---

[1] Aoki, 2001. See also Greif, 1994, 1997; Young, 1998.

[2] Aoki, 2001.

[3] Q(k) lies at the heart of the six modular operators described in Baldwin and Clark, 2000.

The solving of problems creates new designs, which in turn cause the technological substrate of designs to evolve. In this fashion, a dynamic substrate is an important feature of the the designers' world. It is at once their collective product, and something that no single designer controls. In this setting, institutions are needed in order to promote information sharing; to coordinate effort; and to reward designers for their effort and accomplishments.

In our investigation, we will consider three paradigmatic institutional forms in relation to the technological substrate of design:

- A "guild," which might also be called a professional association;
- A monopolistic firm; and
- A cluster of firms.

The modeling approach used in this analysis is evolutionary game theory. This approach presumes that there is a population of types (agents, animals, genes, designers) that interact with their environment and each other in some regular way. Because we are after an understanding of stable insitutional forms, we will pay special attention to the equilibria that emerge out of these interactions.[4]

The designers' types are the above-named institutional forms: "Guild," "Monopoly," and "Cluster." One can think of each designer as belonging to, or being employed in, one of the three institutional forms.

Our work on the institutions that support design processes is organized into a series of papers as follows. This one, the first, sets up a "model world," and a base case. In Section 2, we characterize the technological substrate of design activity and begin to define the designers' preferences. In our specification, we make a crucial assumption: that designers have preferences with respect to the substrate, as well as preferences over standard variables, such as wealth and effort. In Section 3, we define a base case: an economic environment that "puts the designers to work," but causes their efforts to be separate and uncoordinated. In Section 4, we show how the pooling of design solutions and the coordination of designers' efforts generates a (potentially large) surplus for designers and society. We also consider what is necessary for the pooling of

[4] An Evolutionary Stable Strategy (ESS) is an equilibrium concept in evolutionary game theory, which was first defined by Maynard Smith and Price, 1973. In essence, an ESS is a distribution of types in the

solutions and the coordination of effort to occur. These are technical requirements that must be satisfied by effective institutional forms that "operate" on this substrate.

The technical requirements of institutional forms define their purpose in a particular setting. To the technical requirements, we add an economic requirement: the institutional form must "break-even" in a financial sense vis a vis the greater economy. By this we mean that the financial cost of sustaining the institutional form must not exceed the benefits obtained by those who directly pay for it. We label this the "sufficiency constraint."

In our model world, institutional forms must fulfill a purpose in relation to their substrate and they must be financially sufficient. Then, and only then, can they form the basis of self-sustaining shared beliefs, hence be institutions under Aoki's definition.

In subsequent papers, we will define the three institutional forms named above; evaluate their effectiveness; and characterize their behavior (performance) under binding sufficiency constraints. We do this first for a guild (G), then for a monopoly (M), then a cluster (C).

Initially, we consider competition between institutional forms within a single Q(k) design process. Later, however, we will enlarge our view to consider multiple and heterogeneous Q(k) processes such as would characterize an evolving modular system (cf. Baldwin and Clark, 2000). We will show that each of the three institutional forms is best adapted to a different set of Q(k) processes. This means that *the three forms can coexist in equilibrium in a sufficiently large and diverse modular system*.

However, we will also show that a monopoly has "evolutionary interests" that are opposed to those of the guild and the cluster. In other words, on the same technological substrate, a system of designs controlled by a monopoly will evolve toward a different structure (architecture) than a system of designs controlled by a guild or a cluster. We then consider the implications of this result for inter-system competition.

---

population that is stable in response to "invasions" by competing types. The biological idea is that an ESS is robust against mutations: the alternatives types cannot gain a foothold in the population.

2 The Technological Substrate of Design

Our analysis takes place in the following technological context (a "technological substrate"):

2.1 Designers and Their Problem-Solving Domain

1) In some specific economic domain, there are problems whose existence is known, but whose solutions are unknown *ex ante*.
2) There is a population of professional "problem solvers" in the domain, whom we call designers.
3) The class of designers is closed: to be a designer in the domain requires specialized talents and training.
4) The number of problems known to the population of designers at any given time is denoted M. Designers have bounded cognition, thus M is finite but may be quite large. Moreover, as particular problems are solved, new ones may come to be perceived by the designers, thus replenishing the pool of problems.
5) For each *specific* problem (each member of the pool of problems), there are multiple problem-solving paths leading to different solutions. Solutions to problems can be generated at the cost of effort.
6) The same problem can arise many times. We will call the separate occurrences of a single problem the *instances* of that problem.

2.2 Design Effort and the Use of Solutions

7) Each qualified designer is capable of generating one and only one solution in a given amount of time. Only solutions obtained through such effort are available.
8) Only one solution (one design) can be applied to any given *instance* of a problem. However, there are usually many instances of a problem. The use of a particular solution in one instance does not preclude the same solution being used in another instance. (Designs are non-exclusive information goods.)

9) However, in addition to being available, a solution must be known to a designer in order to be used to solve an instance of a specific problem.

2.3 The Quality of Solutions

10) The quality of solutions to a given problem can be measured. Let the scale of measurement be called "x." The x-scale ranges over positive and negative numbers, and is continuous. Higher x's are better.

11) For convenience, the best prior solution to the problem is located at the origin x=0. As part of their training, all designers are assumed to know this prior solution (the "textbook solution").

