

The Theory of the Firm

MICROECONOMICS WITH ENDOGENOUS
ENTREPRENEURS, FIRMS, MARKETS,
AND ORGANIZATIONS



Daniel F. Spulber

CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521517386

This page intentionally left blank

THE THEORY OF THE FIRM: MICROECONOMICS WITH ENDOGENOUS ENTREPRENEURS, FIRMS, MARKETS, AND ORGANIZATIONS

The Theory of the Firm presents a path-breaking general framework for understanding the economics of the firm. The book addresses why firms exist, how firms are established, and what contributions firms make to the economy. The book presents a new theoretical analysis of the foundations of microeconomics that makes institutions endogenous. Entrepreneurs play a central economic role by establishing firms. In turn, firms create and operate markets and organizations. The book provides innovative models of economic equilibrium that endogenously determine the structure and function of economic institutions. The book proposes an “intermediation hypothesis” – the establishment of firms depends on the effects of transaction costs and on the extent of the market.

Daniel F. Spulber is the Elinor Hobbs Distinguished Professor of International Business and Professor of Management Strategy at Northwestern University’s Kellogg School of Management and the founder of Kellogg’s International Business and Markets Program. He is the founding editor of the *Journal of Economics and Management Strategy*. His current research is in the areas of entrepreneurship, international economics, economics of organizations, industrial organization, management strategy, and law. Spulber is the author of 11 books, including *Networks in Telecommunications: Economics and Law* (with Christopher Yoo, 2009), *Global Competitive Strategy* (2007), *Market Microstructure: Intermediaries and the Theory of the Firm* (1999), and *Deregulatory Takings and the Regulatory Contract: The Competitive Transformation of Network Industries in the United States* (with J. Gregory Sidak, 1997), all from Cambridge University Press, and *Management Strategy* (2004), *The Market Makers* (1998), and *Regulation and Markets* (1989), from other publishers.

The Theory of the Firm

Microeconomics with Endogenous
Entrepreneurs, Firms, Markets, and
Organizations

Daniel F. Spulber

Northwestern University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521517386

© Daniel F. Spulber 2009

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2007

ISBN-13 978-0-511-53076-0 eBook (Adobe Reader)

ISBN-13 978-0-521-51738-6 hardback

ISBN-13 978-0-521-73660-2 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>Preface and Acknowledgments</i>	<i>page ix</i>
Introduction	1
PART I. THE THEORY OF THE FIRM	
1 The Consumer	11
1.1 The Intermediation Hypothesis and the Scope of the Firm	12
1.2 Consumer Characteristics	26
1.3 Consumer Cooperation and Transaction Benefits	28
1.4 Consumer Coordination and Transaction Costs	35
1.5 Consumer Organizations and the Separation Criterion	40
1.6 Conclusions	61
2 The Firm	63
2.1 The Separation Criterion	64
2.2 Firms Create and Manage Markets	76
2.3 Firms Create and Manage Organizations	88
2.4 The Development of the Firm	102
2.5 The Social, Legal, and Political Context of the Firm	117
2.6 Conclusions	123
3 The Separation of Consumer Objectives and Firm Objectives	125
3.1 The Neoclassical Separation Theorem	127
3.2 The Separation Theorem with Price-Setting Firms	132
3.3 The Fisher Separation Theorem	136
3.4 The Fisher Separation Theorem with Price-Setting Firms	142
3.5 Conclusions	147

PART II. THE ENTREPRENEUR IN EQUILIBRIUM

4	The Entrepreneur	151
4.1	A Dynamic Theory of the Entrepreneur	153
4.2	The Entrepreneur's Decisions and the Foundational Shift	163
4.3	Type I Competition: Competition among Entrepreneurs	175
4.4	Type II Competition: Competition between Entrepreneurs and Consumer Organizations	177
4.5	Type III Competition: Competition between Entrepreneurs and Established Firms	179
4.6	The Classical Theory of the Entrepreneur	186
4.7	Conclusions	196
5	Competition among Entrepreneurs	197
5.1	Set-Up Costs	198
5.2	Investment	205
5.3	Time Preferences	212
5.4	Risky Projects	214
5.5	Risk Aversion	219
5.6	Endowments and Incentives for Effort	223
5.7	Conclusions	226

PART III. HUMAN CAPITAL, FINANCIAL CAPITAL, AND THE ORGANIZATION OF THE FIRM

6	Human Capital and the Organization of the Firm	231
6.1	The Worker Cooperative versus the Firm	232
6.2	Hiring Workers with Diverse Abilities	237
6.3	Hiring with Moral Hazard in Teams	242
6.4	Market Contracts versus Organizational Contracts	247
6.5	Hiring by the Firm versus a Cooperative with Open Membership	256
6.6	Conclusions	262
7	Financial Capital and the Organization of the Firm	263
7.1	The Basic Model	264
7.2	The Corporation in Equilibrium	268
7.3	The Partnership in Equilibrium	276
7.4	Market Equilibrium and Organizational Form	283
7.5	Conclusions	290

PART IV. INTERMEDIATION BY THE FIRM

8	The Firm as Intermediary in the Pure-Exchange Economy	295
8.1	The Firm and Money	296
8.2	The Firm and the Absence of a Double Coincidence of Wants	302
8.3	The Firm and Market Clearing	308
8.4	The Firm and Time	310
8.5	The Firm and Distance	315
8.6	The Firm and Risk	322
8.7	Conclusions	328
9	The Firm versus Free Riding	329
9.1	The Firm and Economies of Scale	331
9.2	The Firm and Public Goods	347
9.3	The Firm and Common-Property Resources	355
9.4	Conclusions	362

PART V. MARKET MAKING BY THE FIRM

10	The Firm Creates Markets	367
10.1	Market Making and Matchmaking by the Firm: Overview	368
10.2	Market Making by the Firm versus Consumer Search	380
10.3	Matchmaking by the Firm versus Consumer Search	388
10.4	Competition between Market-Making Firms	391
10.5	Competition between Market-Making Firms: Characterization of Equilibrium	404
10.6	Conclusions	415
11	The Firm in the Market for Contracts	417
11.1	Contracts and Incentives to Invest: Firms Create Markets	419
11.2	Moral Hazard: Firm Management of Tournaments versus Bilateral Agency Contracts	440
11.3	Adverse Selection: The Firm Monitors Performance	447
11.4	Adverse Selection: The Firm Certifies Quality	451
11.5	Conclusions	456
12	Conclusion	458
12.1	The Firm	458
12.2	The Entrepreneur	460

12.3	The Intermediation Hypothesis	462
12.4	Markets and Organizations	463
<i>References</i>		465
<i>Author Index</i>		503
<i>Subject Index</i>		511

Preface and Acknowledgments

This book presents a general theory of the firm. *The Theory of the Firm* seeks to explain (1) why firms exist, (2) how firms are established, and (3) what firms contribute to the economy. The book addresses the foundations of microeconomics by making institutions endogenous. In the models presented in the book, the following are endogenous: entrepreneurs, firms, markets, and organizations.

The general theory of the firm begins with the individual consumer. The characteristics of consumers are the theory's exogenous data. Consumers can do practically anything without firms. Consumers can produce goods and services by operating technology. Consumers can transact directly with each other through bilateral exchange. Finally, consumers can form organizations such as clubs, buyers' cooperatives, workers' cooperatives, and basic partnerships.

The firm is an economic institution that differs fundamentally from a consumer organization. This book introduces a new definition of the firm that is highly useful in developing the theory: The firm is a transaction institution whose objectives are separate from those of its owners. Consumer organizations such as clubs and basic partnerships are not firms. The objectives of consumer organizations cannot be separated from those of their owners.

Why do firms exist? *The Theory of the Firm* shows that firms exist only when they improve the efficiency of economic transactions. The efficiency of firms is compared to the alternative of direct exchange between consumers. Direct exchange between consumers involves search, bargaining, barter, and contracts. Direct exchange between consumers also can involve forming consumer organizations. To be economically viable, firms must improve on the efficiency of what consumers can achieve without firms.