12) All designers agree on the metric x. In particular, given a set of candidate solutions, all designers will agree on which is best (has highest x).

13) As indicated, a problem-solving effort uses one unit of one designer's time. It obtains a draw from a distribution G(x), whose support is the real line. Designers are assumed to draw solutions independently of one another from the G(x) distribution.[5]

14) Define R(k) as follows:

$$R(k) = \int_0^{\infty} x\, kG(x)^{k-1} g(x)\, dx \quad ;$$

R(k) is a modified highest order statistic: in words, it is the expected value of the *highest x* among k independent solutions drawn from the distribution G(x). The expectation is truncated in that values of x that are less than zero (the textbook solution's benchmark value) are replaced by zero values. The truncation arises because of a designer's option: if x<0, then the solution drawn from the G(x) distribution is worse than the previously known solution (x = 0). In that eventuality, we assume that a designer can fall back on the the the previously known solution, and achieve x = 0. This is a basic "designer's option."[6]

---

[5] This assumption is less restrictive that it appears: we can partition the quality measure, and define x as that part which is independent across designers.

[6] R(k) and Q(k) are monotonically increasing and concave in k. Lindgren, 1968.

15) R(k) is an *ex ante* expectation. An *ex post* realization of the highest x among k independent solutions drawn from the distribution G(x) will be denoted R*(k). As usual, R(k) = R*(k) + ε, where E(ε) = 0.

2.4 Designers' Preferences

16) We assume that designers care about money, effort *and the quality of the designs in their problem solving domain*. (In subsequent papers, we will have them care about other things, e.g., winning a tournament.)

17) *Ceteris paribus*, all designers prefer more wealth to less; and less effort to more.

18) *Ceteris paribus*, all designers would choose to apply the best (highest x) available solution to any instances of a specific problem they might encounter.

19) Designers can forecast changes in wealth, effort and expected design quality. They have well-defined, rational preferences over all points in the subspace defined by these M+2 variables.[7] Furthermore, in allocating their time, they are one-period-look-ahead utility maximizers, whose per-period utility function is:

$$u(w, x_1, \ldots, x_M, e)$$

where w denotes income received at the end of the period, x_i denotes the end-of-period quality of design solution for the i^{th} known problem, and e denotes intra-period effort. (**Note:** In a dynamic programming context, a designer's short-term preference for higher quality designs may be a direct aesthetic preference, or a derived preference for lower effort in the future. In the latter case, the quality of known design solutions functions like an asset for the designer.)

[7] For simplicity, we assume that the designers are risk neutral with respect to design quality.

3 The Base-Case Local System Has High Transaction Costs, No Information Sharing and No Coordination

Our objective in this section is to specify a "base case" economic environment that can serve as the starting point of our analysis of comparative institutional forms. In our analysis, we want institutions to arise endogenously, as solutions to the problems posed by the technology (in this case, the technology of design). However, it is impossible to start this type of analysis in an institution-free world.[8] Thus for analytic purposes, we shall divide the model world into two parts: the "greater economy" and the "local system."

The *greater economy*, we assume, is a large and complex social and economic system. It has many technologies and institutions, which are exogenous to our model.

The *local system* is our unit of analysis. It consists of:

- a particular technological substrate, which defines opportunities;
- players ("actors") who can realize the opportunities and benefit from their realization; and
- institutional forms that structure the players' actions and interactions.

As the starting point of our analysis, we want to define a local system and endow it with the generic institutional forms characteristic of the greater economy. Thus, if the greater economy uses money and credit, has corporations, and utilizes contracts, we will assume that actors in the local system may use money and credit, create corporations, and enter into contracts.

But, crucially, at the start of the analysis, the local system should not have *specialized* institutional forms that address the specific needs of the particular technological substrate. Instead, we view specialized institutional forms as endogenous: such forms will be the focus of our analysis in sections that follow.

We maintain that in the real world, specialized institutional forms may be constructed on particular technological substrates for the following purposes:

- To reduce frictional transactions costs, including, but not limited to, search costs;
- To enable the collection, sharing and dissemination of information insofar as those activities are valuable within the substrate;
- To facilitate coordination insofar as it is valuable within the substrate.

[8] Aoki, 2001; Greif, 1998.

On this view, for example, supermarkets are a specialized institutional form which, in the late 20th Century in the US, reduced the cost of searching for food and other household supplies; they replaced open-air markets as well as the shops of "butchers, bakers and candlestick-makers," which performed the same functions in earlier centuries.[9] Symmetrically, train, bus and airline schedules are specialized institutional forms that collect and disseminate information to travelers and carriers; and reservation systems are specialized institutional forms that facilitate the coordination of travel plans. As an institutional form, supermarkets were constructed on a complex technological substrate of food supply and distribution; schedules and reservation systems rest on the technological substrates of their particular modes of transport and communication.

If our assumption about the purpose of specialized institutional forms is true in the real world, then in our model world, the base-case local system (the local system without specialized institutional forms) should have relatively high frictional transactions costs; high information costs; and low degrees of coordination within the substrate. *Specialized institutional forms can then be introduced into the local system,* where they will act as "mutations" in the standard evolutionary game theoretic manner.[10] In general, useful forms will "move" the "local system" (consisting of the technological substrate, the players, and the forms themselves) closer to the neo-classical ideal of frictionless transactions, perfect information, and full coordination.[11]

In what follows, we assume that our designers and their problems are lodged in a market economy of the late 20th Century.[12] In particular, the "greater" economy supports voluntary *bilateral* transactions, *bilateral* formal contracts and *bilateral* relational (self-enforcing) contracts. The greater economy also has consumers, firms, money, and methods of valuing future expected cash flows. However, in the greater economy, searching, transacting, enforcing contracts, collecting and publishing information, coordinating actions, forecasting, and

---

[9] Braudel, 1982.

[10] Maynard Smith, 1982; Gintis, 2000; Aoki, 2001. See also, Dawkins, 1989; Lumsden and Wilson, 1981.

[11] Merton and Bodie argue that "moving the system" closer to the neo-classical ideal of perfect markets is the role of essentially all financial institutions. (Private communication, December 2000; see also, Merton and Bodie, 1995.)

[12] They do not operate in, for example, a tribe, a peasant village, or a Renaissance city.

measuring economic value are all costly. Furthermore, non-standardized, one-off searches, transactions, contracts, etc. are more costly than standardized searches, transactions, contracts, etc. Finally, precise forecasting and accurate measurement of economic value are more costly than imprecise forecasting and approximate measurement.

We must now link the greater economy to the designers' local system. We do this by defining a set of actors called "users." Users create a demand for designers' services and allow them to be paid for what they do.

3.1 Users in the Problem Solving Domain

1) In the greater economy, there are entities called "users." Users are willing to pay designers to solve (instances of) problems. In some circumstances, users will be willing to pay more for better solutions to their problems. (We will discuss what those circumstances are below.)
2) With respect to designs, users are risk-neutral expected value maximizers. For most purposes, they can be thought of as firms.

3.2 Users' Economic Payoffs

3) Consider a specific problem, with its associated x-scale (defined above). We assume there is a mapping $x \rightarrow v(x)$, where $v(x)$ represents the money a user is (or should be) willing to pay for a solution of quality x to that problem.[13] We further assume that $v(x)$ is non-decreasing and quasi-concave in x, and that $v(0) = w$, a known value.
4) Let $F(v(x)) = G(x)$. In other words, the distribution $F(v)$ is induced on the scale v (measured in dollars) by the distribution $G(x)$. $F(v)$ is the distribution of *economic payoffs* associated with the designers' efforts to solve problems.
5) For future reference, define $Q(k)$ symmetrically with $R(k)$ as follows:

$$Q(k) = w \int_{-\infty}^{w} kF(v)^{k-1} f(v)\, dv + \int_{w}^{\infty} v\, kF(v)^{k-1} f(v)\, dv \quad ;$$

[13] Strictly speaking, willingness-to-pay function might vary from one user to the next. For simplicity, in this analysis we assume that users are identical in terms of their willingness-to-pay.

In words, Q(k) is the expected value of the *most valuable solution* among k independent solutions drawn from the distribution F(v). As before, solutions that are worth less than w (the benchmark) are replaced by w. This truncation reflects the designer's option to fall back on the previously known, textbook solution, whose value is w.

6) Symmetrically with R*(k), define Q*(k) as the *ex post* realization of the highest v among k independent solutions drawn from the distribution F(v).

3.3 Users' Information and Contracting Constraints

7) Before the fact, users know they have problems, but they do not know x (how to compare solutions) nor G(x) (the distribution of performance as a function of designers' efforts). Therefore, when contracting to get their problems solved, users do not know F(v(x)) in any detail.
8) Nevertheless, designers must be paid *ex ante* for their efforts (they have to eat!).
9) After the fact, a user can compare the actual values of a designer's solutions to actual payments made to that designer (the designer's remuneration). However (as is usual in this economy), more precise measurements of value are more costly than less precise measurements.

3.4 Completing the Base Case: Bilateral Contracts in a "Robinson Crusoe" Design Economy

At this point in our analysis, the technological substrate of our local system is the so-called "Q(k) design process", and the actors are the designers and users. We now need to specify the way in which the designers and users interact. The institutional forms that result will round out our specification of the base-case local system.

For simplicity, let us assume that there are n users and n designers, and that each user-designer pair has (the same) M known problems to be solved. After receiving their training, because communication and coordination are costly, the designers do not communicate with each other, nor do they attempt to coordinate their efforts in any way. This is a "Robinson Crusoe" type of design economy.

Under the assumptions we have laid out, one feasible form of contract between designers and users is a simple bilateral "contingent renewal" contract:[14] Under this contract, the user will retain the services of the designer over some fixed contract period, T, where T is a parameter of the contract. The user will pay the designer a per-period fee or wage, ω, which is also a parameter of the contract. During the contract period, the designer will work on the user's problems according to the designer's own priority rankings. At the end of the contract period, the user will assess the economic value of the designer's work, and decide whether to continue to employ the designer.

Our assumptions give designers sole control over the choice of problems and the development (or choice) of solutions. We think of designers as highly skilled professionals with expertise not available to the users. Hence, the way designers rank problems and allocate effort to them is an important feature of the local system.

We have said that designers place positive values on wealth, and the quality of solutions, and negative values on effort. We also stipulate that, in order to earn his or her remuneration, ω, each designer must put in some basic level of effort, e, in each problem-solving interval; however, this condition can be satisfied by applying the "textbook solution" to each problem. This is implied by the assumption that $v(0)=w$, as long as $\omega \leq w$.