How are firms established? Individual consumers can choose to become entrepreneurs and establish firms. *The Theory of the Firm* thus makes the entrepreneur endogenous in microeconomics. Because entrepreneurs establish firms, the firm also is endogenous in microeconomics. Entrepreneurs and firms arise based on the underlying characteristics of consumers who possess the judgment, knowledge, skills, and technology that are needed to set up a firm. Individuals

provide the effort, investment, and planning that are needed to start up a business. If firms will enhance economic efficiency, entrepreneurs can earn a return from establishing a firm.

What do firms contribute to the economy? Firms are institutions that coordinate transactions by acting as intermediaries. Among the many instruments that firms use to coordinate transactions are two major ones. First, firms intermediate exchange by creating and operating markets. This makes markets endogenous in the theory of the firm. Firms create markets by marketing and selling goods and services, by setting up facilities such as stores and Web sites, and by arranging exchanges for commodities and financial assets. Firms adjust prices to balance their purchases and sales and thereby clear markets. Second, firms create and manage organizations that employ personnel and financial capital; intermediate transactions; internally allocate capital, labor, and resources, and carry out production. This makes organizations endogenous in the theory of the firm.

The theory of the firm constitutes a unified field with its own set of questions. The analysis departs from the neoclassical general equilibrium framework that takes both firms and markets as given exogenously and that does not consider either entrepreneurs or organizations. The theory of the firm incorporates advances in the study of firms from industrial organization, contract theory, game theory, law and economics, institutional economics, the economics of organizations, and finance.

The general theory of the firm is not based on a specific “silver bullet” theory of why firms exist. The general theory of the firm includes the full range of transaction costs, including the absence of a double coincidence of wants, communication costs, search costs, bargaining costs, moral hazard, adverse selection, contracting costs, and free riding.

Microeconomics seeks to address the purpose and functions of firms, markets, and organizations. Understanding why firms exist, how firms are established, and what firms contribute to the economy is essential to this task. The framework develops some critical empirical implications that require further investigation. In addition, the general theory of the firm helps to understand management decision making. The field of management strategy seeks to develop policies for managers, which require a framework that can evaluate the effectiveness of alternative strategies.

A general theory of the firm also is useful for teaching economics. Economics courses, including principles of economics, intermediate microeconomics, and graduate microeconomics, rarely mention entrepreneurship. In the neoclassical economics course, firms and markets are given exogenously. Firms lack an explicit organizational structure and are fully described by their production technology. Markets are operated by an invisible hand. Students are often perplexed, because firms are said to be price-takers and yet, at the same time, firms often are said to adjust prices in response to surpluses or shortages, an obvious contradiction to price-taking behavior. The theory of the firm contributes to teaching economics

by introducing a more complete picture of the economy. The contributions of entrepreneurs often provide appealing narratives. Discussion of eBay's Internet auctions or the now publicly traded New York Stock Exchange yields insights into the market-making activities of firms.

The book is organized as follows. Part I of the book provides the foundation for the endogeneity of the firm. Chapter 1 provides the exogenous preconditions for the theory of the firm by defining the consumer. The characteristics, endowments, and transaction costs encountered by the consumer form the basis for the endogenous decisions of consumers to become entrepreneurs and establish firms. The chapter defines direct exchange between consumers and also explains why consumer organizations are not firms. Chapter 2 explores the formal definition of the firm, examines the separation criterion, and introduces the intermediation hypothesis. Chapter 3 introduces the separation theorems, which explain the separation of consumer decisions from those of the firm. The chapter extends the neoclassical and Fisher separation theorems to a model with oligopoly competition between price-setting firms.

Part II of the book introduces the entrepreneur as the central figure in microeconomics. Chapter 4 presents a formal definition of the entrepreneur and reviews the literature and historical context of entrepreneurship. The discussion highlights the critical importance of the entrepreneur in the economy and emphasizes the role of the entrepreneur in establishing firms. The chapter identifies three types of competition faced by the entrepreneur: competition among entrepreneurs, competition between the entrepreneur and direct exchange between consumers, and competition between the entrepreneur and established firms. Chapter 5 presents a set of models in which entrepreneurs establish firms in economic equilibrium. Entrepreneurs compete to establish firms, with various factors determining the number of entrepreneurs, including such factors as set-up costs, rates of time preference, risk aversion, and wealth.

Part III of the book considers the role of the firm in obtaining human capital and finance capital. Chapter 6 contrasts management of the firm with worker cooperatives and examines the implications of human capital for size and structure of organizations. Chapter 7 considers how financing the firm's capital investment affects the organization of the firm and compares sole proprietorships, partnerships, and corporations.

Part IV of the book develops the economic role of the firm as an intermediary. In Chapter 8, the firm alleviates the absence of a double coincidence of wants and provides a substitute for money. The absence of a double coincidence of wants is examined in the context of transportation and travel costs, allocation over time, and uncertainty. In Chapter 9, the firm addresses the free rider problem when joint production involves economies of scale, public goods, or common property resources.

Part V of the book considers the economic contribution of the firm as a market maker. In Chapter 10, the firm acts as a market maker and a matchmaker in markets

Introduction

The purpose of this book is to present a general theory of the firm. The theory provides a microeconomic framework in which entrepreneurs, firms, markets, and organizations are endogenous. The models help to explain why firms exist, how firms are established, and what firms contribute to the economy. Because firms create and operate markets, *The Theory of the Firm* helps to explain how markets arise and how they work, and provides a basic analysis of the formation and design of organizations.

The structure of the theory of the firm is as follows. Consumers and their preferences, endowments, and intellectual property, are exogenous to the model. Consumers choose to become entrepreneurs by working to establish firms, which makes entrepreneurs endogenous. Through the actions of entrepreneurs, firms are established endogenously. Firms act as market makers by creating and operating markets, so that markets also are endogenous. Firms also create and manage organizations that transact internally and in the marketplace, making organizations endogenous. Economic equilibria are the result of consumer-entrepreneurs who establish firms and, in turn, of firms that create and manage markets and organizations. Firms, markets, and transactions are the results of economic equilibria. This framework is summarized in Figure I.1.

Economics is a social science. Economic relationships between individuals in society are its essential elements. Accordingly, the theory of the firm rests upon the characteristics and actions of individual consumers and their relationships. Acting together they form families, communities, social institutions, and governments. Consumers carry out market transactions through direct exchange. Consumers generate a wide variety of organizations that are alternatives to firms.

The critical first step in the theory of the firm is to begin with the individual consumer – and without firms or markets. Consumers have preferences over consumption bundles. They have initial endowments of factors of production and of goods and services. Consumers also have technological knowledge regarding production processes, product designs, transaction mechanisms, and organizational

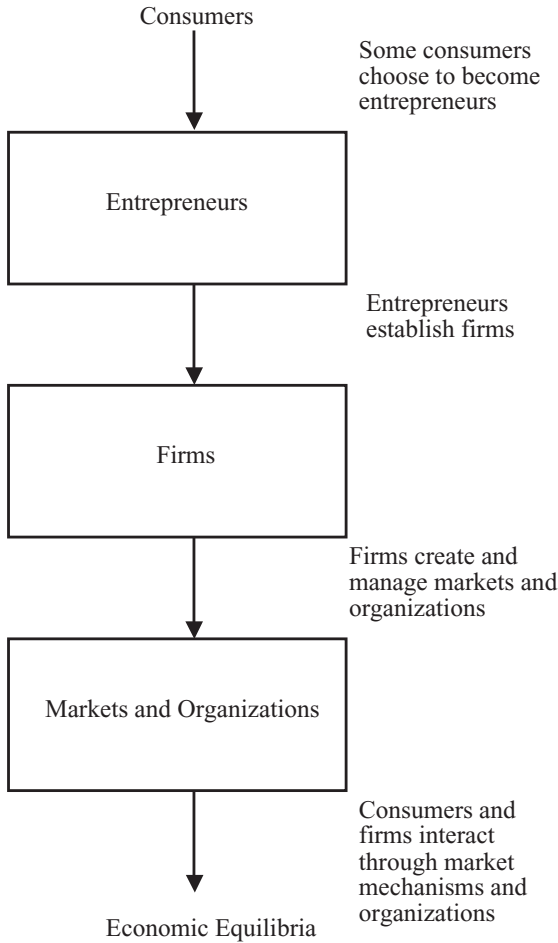


Figure I.1. Microeconomics with endogenous entrepreneurs, firms, markets and organizations.

management. They make purchasing decisions and labor supply decisions. Consumers are inventors, artists, managers, investors, and property owners. Consumers can act as producers and operate manufacturing technology, either individually or collectively. They are capable of technological and scientific invention, and they can commercialize their inventions.