Alternatively, a designer can try to devise a new solution to a specific problem: each attempt to find a new solution requires an additional quantity of effort, Δe. Under our base-case assumptions, new solutions will not affect the designer's remuneration, but they will affect the designer's utility. We derive the designer's optimal decision rule below.

This very basic bilateral contract – user hires designer for T periods; designer selects problems and allocates effort to find or use solutions; user pays designer and claims value (such as it is) – completes the base-case local system. We have a technological substrate, the set of players and an institutional form. But we have yet to determine whether this local system is

[14] Gintis, 1976; Shapiro and Stiglitz, 1984; Bowles, Gintis and Osborne, 2001.

stable. In other words, does the system have the properties of an Aoki-type institution? In fact, there are two parts to the question: 1) Does the bilateral contract we have specified give rise to an equilibrium in the game played between users and designers? And 2) does the local system create sufficient value to warrant its continuation? To address these questions we must define more carefully the nature of the game and the potential payoffs to the designers and users.

3.5 The Contingent Renewal Contract Game

Because of the "contingent renewal" character of the bilateral contract, we can frame the relationship between user and designer as a simple game. At the beginning of the contract, a specific user-designer pair can choose end dates ranging from 1 (the time needed to solve one problem) to T_{max} (the designer's lifetime). At the end of each contract interval, the user will measure the economic value of the designer's work incurring measurement costs. The designer and user will then have symmetric options to continue the relationship, or split up and search for new partners (incurring search costs). *In addition, the user will have the option of not employing a designer over the next contract interval.*

Figure 1 shows the players and moves of this game between designer and user: the payoffs associated with each cell will be derived below.

Figure 1
Domain of the Contingent Renewal Contract Game

User: \ Designer:	Stay	Search
Stay	,	,
Search	,	,
Discontinue	,	,

The bilateral contract form offers a spectrum of possible wage rates and contract intervals. In the spectrum of contract intervals, T=1 corresponds to a spot market in problems;

$T=T_{max}$ is equivalent to lifetime employment. In general, the optimal contract interval (optimal T) will be a function of search, measurement, and other costs of returning to the market or renegotiating the contract's terms.

We examine below the payoffs and equilibria associated with this game. We consider first, the designer's optimal choice of action within the contract interval; and second, the user's financial considerations. We then show that the base-case local system can support a financially sufficient contractual equilbrium. As a result, the local system can be an institution in the sense defined by Aoki:

> An institution is a self-sustaining system of shared beliefs about how the game is played. Its substance is a compressed representation of the salient, invariant features of an equilibrium path, perceived by almost all the agents in the domain as relevant to their own strategic choices. As such it governs the strategic interactions of agents in a self-enforcing manner and in turn is reproduced by their actual choices in a continually changing environment.[15]

3.6 The Designer's Decision Rule in a "Robinson Crusoe" Design Economy

We have assumed that a designer can choose (1) to solve one problem per problem-solving interval, or (2) to apply the textbook solution to some problem. Although there are T problem-solving intervals in each contract interval, the designer can select problems one at a time. An individually-rational, one-step-look-ahead rule for allocating effort is then:

> At the beginning of each problem-solving interval, let each designer review the M known problems, and rank them on the basis of R(1), the expected benefit obtained from *one draw* from the distribution, G(x):
>
> $$R(1) = \int_0^\infty x\, g(x)\, dx \quad ;$$
>
> Without loss of generality, let the first x-scale, x_1, in the designer's utility function correspond to the problem with the highest R(1). Since all problems demand the same

[15] Aoki, 2001, p. 26.

amount of effort, and solutions have no effect on income, a utility-maximizing designer will choose to work on that problem, if any.

Let each designer then compare the utility of working on that problem, with the utility of using the textbook solution there and elsewhere. It follows that the designer will voluntarily work on the problem, if and only if:

$$u(w, R(1), \ldots, 0, e + \Delta e) > u(w, 0, \ldots, 0, e)$$

Otherwise, the designer will expend less effort, and use the textbook solution to solve a randomly selected "problem of the moment." The designer will also be aware of the textbook solutions to the M-1 known problems.

Because we have framed the designer's problem as a discrete choice, the above procedure is a utility-maximizing decision rule as long as the designers are isolated one from another.

Note that for the designers in the base-case local system, extra effort expended on solving problems is rewarded by an increase in the quality of those specific designs. Such design improvements directly increase the designer's utility, hence are an end in themselves. From the designer's point of view, the economic benefits that may accrue to the user and/or society are strictly incidental: they a pure byproduct of the designer's efforts.[16]

3.7 The User's Financial Considerations; Sufficiency

We have said that the "Q(k) design process", the users and the designers and the institutional form of a bilateral, contingent renewal contract constitute "the base-case local system." In a market economy, every local system has a financial "footprint," that is, it has cash inflows and outflows, and corresponding payors and payees. In our base-case local system, the users are the payors and the designers are payees.

We will say that a local system (with attendant institutional forms) is *financially sufficient* (or just sufficient), if the payors' willingness to pay for the products of the system exceeds the

[16] Notice that there is a free lunch here. It arises because designers are both producers and consumers of their designs; and users are consumers of designs. This may be an example of "incentive-enhancing preferences." (Bowles, Gintis, Osbourne, 2001.)

financial outflows from the system. We label the difference between willingness to pay and cash outflows the "net financial value of the local system," NFV_{LS}:

$$NFV_{LS} \equiv \Sigma \text{ Payors' Willingness to Pay} - \Sigma \text{ Cash Outflows from Payors to Payees.}$$

If $NFV_{LS} > 0$ => Local System is financially sufficient.

By definition, financially sufficient local systems are in a *positive financial balance* with the rest of the economy. The payors' willingness to pay exceeds the money flowing out in payment for costly resources. The difference is a local surplus, which accrues to the payors, and justifies the continued existence of the system.

In market economies, which rely on voluntary transactions and contracts, sufficiency is a basic pre-requisite for stability in a local system.[17] By definition, insufficient local systems are either bankrupt or being subsidized. If bankrupt, they cannot participate in market exchanges. Bankrupt local systems must either be reorganized so as to achieve sufficiency, or go out of existence. By the same token, subsidized local systems can continue to exist only so long as the subsidies continue.

Conversely, in market economies, sufficient local systems generally have (and are assumed to have) the right to exist without interference.[18]

The patterns formed by local systems in a large economy may be quite complex. One common pattern is for a smaller local system to be "nested" inside a larger one. This is, in fact, the relationship between our base-case local system and the corresponding user firm. Each designer-user contractual dyad resides within a specific user firm. The user firm in turn is a local system (of linked institutions) that resides in a greater economy.

---

[17] Sufficiency is at once weaker, less ambiguous, and more universal than the better-known economic criterion of "efficiency." Efficiency implies an "optimal" allocation of resources, a maximization of something (see Samuelson, 1970). However, the optimization often lies in the eye of the beholder: A and B value different things, hence would optimize the allocation of resources in different ways.

[18] Some sufficient local systems may be deemed illegal, hence not have the right to exist. For example, Napster was a new and sufficient local system that turned out *not* to have the right to exist! But in a market economy, the presumption is that most sufficient local systems, e.g., contracts, markets, corporations, commercial clusters etc., do have the right to exist. Note that this presumption is itself an institution under Aoki's definition, (quoted above).

In cases of nesting, the test of financial sufficiency must be applied at every level of the hierarchy of institutions. Contracts lodged within bankrupt organizations are not sufficient, hence not stable; organizations lodged withing bankrupt economies are not sufficient, hence not stable.

Formally, we can introduce a “sufficiency constraint” or “sufficiency test” into any economic game by giving those who pay for the system an outside option: to “play” or “not play.” In fact, we have already done this in the contingent contract renewal game (see Figure 1 above): in addition to the strategies “Stay” and “Search”, we have said that the user has the option not to employ a designer over the next contract interval.

A complexity arises, however, because under our assumptions the user must contract for the designer’s services before the fact, but can only measure the designer’s output after the fact. The payoffs in the contracting game are thus *expected payoffs*, which the user and the designer must assess in some fashion. The designer, however, knows the nature of both the technological substrate and his or her own decision rule. The user, in contrast, can only measure (at some cost) what the local system contributed over the past contracting intervals.

We will now prove that the base-case local system can be financially sufficient, according to a user’s *ex post* analysis.

Proposition 1. Consider one designer-user pair as a local system. If the designer applies the individually rational, one-step-look-ahead decision rule described above; if the user-firm can charge its customers their full willingness-to-pay for superior designs; and if the designer’s wage rate ω is such that:

$$\omega < w - m/T \quad ,$$

where m is the user’s measurement cost;[19] then in the user’s *ex post* analysis, the contingent-contract local system will be financially sufficient.

---

[19] We are simplifying here by assuming that the cost of measurement is an exogenous parameter. In reality, measurement systems are institutions in their own right: they have technological substrates and actors, and they require self-sustaining incentives and shared beliefs.

Note, however, that in the base-case model world, the equilibrium measurement system for each designer-user pair is likely to be very imprecise. By assumption, the users will capture the entire financial surplus, thus they do not have incentives to measure the designers’ contributions very precisely. The designers, for

Proof:

The designer's time is the resource being used: it is being paid for at the per-period rate, ω. From the perspective of each designer, ω is a market wage. By assumption it is high enough to allow the designer to live in the greater economy. (Designers are not bankrupt, and their families are not starving.)

Let the contract interval be T. Suppose, for $\tau \leq T$ periods, the designer expends extra effort to seek new solutions to his problems; for T-τ periods, the designer uses the textbook solution, and expends no extra effort. Ignoring the order in which problems are solved, the net financial value of the local system is:

$$NFV_{LS} = \sum_{t=1}^{\tau} Q_t^*(1) + \sum_{t=\tau+1}^{T} v(0) - m - \omega T \quad .$$

Because of the designer's option to reject inferior solutions:

All $Q_t^*(1) \geq v(0) = w$.

Thus by substitution:

$$NFV_{LS} \geq wT - m - \omega T > 0 \quad .$$

The second inequality is strict if $\omega < w - m/T$, as we have assumed. Therefore, for the user, the local system is sufficient in *ex post* analysis: after the fact, it will have generated more financial value than it consumed. **QED**

The premises of Proposition 1 are *sufficient conditions* for the contingent-contract local system to generate a financial surplus for the user-firm. They are not necessary conditions, however. Particular designer-user pairs can make their own local systems financially sufficient in other ways. For example, a designer might bargain for a higher wage rate: $\omega > w - m/T$; and

their part, do not have a methodology by which to measure the financial value of their own problem-solving efforts. Finally, for any one designer-user dyad, the costs of designing a precise and mutually credible measurement system may completely negate the net financial value generated by their contractual arrangement. This is what it means to be in a high-friction, high-transactions-cost, high-information-cost environment!

make it sufficient to her user-firm by delivering a series of easily measured design improvements (Q*'s with low m's).