Consumers also are able to set up and manage organizations. They can obtain benefits by transacting with one another through barter and bilateral contracts. They can form clubs, interest groups, and other associations. Consumers can devise more complex economic institutions, such as workers' cooperatives, consumers' cooperatives, and basic partnerships. Consumers can choose to form all kinds of social institutions to facilitate their economic interaction, and they can carry out economic transactions without the need for firms.

Although consumers realize gains from trade through direct exchange and cooperative agreements, they also encounter transaction costs. Such transaction

costs include the time and effort required to search for trading partners. Consumers must communicate with one another, calculate the benefits from trade, negotiate the terms of exchange, design contracts, and observe the performance of contract terms by their partners. Consumers' net gains from trade will depend on how efficient direct transaction methods are in achieving gains from trade and how costly those transaction methods are.

Consumers, acting as entrepreneurs, choose to establish firms when doing so improves economic efficiency. An entrepreneur spends time, effort, and resources to establish a firm in order to receive the returns to ownership once the firm is established. The value generated by the firm must be greater than the costs of establishing the firm.

Transaction costs help to explain why consumers need firms and markets. A firm is a particular type of social institution that can improve the efficiency of transactions. Firms enhance net gains from trade by offering transaction methods not available to individual consumers or even to groups of consumers. The firm achieves transaction efficiencies by creating markets and organizations. To be economically viable, a firm must increase net gains from trade in comparison with direct exchange between consumers. Although consumers may develop and own such transaction technologies, they must establish a firm to implement them. The firm is an instrument for carrying out transactions.

A critical source of the firm's transaction efficiencies is that its objectives differ from those of its owners. The objectives of the firm are distinct from those of its customers, suppliers, managers, or employees. After establishing the firm, the entrepreneur becomes an owner. Therefore, the objectives of the firm are also separate from those of the entrepreneur who established it.¹ The separation of the entrepreneur and the firm allows for a development of a theory of the entrepreneur.

Separation helps a firm to improve the efficiency of transactions in comparison to direct exchange between consumers. By extending the neoclassical separation theorem, separation implies that the firm maximizes profits. Separation allows the firm to be an independent economic actor and decision maker. The firm is an additional player in the economic game. By adding a new player, an entrepreneur gives the economy additional degrees of freedom and new economic instruments. As an additional economic player, the firm performs economic functions not available to consumers. The firm can handle multilateral transactions simultaneously and thus improve efficiency relative to bilateral transactions between consumers.

1 Richard Cyert and James March (1963, p. 10) once observed with regard to the neoclassical paradigm, "If we take seriously the concept of a firm as something distinct from an individual entrepreneur, there is no consensus on a theory of the firm." Cyert and March (1963), in their behavioral approach to the firm, criticized the basic profit maximization framework of neoclassical economics, which in their view was based on the notion that the entrepreneur and the firm were one and the same. By separating the entrepreneur and the firm, the analysis presented here provides a theory of the firm that is consistent with optimization by rational economic agents.

The firm is much more than a nexus of contracts. It is an autonomous player that acts as the counterparty in contracts with customers, suppliers, investors, and employees. The firm enters into many contracts simultaneously as a central player. By acting as a contracting hub, the firm can achieve more than would a complete set of contracts directly linking its trading partners. The firm is a market maker, with instruments such as posted prices and auctions that can aggregate and balance supply and demand. The firm is a matchmaker that can coordinate and connect buyers and sellers. The firm is a centralized clearinghouse that can aggregate transactions and process information.

Because the firm's objectives are separate from those of its owners, it offers incentive mechanisms without the budget-balancing requirement of a buyers' cooperative, a workers' cooperative, or a business partnership. Budget-balancing requirements limit the efficiency of incentive mechanisms. The firm's ability to earn a profit enhances its ability to design incentive contracts for consumers, suppliers, and employees. The firm manages an organization with internal transactions that motivate managers and employees.

Because its objectives are separate from those of its owners, the firm offers other advantages. The firm has longevity, with a lifetime that stretches beyond its particular economic relationships and exceeds the lifetimes of its trading partners. The firm also can transcend geographic limits on its trading partners by operating in a convenient central location or by operating simultaneously in multiple locations. The firm has a brand identity and business reputation beyond those of its individual owners, managers, and employees.

The firm, in the form of a corporation or complex partnership, allows its investors to have both limited liability and liquidity. Investors in corporations are residual claimants. The analysis shows that corporations can provide incentives for managers that cannot be achieved within a basic partnership. Investors can withdraw their capital by selling their shares without disrupting the corporation's business and without selling the company to realize the value of its assets.

The theory of the firm as presented here provides an important insight about economic institutions. In a variety of economic settings, the theory of the firm yields the following result. The establishment of complex economic institutions depends on the extent of the market. The greater the extent of the market, the more economic institutions such as firms, markets, and organizations are established. This recalls Adam Smith's observation that the realization of economies of scale depends on the extent of the market. The greater the extent of the market, the greater the contribution of firms in improving transaction efficiencies. This suggests why the establishment of firms, markets, and organizations is associated with economic growth, economic development, and international trade.

A key variable repeatedly emerges in a variety of different contexts – the number of consumers. When there are few consumers, direct exchange may be the most efficient economic process. Consumer transactions and consumer organizations are most efficient. When there are many consumers, firms intermediating exchange

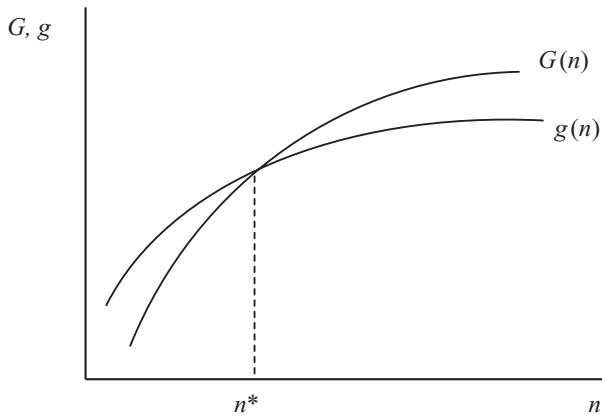


Figure I.2. Entrepreneurs establish firms when the number of consumers is sufficiently greater than n^* . The establishment of complex economic institutions depends on the extent of the market.

becomes the most efficient way to organize the economy. Establishing firms and creating markets and organizations provide economic efficiency relative to direct exchange. The greater the size of the economy, in terms of the number of consumers, the greater are the efficiencies that can be realized by establishing firms.

When there are many consumers, direct exchange may not perform as effectively as intermediated exchange due to transaction costs associated with search, communication, and bargaining, and due to such problems as free riding. As the number of consumers increases, the economy realizes benefits from setting up firms, which are due to economies of scale in transaction technologies. The economy also realizes returns to centralization of exchange through the use of market mechanisms and organizations.

In many models, there is a critical size at which the economic equilibrium switches from direct exchange between consumers to intermediated exchange through firms. It is useful to illustrate the effects of the extent of the market on economic institutions. Let $g(n)$ represent the consumers' total net gains from trade with direct exchange when there are n consumers in the economy. Let $G(n)$ represent the total net gains from trade when there is intermediated exchange through firms. As the number of consumers increases, net gains from trade with direct exchange may increase more slowly than net gains from trade with intermediated exchange through firms. The effects of the extent of the market are illustrated in Figure I.2. Entrepreneurs become active when the size of the economy is greater than n^* . When there are costs to establishing firms, there must be a sufficient increase in net gains from trade relative to direct exchange.

The importance of the size of the economy for the formation of firms has not been recognized elsewhere. The main point is that direct exchange between consumers through decentralized institutions may be preferable with a small number of people. Intermediated exchange through firms becomes preferable when there is

a large number of people. One forerunner of the idea that transaction costs depend on the number of people is Ronald Coase's (1960) discussion of externalities in the classic article "The Problem of Social Cost." Coase observes that with a smoke nuisance "a large number of people are involved and . . . therefore the costs of handling the problem through the market or the firm may be high." Coase declines to suggest whether markets or administrative decisions within firms are better at handling many transactions.

The Theory of the Firm explores the great variety of economic conditions under which the establishment of firms may or may not be worthwhile. The theory provides a unified framework for studying firms and markets that attempts to integrate much of the previous work on the firm. Within a general framework, different reasons for firms to undertake an activity will be valid depending on economic conditions. Some explanations may have greater resonance than others in empirical applications. All of these explanations are highly valuable for organizing our thoughts about firms. The theory of the firm should not advocate any particular explanation for the boundaries of the firm. Rather, it provides a set of methods that can be used to compare and evaluate the various explanations for firms' activities.

The Theory of the Firm introduces a general analysis of the entrepreneur. The analysis builds on the classic work of Richard Cantillon, Jean-Baptiste Say, Frank Knight, and Joseph Schumpeter. Knight (1971) emphasizes the essential role of the firm as a means of handling incentives. Entrepreneurs exercise judgment in the face of uncertainty and earn profit as a residual return that depends on the uncertain outcome.

The role of the firm in creating and operating markets is central. The firm is an intermediary that reduces the costs of market transactions; see Daniel Spulber (1996a, 1996b, 1999). The firm is instrumental in solving problems due to the absence of a double coincidence of wants. It allows consumers to engage in risk sharing. The firm coordinates exchange over time and provides services that substitute for money in an economy without fiat money or without a store of value. The firm acts as a matchmaker to alleviate the costs of search. The firm sets up markets to improve coordination between buyers and sellers. The firm creates markets for contracts that improve efficiency in comparison to bilateral contracting.

In the theory of the firm, firms also coordinate transactions within organizations. Governance within the firm provides an alternative to market contracts. Ronald Coase (1937, 1988, 1994) explains the firm as an oasis of planning and organizational administration that is an alternative to transactions in the market. Firms choose whether to buy an input and thereby incur a transaction cost or to make the input in-house, thereby benefiting from management and control of activities within the organization. Coase's insights, particularly the introduction of transaction costs into economics, are reflected throughout the discussion.

Firms manage investments in risky projects. A potential advantage of firms over direct exchange is that firms are mechanisms for separating ownership and

control. The discussion includes the classic work on the theory of the firm due to Michael Jensen and William Meckling (1976). They characterize the public corporation as a “nexus for a complex set of voluntary contracts among customers, workers, managers, and the suppliers of materials, capital, and risk bearing.” In their view the firm is a mechanism for allocating control when there is economic risk. “Corporations, like all organizations, vest control rights in the constituency bearing the residual risk” (Jensen 2000, p. 1).

The discussion examines the financial structure of the firm and compares the sole proprietorship, the partnership, and the corporation. Research on the corporation in economics and finance tends to emphasize moral hazard and adverse selection problems that result from delegating authority to the CEO. In the traditional framework, the corporation’s investors are principals and must design incentives for the manager, who acts as their agent. The financial literature examines the inefficiencies that result from the separation of ownership and control using the benchmark of first-best efficiency. In contrast, it is suggested here that the shortcomings of the corporation should be measured against the yardstick of the best organizational alternative. The organizational alternative that combines ownership and control is the partnership. The discussion gives conditions under which either the partnership or the corporation is more efficient, based on Pobleto and Spulber (2008b).

An important task of the firm is the employment and management of labor. Firms manage production operations whether they are used to manufacture goods and services or to generate transactions. The role of the firm is examined in comparison to that of the labor-managed organization. The firm chooses employment at the margin to maximize profit. In contrast, the labor-managed firm chooses employment to maximize the average benefit of its members. The firm provides efficiencies as a consequence of profit maximization.

The general theory of the firm differs from neoclassical economics in a number of critical areas.² In the general theory of the firm, the entrepreneur is the central player because the entrepreneur establishes the firm. In neoclassical economics, there are no entrepreneurs because firms already are established.

In the general theory of the firm, the firm is responsible for intermediating transactions. In contrast, neoclassical economics limits the function of the firm to that of being a producer that operates a technology. Arrow and Hahn (1971, p. 17),

2 The classical economists include Adam Smith, David Ricardo, and John Stuart Mill. See Sowell (2006) for a highly illuminating overview and discussion of classical economics. By neoclassical economics, I refer to economics since the marginalist revolution, including the work of Léon Walras, William Stanley Jevons, and Carl Menger. I also use the term neoclassical economics as shorthand for the general equilibrium model that was developed by Kenneth Arrow, Gerard Debreu, Robert Hahn, and many others. The neoclassical theory of value refers to the concept that the value of goods and services is determined by a market equilibrium in a general equilibrium setting based on consumer preferences, initial consumer endowments, and producer technologies.

for example, observe that households and firms “are distinguished by the property that firms do, and households do not, take production decisions.” Firms take prices as given and make no organizational, management, purchasing, or marketing decisions. Firms do little more than choose the input-output mix that maximizes profits.

The general theory of the firm is founded on transaction costs. In contrast, neoclassical economics is without frictions because markets are established and operate without transaction costs. In the neoclassical general equilibrium framework, a fictitious Walrasian auctioneer clears markets by exogenously selecting prices. Neoclassical economics presents a theory of value that does not consider the institutions of exchange. The market equilibrium with endogenous firms provides a more general theory of value because it reflects transaction costs.

In neoclassical economics, the firm is a veil that covers the decisions of consumers as well as those of suppliers, customers, and owners of firms. Once the veil of firms is removed, all the action in the neoclassical tradition is on the consumer side. Consumers own firms and unanimously agree on the profit-maximizing input choice, so the production decision coincides with the consumer-owners’ decision. Consumers are the suppliers of firms, providing the firm with all of its primary factors of production, because consumer endowments include labor and all other resources. Consumers purchase all of the outputs of firms that are not used as inputs for other firms. The transformation of inputs into outputs through production is absorbed into the economy’s excess demand function. The production activity could just as well be owned and operated by consumers, who then supply outputs to firms or to other consumers.

The general theory of the firm draws from neoclassical economics and institutional economics. Neoclassical economics offers an analysis of the benefits of transactions through its study of the theory of value. Institutional economics offers an analysis of the costs of transactions through its study of the mechanisms of exchange. By considering both the benefits and costs of transactions, the general theory of the firm combines the strengths of these two important traditions.

The general theory of the firm presented here contains a number of new models and results. The discussion presents equilibrium models in which entrepreneurs choose to establish firms. The discussion examines equilibrium models that contrast the decisions of consumer organizations with those of firms. The analysis compares how consumer organizations and firms choose the size of the organization, obtain capital investment, provide incentives, and allocate resources. The discussion considers how firms create and operate markets. The general theory of the firm helps to explain the vital economic role played by the firm in the contemporary economy. Microeconomics with endogenous entrepreneurs, firms, markets, and organizations provides insights into the nature and function of economic institutions.