3.8 The Base-Case Local System Supports an Aoki-type Institution If Users Believe It is Sufficient

We can now specify the payoffs in each cell of the contract renewal game between the user and the designer. The normal form of the game is shown in Figure 2, which should be read as follows:

- The user's choices are to stay with the present designer, search for a new designer, or discontinue the employment of any designer. The user's payoffs are expectations, denominated in dollars over the next contract interval.
- The designer's choices are to stay with the present contract or search. The designer's payoffs are expected utilities over the next contract interval.
- If both stay, neither incurs search costs.
- If the designer searches, then he or she incurs search costs: in expectation, wealth goes down by S, and effort goes up by s, relative to staying with the present contract.
- If the user searches, it incurs a search cost, S, which reduces the expected NFV of the local system; if the user stays, but the designer searches, the user incurs the same search cost. Finally, if the user discontinues the contract, it receives a zero payoff.
- If the user searches or discontinues but the designer stays, the designer is unilaterally terminated. The designer obtains the lowest possible utility, u_{min}, in that case.

In the presence of positive search costs, there are two potential Nash equilibria in this game: Stay-Stay and Discontinue-Search. They are denoted by dark borders around the corresponding cells.

The user's perception of the sufficiency of the local system determines the nature and number of Nash equilibria of this game. If the user believes that the local system will be sufficient, then $E(NFV_{LS}) > 0$, and the unique Nash equilibrium is Stay-Stay. In contrast, if the user believes that the local system will be strictly insufficient, then $E(NFV_{LS}) < 0$, and the unique Nash equilibrium is Discontinue-Search. Finally, if the user believes that $E(NFV_{LS}) = 0$, then there are two Nash equilibria.

Figure 2
Normal Form of the Contingent Renewal Contract Game: Base Case

User:	Designer: Stay	Search
Stay	$E(NFV_{LS})$, $Eu(\omega T, \ldots, e)$	$E(NFV_{LS} - S)$, $Eu(\omega T\text{-}S, \ldots, e+s)$
Search	$E(NFV_{LS} - S)$, u_{min}	$E(NFV_{LS} - S)$, $Eu(\omega T\text{-}S, \ldots, e+s)$
Discontinue	0, u_{min}	0, $Eu(\omega T\text{-}S, \ldots, e+s)$

Thus we have shown that the base-case local system consisting of:

- the Q(k) design process,
- the designers,
- the users, and
- a bilateral contingent renewal contract;

can be sufficient in a Robinson Crusoe design economy. Moreover, if the local system is perceived to be sufficient by the users, it will support a contractual equilibrium between users and designers. Finally each local system that supports an equilibrium can be an institution as defined by Aoki (2001).

Still the Robinson-Crusoe-type design economy, with its associated institutions, is a long way from being "as good as it could be." Indeed, there are two immediate targets of opportunity in the technological substrate. They are:

- pooling designers' solutions;
- coordinating designers' choice of problems.

In the next section, we will show how pooling and coordination unlock value from the technological substrate. Such value may be sufficient to "pay for" the creation of new, specialized institutional forms, whose purpose is to reduce the frictional costs of pooling and coordination.

4 Solving The Problems of the Q(k) Technological Substrate

4.1 Pooling Solutions

Suppose that by some low-cost mechanism designers can publish their solutions and read other designer's solutions to the M known problems. Each designer can then apply, not just his or her own solution, but the *best* solution devised by any designer. Because solutions are non-exclusive information goods, there is potentially a very large social surplus in such a pooling arrangement.

To fix ideas, let the time pointer be set at the end of a problem-solving interval. Assume that, during the previous period, each of the n designers worked on the first of the M problems. Individually they have obtained solutions with varying x-values. Those obtaining x-values less than zero have discarded their solutions in favor of the textbook solution. Now suppose a new institutional form "drops from the sky." It allows all designers to publish their solutions, and to read and use the solutions of others. Participation in this new institutional form is voluntary.

Proposition 2.

a. If all designers worked on the same problem, then an institution that allows designers to publish their solutions increases the *ex post* utility of every designer except one.

b. An institution that allows designers to publish their solutions increases the *ex ante* utility of all designers.

Proof:

2a. A designer, i, who accesses the pool gets a higher utility than one who does not if:

$u(w, R_1^*(n), \ldots, e + \Delta e) > u(w, R_1^*(1), \ldots, e + \Delta e)$.

By definition:

$$R_1^*(n) = \max_n R_1^*(1) \quad .$$

Thus, for n-1 of the designers, participation in the pool strictly increases their *ex post* utility. For one designer, R*(1) = R*(n), and participation in the pool is a break-even proposition.

2b. *Ex ante* a designer does not know where his solution will rank in the pool of all solutions. By the properties of order statistics, the truncated expectation of the best of n solutions is greater than the truncated expectation of one solution:

$$u(w, R_1(n), \ldots, e + \Delta e) > u(w, R_1(1), \ldots, e + \Delta e) \quad .$$

Thus *ex ante*, all designers have something to gain from being in the pool. **QED**

Note 1: Result b implies (under suitable assumptions about utility) that the designers will voluntarily bear some cost in terms of money and/or effort in order to create and maintain the pool. Therefore, if its costs are low enough, the pool will be sufficient!