PART I

THE THEORY OF THE FIRM

The Consumer

The main tasks of the economic theory of the firm are to determine why firms exist, how firms are established, and what functions firms perform. Firms are economic institutions whose objectives, decisions, and activities are the result of fundamental economic forces. To explain the economic role of firms, economic analysis must derive firms endogenously from initial conditions.

Take firms away from microeconomics and what is left? – consumers! This immediately suggests that the theory of the firm should begin with an economy in which there are only consumers. Starting with consumers as the givens of the model has a key implication. Consumers will establish firms if and only if doing so improves economic efficiency. The theory of the firm necessarily derives the existence of firms from fundamental assumptions about the characteristics of consumers who have exogenously given preferences and endowments. Therefore, consumers are the basic building blocks of the theory of the firm.

The purpose of this chapter is to examine the characteristics of consumers and economic interaction between consumers. Consumers are defined by their preferences and endowments. Consumers own ideas, business methods, technologies, and other intellectual property. Consumers can engage in research and development (R&D) to find new technology. Consumers own production technologies and they can act as producers by operating the production technology themselves. Consumers obtain benefits from exchange with one another and incur transaction costs.

This chapter introduces the *intermediation hypothesis*. Firms play an economic role when intermediated exchange is more efficient than direct exchange. *Intermediated exchange* refers to transactions between consumers that go through firms. *Direct exchange* refers to transactions between consumers that do not involve firms. Consumers have the ability to engage in transactions with each other, in the form of either spot transactions or contracts. Consumers can form various organizations to coordinate their transactions, including clubs, cooperatives, associations, non-profits, and partnerships. There are economic benefits to establishing firms when they improve efficiency in comparison to direct exchange.

The distinction between the firm and direct exchange depends on the *separation criterion*. The critical difference between firms and direct exchange is the separation of the firm's objectives from those of its consumer-owners. A transaction institution is defined to be a *firm* if its objectives can be separated from those of its owners. The next chapter examines how the objectives of firms differ from those of their owners. Consumer organizations have objectives that cannot be separated from those of their owners, so they are not firms. Direct exchange between consumers is the benchmark for evaluating the economic contribution of the firm. This chapter examines direct exchange between consumers as the alternative to intermediation by firms.

1.1 The Intermediation Hypothesis and the Scope of the Firm

Consumers in the economy are connected by all kinds of personal interactions – “no man is an island, entire of itself.”¹ Consumers are linked economically through their roles as buyers, sellers, workers, entrepreneurs, managers, investors, and owners. Their economic relations are embedded within complex ties of family, friendship, society, and nation. Yet, despite these many connections, there remain economic forces that divide individuals – these forces are transaction costs. Firms play an economic role by mitigating transaction costs and improving the efficiency of transactions. This section presents the intermediation hypothesis.

1.1.1 Transaction Costs and the Intermediation Hypothesis

A *transaction* is the creation of value by voluntary cooperation between two or more economic actors. The value created by a transaction equals the benefits to the parties that are generated by their cooperation minus the costs to the parties of arranging the cooperation. Economic actors structure transactions to maximize their net gains from trade, that is, transaction benefits net of transaction costs.

Consumers encounter transaction costs when they engage in direct exchange. Consumers encounter transaction costs in pure exchange of goods. Consumers also encounter transaction costs in forming and operating organizations such as clubs, consumer cooperatives, worker cooperatives, nonprofits, and basic partnerships. Organizational transaction costs sometimes are referred to as governance costs. Consumers incur transaction costs of communication, information processing, search, matching, bargaining, moral hazard, adverse selection, free riding, and contracting.

The key question is “Why do consumers need firms?” To be profitable, firms must provide better transactions in comparison with direct exchange between consumers. To improve the efficiency of transactions, the firm creates both markets

1 John Donne, Meditation XVII, 1975 [1572–1631].

Table 1.1. *The firm improves the efficiency of transactions by creating and managing markets and organizations*

Consumer transaction costs	The firm's transaction mechanisms	
	Market transactions	Organizational transactions
Communication and information processing	Communication networks and platforms	Organizational hierarchies and information technology
Search and matching	Centralized marketplace Match-making Intermediation	Internal coordination Assignments to tasks Internal allocation
Bargaining	Posted prices, auctions	Relational contracts
Moral hazard	Monitoring and incentives	Monitoring and incentives
Adverse selection	Incentive contracts and quality certification	Incentives and quality management
Free riding	Incentive contracts	Incentives and coordination
Contracting	Market for contracts	Vertical integration

and organizations. A *market* is a mechanism that brings buyers and sellers together. A market can be a store, a Web site, a matchmaker, or an auction. The firm's *market transactions* are those that the firm conducts with individuals and firms outside of its organization. The firm creates markets by designing the institutions of exchange.

An *organization* can involve hierarchies, bureaucracies, groups, teams, and networks. The firm's *organizational transactions* are those that it conducts with its managers, employees, and business units. The firm creates organizations by hiring managers and delegating authority to them, by establishing internal rules and mechanisms for communication, by defining tasks for employees, and by assigning employees to those tasks. The firm provides incentives for performance to its managers and employees and monitors their performance.

The *scope* of the firm's activities is the *combination* of its market-making and organizational activities. The firm expands its scope by increasing its market transactions and its organizational transactions. The scope of the firm is determined by its ability to improve the efficiency of economic transactions relative to direct exchange and consumer organizations. The firm plays an economic role if it provides greater efficiency than direct exchange between consumers. Firms intermediate exchange through markets and through organizations. Table 1.1 provides an overview of how the firm addresses transaction costs that consumers face in direct exchange. The general theory of the firm suggests a set of empirical implications that are referred to as the "intermediation hypothesis."

Intermediation Hypothesis. Higher consumer transactions costs of direct exchange, in comparison to firms' transaction costs, lead to firms intermediating exchange between consumers.

The intermediation hypothesis compares the costs of direct exchange between consumers to the costs of intermediated exchange through firms. The hypothesis makes predictions about what types of institutions of exchange will be observed.

The intermediation hypothesis involves comparison of alternative institutions of exchange. The comparison applies whether costs are rising or falling. The development of electronic commerce, including improvements in telecommunications, information systems, and the Internet, lowers transactions costs for a wide variety of transactions. This implies that the costs of direct exchange and the costs of intermediated exchange are both likely to fall. If improvements due to electronic commerce lower transaction costs of intermediated exchange more than those for direct exchange, the intermediation hypothesis predicts more intermediation by firms. The advent of electronic commerce thus may explain the entry of a wide variety of highly specialized firms that intermediate transactions. Even though the costs of direct exchange decrease, the costs of intermediated exchange decrease even more, leading to a restructuring of institutions of exchange.

1.1.2 The Firm versus Direct Exchange

Transaction costs can be either explicit or implicit. Transaction costs such as those associated with communication, information processing, and search involve resource expenditures. The presence of explicit transaction costs can result in additional implicit transaction costs in the form of inefficient transactions. Transaction costs associated with matching, bargaining, moral hazard, adverse selection, free riding, and contracting can be implicit and appear only in the form of efficiency distortions. To highlight the role of the firm, it is useful to consider the basic framework with implicit transaction costs.

Suppose that there are n consumers. Denote each consumer's allocation by q_i , which may be a vector. The allocation takes values in the set Y_i . Let q be the vector of consumer allocations, which takes values in the set $Y = Y_1 \times \cdots \times Y_n$. Each consumer has an additively separable utility function of the allocation vector, q , and a scalar numeraire good, x :

$$(1) \quad U_i(q, x_i) = U_i(q) - x_i.$$

The vector of allocations, q , can represent externalities and public goods. In the case of private goods, the effect on utility of the vector of allocations, q , is restricted to consumer i 's allocation, q_i . Let u_i be consumer i 's opportunity costs of trade. For ease of presentation, the opportunity costs of trade are assumed to be independent of allocations. The discussion can be extended to include more general opportunity costs.

Let $C(q)$ represent the costs of production and other joint costs. Then the gains from trade can be represented by

$$(2) \quad \Gamma(q) = \sum_{i=1}^n U_i(q) - C(q) - \sum_{i=1}^n u_i.$$

Denote a socially optimal allocation by q^* , where

$$q^* \in \arg \max_q \Gamma(q).$$

Consider direct exchange between consumers. Let q_{-i} denote the vector of allocations excluding consumer i 's allocation. Define the payment function $p_i(q_i, q_{-i})$ for consumers $i = 1, \dots, n$. The gains from trade for an individual consumer equal

$$(3) \quad V_i(q_i, q_{-i}, p_i) = U_i(q_i, q_{-i}) - p_i(q_i, q_{-i}) - u_i.$$

Consumers participate in direct exchange only if they obtain gains from trade. Consumer i 's individual rationality condition is

$$U_i(q_i, q_{-i}) - p_i(q_i, q_{-i}) \geq u_i.$$

In direct exchange, the payments are subject to a budget-balancing condition,

$$(4) \quad \sum_{i=1}^n p_i(q_i, q_{-i}) = C(q).$$

Combining the individual consumer gains from trade definition with the budget-balancing condition gives total gains from trade

$$(5) \quad \sum_{i=1}^n V_i(q_i, q_{-i}) = \Gamma(q).$$

Consider direct exchange between consumers in the absence of either explicit or implicit transaction costs. Consumers can coordinate their choices of allocations and payments. The allocations, q , are observable and consumers can form contractual agreements. Consumers can play a cooperative game in which they choose among allocations that maximize their net gains from trade. For example, allocations and payments can be determined by outcomes in the Core of a cooperative game. It follows from equation (5) that consumers will maximize net gains from trade.

Proposition 1. Without transaction costs, direct exchange between consumers maximizes gains from trade.

Suppose now that there are explicit or implicit transaction costs of direct exchange. Transaction costs such as search, moral hazard, adverse selection, free riding, or contracting costs can prevent consumers from attaining the efficient outcome. These costs prevent consumers from coordinating their activities to choose cooperatively among allocations that maximize gains from trade. One possibility is that direct exchange entails explicit transaction costs. Sufficiently large transaction costs result in autarky, with consumers choosing their outside option u_i .

Suppose that there are implicit transaction costs. These can limit exchange to bilateral transactions or they can limit the types of payment functions that consumers can choose. Another implicit form of transaction costs requires that consumers can only choose allocations in a noncooperative fashion. Thus, a Nash noncooperative equilibrium, q^0 , is defined by

$$q_i^0 \in \arg \max_{q_i} V_i(q_i, q_{-i}^0), \quad i = 1, \dots, n.$$

Designate the equilibrium gains from trade with direct exchange by

$$(6) \quad g = \Gamma(q^0).$$

Suppose that the payment functions are given exogenously. For many payment functions, the noncooperative equilibrium does not correspond to the efficient outcome. Due to the budget balancing condition, there are settings in which there do not exist payment functions such that the noncooperative outcome is efficient. This implies the following result.

Proposition 2. With transaction costs, direct exchange between consumers may not maximize gains from trade.

For example, in the classic prisoners' dilemma game, the outcome fails to be efficient. The prisoners' dilemma has exogenous payments that result from legal penalties. The payoffs are represented in the table below. The unique outcome of the game occurs when the prisoners do not cooperate with each other, they both confess and receive the payoffs (0, 0). Transaction costs appear in the form of infinite costs of communication. The prisoners cannot communicate with each other to make an agreement on their strategies. If the two prisoners could form an agreement on their strategies they would cooperate with each other and receive the payments (2, 2).

	Cooperate	Do not cooperate
Cooperate	2,2	-1,3
Do not cooperate	3,-1	0,0

Suppose that the players in the prisoners' dilemma game can make binding agreements with action-contingent side payments. There exist side payments that transform the payoffs in the game; see for example Gutman (1978, 1987). As a result of these transfers, the players achieve the efficient outcome, which is cooperate-cooperate, as a noncooperative equilibrium. However, Jackson and Wilkie (2005) show that when players choose side payments noncooperatively, there is no equilibrium that results in the efficient outcome. Thus, even with endogenous action-contingent side payments, the outcome is inefficient. Again, when transaction costs result in noncooperative behavior, the outcome is inefficient.

The firm is able to intermediate exchange. The firm offers consumers different payment schedules $P_i = P_i(q_i, q_{-i})$. The firm's objective is to maximize profit,

although this objective is derived endogenously in later chapters. The firm's profit equals

$$(7) \quad \Pi(q) = \sum_{i=1}^n P_i(q) - C(q).$$

Define the allocation achieved by a profit-maximizing firm by

$$q^F \in \arg \max_q \Pi(q).$$

Let gains from trade with a profit-maximizing firm be denoted by

$$(8) \quad G = \Gamma(q^F).$$

Suppose that the firm chooses allocations and payments that are subject only to the consumers' individual rationality constraints,

$$U_i(q_i, q_{-i}) - P_i \geq u_i.$$

Suppose that the firm exercises monopoly power by choosing allocations and payments such that consumers' individual rationality constraints are strictly binding. For example, the firm can engage in first-degree price discrimination, setting payments that extract all of the consumers' gains from trade. Then, substituting from the individual rationality constraints into the profit function (7), it follows that the firm's profit function exactly equals total gains from trade,

$$(9) \quad \Pi(q) = \Gamma(q).$$

This implies that the profit-maximizing firm chooses the efficient allocation. This result is standard for firms with private goods that can engage in first-degree price discrimination; see for example Spulber (1979). For a similar result with externalities in a full information setting, see Segal (1999).

Proposition 3. When the firm can choose both allocations and payments subject only to consumers' individual rationality constraints, the profit-maximizing allocation is efficient, $q^F = q^*$.

This result implies that when consumers cannot form cooperative agreements, firms play an economic role by providing coordination.

More generally, suppose that the firm has incomplete information about the characteristics of consumers. The firm chooses allocations and payment schedules to maximize profits subject to consumer individual rationality constraints and incentive compatibility constraints,

$$\max_p \sum_{i=1}^n P_i(q) - C(q),$$

subject to

$$U_i(q_i, q_{-i}) - P_i(q_i, q_{-i}) \geq u_i, \\ q_i \in \arg \max_{q_i} U_i(q_i, q_{-i}) - P(q_i, q_{-i}), \quad i = 1, \dots, n.$$

In a multiple-agent model with externalities and asymmetric information, Spulber (1988b) shows that the full-information optimum can be achieved by incentive contracts if and only if the individual rationality conditions for agents are nonbinding. In a regulation model with asymmetric information about both supply and demand, Spulber (1988a) shows that the full-information optimum can be achieved with sufficient gains from trade. In both of these analyses, the principal must give agents information rents to induce truthful revelation of information. When gains from trade at the full-information optimum are sufficient to cover information rents, the optimum is attained. Conversely, when gains from trade at the full-information optimum do not cover information rents, the optimum is not attained. The profit-maximizing firm will depart from the social optimum because of the trade-off between increasing profits and inducing revelation of information.

The comparison of the firm with direct exchange between consumers has two main implications. First, the firm can serve as a means of exchange. Suppose that the transaction costs of direct exchange are sufficiently large so that they result in autarky. Then the firm plays a role when $G \geq 0$ by providing gains from trade. This role is the basis of the neoclassical separation theorem, which essentially identifies the firm as a vehicle for achieving gains from trade.

Second, the firm provides intermediated exchange that competes with direct exchange. Generally, with transaction costs, both direct exchange between consumers and intermediated exchange with firms depart from the optimum. The value created by the firm is G , which equals gains from trade adjusted for the firm's transaction costs. If direct exchange is feasible, $g \geq 0$, the firm's added value relative to direct exchange is $G - g$. Then, the firm improves economic efficiency if and only if it increases gains from trade,

$$G \geq g.$$

This fundamental inequality describes the economic role of the firm. What distinguishes direct exchange from intermediated exchange is the intervention of an additional player with a distinct objective. If the firm adds value relative to direct exchange, there are benefits to consumers from establishing an institution with a distinct objective.