Note 2: Result b can also serve as the basis of an ongoing relational contract between individual designers and the pool. Even if *ex post* the designer with the highest-value solution only breaks even, in expectation, he/she still benefits from membership.

Note 3: This is a new game between designers and the pool of all other designers. The new game is "orthogonal" to the bilateral contract game previously defined between designers and user-firms.

The pooling of solutions "solves" one of the problems of the technological substrate: how to get the best known solution to each problem used everywhere that it is applicable. Proposition 2 implies that it may be possible for the designers to create and sustain a pool through a self-enforcing relational contract.

The key factors that enable such a contract are:

- design solutions are non-exclusive, i.e., their use in one place does not preclude their use in another;
- designers individually value better solutions regardless of their authorship; and
- *ex ante,* designers do not know how good their future solutions will be relative to those of others in the pool.

If any of these three conditions is violated, the basis for a voluntary, self-enforcing, relational contract will disappear.

The existence of a designers' pool greatly increases the economic value of designers as a class. In the base case, the n user-firms "captured" the net financial value of the designs, and the NFV of each dyadic local system was:

$$NFV_{LS}(BC) = \left[\sum_{t=1}^{\tau} Q_t^*(1) + \sum_{t=\tau+1}^{T} v(0) - \omega T \right] > 0 \quad .$$

(where BC stands for Base Case).

In contrast, if there is a pool, and the user-firms continue to capture the net financial value of the designs, then:

$$NFV_{LS}(P) = \left[\sum_{t=1}^{\tau} Q_t^*(n) + \sum_{t=\tau+1}^{T} v(0) - \omega T \right],$$

$$> \left[\sum_{t=1}^{\tau} Q_t^*(1) + \sum_{t=\tau+1}^{T} v(0) - \omega T \right]$$

$$> 0 \quad .$$

Here P stands for Pool.

By definition, for each problem:

$$Q_1^*(n) = \max_n Q_1^*(1) \quad ;$$

and, almost certainly, different designers would have been responsible for the best solution in different problem-solving intervals (recall that the contract interval, T, spans several problem-solving intervals). Therefore, every user-firm is better off under the pooled arrangement than under the Robinson-Crusoe base case.

4.2 The Potential for a "Dissipated " Surplus

We must acknowledge a subtlety here, however. In the base case, users *applied different solutions* to their instances of particular problems. Many would have used the textbook solution; others would have used new solutions devised by their own designers. In the pooled arrangement, however, users all apply the same solution to each instance of a particular problem.

Competition thrives on differences. In the Robinson Crusoe design economy, users with better solutions might have charged higher prices for their products. "Rents" on superior designs would then have been captured by those user-firms. In contrast, in the pooled-solution design economy, there are no differences in the quality of solutions across user-firms. As a result, superior designs cannot be a source of competitive advantage, and therefore user-firms may not be able to capture rents from them.

The financial value of the superior designs is not lost, however: it is passed through to the end-users, i.e., the customers of the user-firms, as a consumer surplus.

Even if the design surplus is captured by end-users, however, the bilateral contract between user-firms and designers remains sufficient. No user-firm can afford not to have a designer: in fact, to be competitive, each must have a designer who is "in the pool." Thus, in the pooled-solution design economy, designers' services may become a "cost of doing business," but not a source of competitive advantage for the user-firms.

This example, which we label a "dissipated surplus," illustrates some of the complexities caused by nested local systems. A dissipated surplus can be an enticing target of opportunity for a rent-seeking union or monopolist. However, such rent-seekers will need to invent new, specialized institutional forms to enable them to capture the dissipated surplus, and capitalize on the net financial value of the pooled designs.

4.3 Allocating Designers' Effort

In the absence of a means of pooling solutions, there is no value to be gained by coordinating designers' efforts. Absent a pool, each designer leads a "Robinson Crusoe" type of existence: the only way to obtain a better solution is to devise it oneself. In the presence of a pool, however, the uncoordinated choice of target problems is very inefficient. For all probability distributions, the order statistic functions, R(k) and Q(k), exhibit diminishing returns to k, the measure of designer effort.[20] Thus if all designers *ex ante* choose to work on the same problems, in aggregate, they will "overinvest" their effort in some problems, and "underinvest" in others.

How much effort on a given problem is too much? This turns out to be a tricky question. Let us approach it first from a designer's point of view. Recall that there are M known problem types, each of which has n instances for a total of Mn problems. If all designers work on the same problem type (e.g., the first) then the expected value obtained is $R_1(n)$. Each designers' incremental effort is Δe. However, suppose for the last, i.e. nth, designer:

$$u(w, R_1(n-1), \ldots, e) \quad (1)$$
$$> \quad u(w, R_1(n), \ldots, e + \Delta e)$$
$$> \quad u(w, R_1(1), 0, \ldots, e + \Delta e) \quad .$$

From the perspective of the nth designer, his or her incremental contribution to the pooled solution does not justify his or her incremental effort. The nth designer should then choose not to work. This looks very much like a classic free rider problem!