1.1.3 Example 1: Explicit Transaction Costs

Suppose that transaction costs are explicit. The basics of the model can be illustrated for a simple economy with two consumers and two goods, q_1 and q_2 . Suppose that the first consumer derives a per-unit benefit v from consuming the first good and

has a unitary marginal utility of the second good, so that the consumer's utility function is

$$U^1(q_{11}, q_{12}) = vq_{11} + q_{12}.$$

The second consumer derives a per-unit benefit c from consuming the first good and has a unitary marginal utility of the second good,

$$U^2(q_{21}, q_{22}) = cq_{21} + q_{22}.$$

The first consumer has an endowment $(0, \omega)$ and the second consumer has an endowment $(1, 0)$. The two consumers can obtain gains from trade by exchanging some or all of their endowments.

Suppose that the first good is not divisible and is available in discrete units, such as hats or chairs. The second good is a perfectly divisible commodity, such as wheat or wood. The second good can be used as commodity money providing a store of value, a medium of exchange, and a unit of account. In the pure-exchange economy with two consumers, the first consumer can be seen as a buyer and the second consumer as a seller.

Suppose that the two consumers live on different islands. If the buyer and seller engage in trade, they jointly encounter transaction costs that consume t units of the second good.² Transaction costs are distinct from the buyer's enjoyment of the good and from the seller's cost of providing the good. Suppose that the initial endowment of the divisible good is sufficient to cover transaction costs and to compensate the second consumer for transferring the first good, $\omega > t + c$.

Consider first the state of autarky. The buyer's initial benefit is $U^1(0, \omega) = \omega$. The seller has an endowment that consists of one unit of the discrete good and no endowment of the monetary good, so that the seller's initial benefit is $U^2(1, 0) = c$.

Next, compare the state of autarky with direct exchange between the two consumers. For simplicity, assign all of the transaction costs to the seller. If the buyer obtains a unit of the discrete good and pays a price $p < \omega$ of the monetary good, the consumer's benefit is $U^1(1, \omega - p) = v + \omega - p$. If the seller trades the unit of the discrete good to the first consumer in return for p units of the monetary good and incurs the joint transaction costs, the second consumer's benefit is $U^2(0, p - t) = p - t$.

The first consumer will wish to trade only if there are gains from trade,

$$U^1(1, \omega - p) - U^1(0, \omega) = v - p \geq 0.$$

The second consumer also will wish to trade only if there are gains from trade,

$$U^2(0, p - t) - U^2(1, 0) = p - c - t \geq 0.$$

For both consumers to obtain positive gains from trade requires that there are positive net gains from trade, $v - c - t > 0$. Both consumers have positive gains

2 Alternatively, transaction costs might be consumer-specific, where $t_1 + t_2 = t$.

from trade for p in $(c + t, v)$. The buyer makes any payment p that does not exceed the minimum of the buyer's benefit and the initial endowment of the monetary good, $p \leq \min\{v, \omega\}$. The seller accepts any payment that covers the opportunity cost of selling the good plus transaction costs, $p \geq c + t$.

Direct exchange between the buyer and seller yields transaction benefits net of transaction costs,

$$g = U^1(1, \omega - p) - U^1(0, \omega) + U^2(0, p) - U^2(1, 0) = v - c - t.$$

Direct exchange takes place only if there are net gains from trade, $g \geq 0$. This requires that gains from trade cover transaction costs, $v - c - t \geq 0$. If the net gains from trade are negative, the buyer and seller choose not to trade; that is, autarky results. Transaction costs are hidden under autarky because no trade occurs. Thus, implicit transaction costs are the forgone returns to trade, $v - c$.

The value created by a transaction equals the net gains from trade g between the buyer and the seller. The economy's transactions create value that equals the benefits of transactions net of the costs of organizing those transactions. Many of the economy's institutions of exchange are determined endogenously by the activities of consumers and firms maximizing the net benefits of transactions. The value created by the economy depends on the realization of gains from trade and the types of institutions of exchange that develop.

Firms establish institutions of exchange that supplement direct transactions between individual consumers. A firm offers the seller and the buyer an alternative set of transaction opportunities and mechanisms of exchange. To represent this, suppose that when transacting with the firm, the buyer obtains value V from the transaction and the seller incurs cost C in providing the good. This means that firms can improve the goods and services received by consumers through product and transaction innovations, potentially raising V . Also, firms can improve productivity and efficiency through process and transaction innovations, potentially lowering C .

The firm creates a transaction in many different ways. The firm can purchase a good from the seller and resell it to the buyer. The firm can match buyers and sellers and broker transactions. The firm can bring buyers and sellers together by transforming the good in some way through manufacturing, transportation, and so on. The transaction technology T refers to the firm's mechanisms of exchange that are distinct from those available for direct exchange between consumers. For example, consumers can engage in bilateral negotiation to complete an exchange. However, a firm can employ multilateral pricing mechanisms such as auctions. The types of transaction mechanisms specific to firms are discussed more fully in subsequent chapters.

Consider a firm that performs the basic functions of a dealer. The firm obtains the good from the seller and in return makes a payment to the seller. The firm supplies the good to the buyer and collects a payment from the buyer. The firm's intermediation activities consume resources. Let T represent the total of per-transaction costs for the seller, the buyer, and the firm. Transaction costs require using T units

of the divisible good. Assume that there is a sufficient endowment of the divisible good to cover the transaction costs and compensate the seller, $\omega > C + T$.

For ease of presentation, suppose that the firm bears all of the transaction costs in its dealings with the buyer and the seller. This is equivalent to assuming that the firm compensates the seller and the buyer for the costs that they incur. The first consumer pays P to the firm and obtains gains from trade relative to autarky equal to

$$U^1(1, \omega - P) - U^1(0, \omega) = V - P \geq 0.$$

The second consumer receives W from the firm and obtains gains from trade relative to autarky equal to

$$U^2(0, W) - U^2(1, 0) = W - C \geq 0.$$

The firm's profit equals the bid-ask spread net of transaction costs,

$$\Pi = P - W - T.$$

The firm's creation and coordination of transactions through intermediated exchange yields gains from trade net of transaction costs equal to

$$G = U^1(1, \omega - P) - U^1(0, \omega) + U^2(0, W) - U^2(1, 0) + \Pi = V - C - T.$$

In comparison to autarky, the buyer and seller will trade through the firm only if the gains from trade are greater than total transaction costs, $G \geq 0$. The firm improves economic efficiency relative to direct exchange if and only if it increases net gains from trade, $G \geq g$, or equivalently,³ $V - C - T \geq v - c - t$.

Suppose that there are n consumers and that consumers are restricted to bilateral transactions due to limits on their time and the costs of effort devoted to transactions. Suppose also that a firm can implement a transaction technology that permits multiple transactions. The firm's transaction technology has a fixed cost, T . Then, if the firm's transactions and the consumers' direct transactions are otherwise identical, the firm improves on direct exchange through economies of scale in transactions. Thus, $G \geq g$ when the firm offers lower total transaction costs in comparison to n bilateral transactions, $T \geq nt$. When the economy is sufficiently large, intermediated exchange through firms replaces direct exchange.

Figure 1.1 summarizes the comparison between the firm and direct exchange. The transaction triangle shows the alternatives. The line connecting the seller and the buyer represents direct exchange. The lines connecting the seller and the buyer to the firm represent the intermediated transactions. After a firm is established, the buyer and the seller choose between direct exchange with gains from trade g and intermediated exchange with gains from trade G . The firm plays an economic role if and only if intermediation of transactions yields greater benefits than does direct exchange.

3 If direct exchange is not feasible, the firm's added value equals its value created, G . Thus, if direct exchange is not feasible, the firm improves the efficiency of exchange relative to autarky if and only if $G \geq 0$.