However, the designer who chooses NOT to work on the first (most salient) problem, may choose to work on a different problem (say the second). In fact, if:

$$u(w, R_1(n-1), R_2(1), \ldots, e + \Delta e) \quad (2)$$
$$> \quad u(w, R_1(n-1), 0, \ldots, e) \quad ,$$

then, by the same argument as was used in Section 3, the designer will voluntarily elect to increase his or her effort in hopes of devising a better solution to the second problem.

The second problem in turn becomes a candidate for the pooling of solutions. And, if too much effort is expended on that problem, some designers may choose to work on the third, fourth, fifth or less important problems.

---

[20] Lindgren, 1968.

Expressions (1) and (2) are calculations that rational designers might make in the presence of a pool. Interestingly, therefore, although a pool makes all designers better off, it also "disequilibrates" the base-case local system. Specifically the individually rational, one-step-look-ahead rule by which designers chose their target problems in the Robinson-Crusoe context, is not individually rational in the presence of a pool. And, as if that were not enough, the pool creates the potential for free-riding of some designers on the solutions of others.

Thus even as the pool solves one set of problems in the technological substrate (how to get the best solution applied to all instances of a problem), it creates others (how to coordinate designers' choices of what problems to work on; how to manage free-riding).

5 Summary: The Technology of Design and its Problems

Let us now pause and summarize the generic problems posed by the Q(k) technological substrate. As we have seen, the substrate "needs" institutional forms that provide:

- a method of assessing the benefits of solving problems in the eyes of designers and in terms of value in the greater economy: R(1) and Q(1));
- a method of assessing benefits from parallel, independent efforts: R(k) and Q(k) for k > 1;
- a method of publishing and transferring solutions (a means of *ex post* pooling); and
- a method of coordinating designers' choices of target problems, or assigning designers to work on specific problems.

Institutional forms that address these problems have the potential to increase:

- the designers' individual utilities;
- the Net Financial Value (NFV_{LS}) of the local system; and/or
- the "dissipated surplus" of end-users of the designs.

Such forms will be useful: they will increase the value to society of designs and designers as a class, as well as the utilities of individual designers. Hence, such forms may be able to pay for themselves out of the surpluses they generate.

The institutional forms named at the beginning of this document—the guild, the monopoly, and the cluster — each "solve" the problems of the Q(k) substrate in different ways. How these forms work, the nature of the equilibria they support, and how they interact or compete withone another will be the focus of our analysis in the remaining papers of this series.

References

Aoki, Mashiko (1999) "Information and Governance in the Silicon Valley Model," Stanford University, http://wwwecon.stanford.edu/faculty/workp/swp99028.html, viewed Jan 12, 2001.

_______ (2001). *Towards a Comparative Institutional Analysis*, MIT Press, Cambridge, MA.

Baldwin, Carliss Y. and Kim B. Clark (2000). *Design Rules, Volume 1, The Power of Modularity*, MIT Press, Cambridge MA.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, Barry Weingast (1998). "Introduction" and "Conclusion," *Analytic Narratives*, Princeton University Press, Princeton, NJ.

Bowles, Samuel, Herbert Gintis and Melissa Osborne (2001). "The Determinants of Earnings: A Behavioral Approach, *Journal of Economic Literature* XXXIX (December): 1137-1176.

Braudel, Fernand (1982). *The Wheels of Commerce*, (tr. Sian Reynolds) Harper & Row, New York, NY.

Dawkins, Richard (1989). *The Selfish Gene, 2nd ed.* Oxford University Press, Oxford UK.

Gintis, Herbert (1976). "The Nature of Labor Exchange and the Theory of Capitalist Production," *Review of Radical Political Economy*, 8(2): 36-54.

________ (2000). *Game Theory Evolving*, Princeton University Press, Princeton, NJ.

Greif, Avner (1994). "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy* 102: 912-50.

________ (1998a). "Historical and Comparative Institutional Analysis," *American Economic Review*, 88 (May): 80-84.

________ (1998b). "Self-Enforcing Political Systems and Economic Growth: Late Medieval Genoa," in *Analytic Narratives*, (ed. Bates *et al*) Princeton University Press, Princeton, NJ.

Lumsden, Charles J. and Edward O. Wilson (1981), Genes, Mind and Culture, Harvard University Press, Cambridge, MA.

Maynard Smith, John (1982). *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, UK.

Maynard Smith, J. and G. Price (1973). "The Logic of Animal Conflict," *Nature, London*, 246: 15-18.

Merton, Robert C. (1973) "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4(Spring): 141-183; reprinted in *Continuous Time Finance*, Basil Blackwell, Oxford, UK, 1990.

_________ and Zvi Bodie (1995). "A Conceptual Framework for Analyzing the Financial Environment," in *The Global Financial System: A Functional Perspective*, (ed. Crane *et al*) Harvard Business School Press, Boston, MA.

North, Douglass C. (1990) *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge, UK.

Samuelson, Larry (1997) *Evolutionary Games and Equilibrium Selection*, MIT Press, Cambridge, MA.

Samuelson, Paul (1970). "Maximum Principles in Analytical Economics," Nobel Memorial Lecture, December 11, 1970. Reprinted in *Nobel Lectures: Economic Sciences 1969-1980* (ed. Assar Lindbeck) World Scientific, London, UK.

Shapiro, Carl, and Joseph Stiglitz (1984). "Equilibrium Unemployment as a Discipline Device," *American Economic Review*, 74 (3): 433-44.

Williamson, Oliver E. (1985), "The Economic Institutions of Capitalism," Free Press, New York, NY.