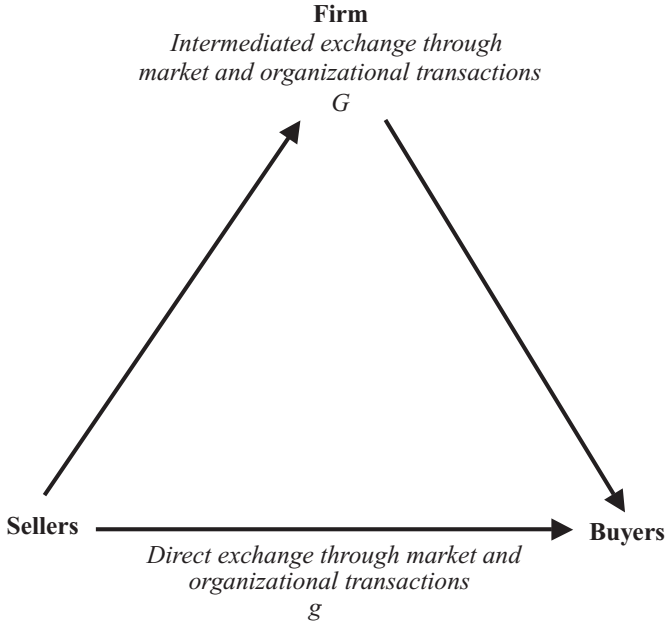


Figure 1.1. Comparison of intermediated exchange by firms with direct exchange between consumers. Intermediated exchange through the firm is efficient if and only if $G \geq g$ for $g \geq 0$, and $G \geq 0$ otherwise.

1.1.4 Example 2: Search versus Market-Making Firms

To illustrate the firm's role in intermediating exchange, consider an example with imperfect matching in a search market. Transaction costs are implicit because they take the form of imperfect matches of buyers and sellers. Suppose that there are two buyers with willingness to pay v_L and v_H , where $v_L < v_H$. Also, suppose that there are two sellers with costs c_L and c_H , where $c_L < c_H$.

Denote the expected value of the buyers' willingness-to-pay levels by

$$v = (1/2)v_L + (1/2)v_H.$$

Also, denote the expected value of the sellers' costs by

$$c = (1/2)c_L + (1/2)c_H.$$

A high-willingness-to-pay buyer can trade with both types of seller, $v_H > c_H$. A low-cost seller can trade with both types of buyer, $v_L > c_L$. Assume that $v_L < c_H$, so that a high-cost seller cannot trade with a low-willingness-to-pay buyer. This assumption means that the supply curve crosses the demand curve.

Before entering the market, buyers and sellers do not know the type of their trading partners. Assume that after a buyer and a seller decide to trade, they learn each other's type. At that point, trade occurs if and only if they have gains from trade. A buyer and a seller split the gains from trade evenly.

Given $v_L < c_H$, the search market is inefficient. It is easy to demonstrate that otherwise, if $v_L \geq c_H$, all matches are feasible, so that direct exchange is efficient. Moreover, if $v_L \geq c_H$, there do not exist any prices at which a monopoly firm is viable.⁴ Thus, consumers would have no need for a firm to intermediate exchange.

With random matching, the expected gains from trade for the high-willingness-to-pay buyer equal $(v_H - c)/2$. The expected gains from trade for the low-cost seller equal $(v - c_L)/2$. However, because $v_L < c_H$, a low-willingness-to-pay buyer has expected gains from trade in the direct exchange market equal to $(v_L - c_L)/4$, because he has only one other potential trading partner. A high-cost seller has expected gains from trade equal to $(v_H - c_H)/4$.

Consider a firm with fixed transaction costs, T . The profit-maximizing firm chooses the highest ask price, p , and the lowest bid price, w , that will attract the high-willingness-to-pay buyer and the low-cost seller,

$$p^M = v_H - (v_H - c)/2 \quad \text{and} \quad w^M = c_L + (v - c_L)/2.$$

The firm's profit equals

$$\Pi = p^M - w^M - T = (c_H - v_L)/2 - T.$$

The low-willingness-to-pay buyer and high-cost seller are inactive because they do not gain from direct exchange with each other.

Direct exchange between consumers, that is, exchange without the presence of firms, yields expected gains from trade equal to

$$g = (1/2)(v_H - c_L) + (1/2)[(v_H - c_H) + (v_L - c_L)].$$

The expected gains from trade with a firm equal

$$\begin{aligned} G &= (v_H - p^M) + (w^M - c_L) + p^M - w^M - T \\ &= v_H - c_L - T. \end{aligned}$$

The firm improves economic efficiency relative to direct exchange if and only if it increases net gains from trade. This is equivalent in this case to the requirement that the firm is profitable,

$$G - g = \Pi = (c_H - v_L)/2 - T \geq 0.$$

The firm obtains profit equal to the increase in the gains from trade. The firm improves efficiency if and only if the gap between the high-cost seller and the low-willingness-to-pay buyer is greater than or equal to the firm's fixed transaction costs. A larger gap, all other things equal, reduces the gains to consumers from direct exchange. A larger gap correspondingly increases the profits to the firm from intermediating exchange. This illustrates the intermediation hypothesis.

4 If $v_L \geq c_H$, all combinations of buyers and sellers can trade in the direct exchange market. Then, because they are uninformed about the type of their trading partners, the expected gains from trade of a type- i buyer and a type- j seller from the direct exchange market are respectively $(v_i - c)/2$, $i = L, H$, and $(v - c_j)/2$, $j = L, H$.

1.1.5 Example 3: Clubs versus Firms

Consider a club whose members share the fixed costs of producing an excludable public good. Let the fixed costs of the good equal c . The number of members of the club equals n . Members of the club derive benefits $u(n)$ from consuming the public good. The club is subject to problems of congestion, so that the member benefits $u(n)$ are decreasing and concave in the size of the club. Suppose that individuals may obtain a substitute good without joining the club, with net benefits v .

The identification of the club's objectives with its members' consumption benefits is what distinguishes a club from a firm. The members of the club determine the size of the club. Their objective is to maximize their individual consumption benefits,

$$u(n) - c/n.$$

The optimal size of the club, n^0 , therefore solves

$$-u'(n^0) = c/(n^0)^2.$$

The club is viable if it offers benefits over competing alternatives, $u(n^0) - c/n^0 \geq v$.

Consider a profit-maximizing firm that offers the excludable public good. The firm charges the highest price subject to the constraint from the outside alternative. Thus, the firm's price equals

$$p = u(n) - v.$$

The firm's profit is then

$$\Pi(n) = pn - c = n(u(n) - v) - c.$$

The firm chooses the number of customers to maximize its profit, trading off congestion effects that reduce consumer benefits with the revenue advantages of serving more customers. The firm's profit-maximizing number of customers, n^F , solves

$$-u'(n^F) = (u(n^F) - v)/n^F.$$

The firm maximizes total benefits and thus is more efficient than the club, which maximizes average benefits.⁵ The profit-maximizing firm serves more customers than the club, $n^F > n^0$.⁶

The members of the club would not prefer to organize the club as a firm. This is because club membership maximizes the members' individual benefits. Suppose

5 The fixed costs of producing the public good do not enter into the firm's choice of the number of customers. If there were variable costs associated with the quantity of a public good, these would affect the firm's choices.

6 From the firm's first-order condition, it follows that $-u'(n^F) = (u(n^F) - v)/n^F > c/(n^F)^2$. Comparing to the first-order condition for the club, and noting that $u''(n) < 0$, implies that $n^F > n^0$.

that members could purchase the good from the firm and obtain a share of the firm's profits. Then a member of the club would obtain greater benefits than a customer/owner of the firm:

$$u(n^0) - c/n^0 \geq u(n^F) - c/n^F = u(n^F) - p^F + \Pi(n^F)/n^F.$$

Suppose, however, that the original members of the club are considered as owners of the club, and that there is a market for club memberships. A membership market for labor-managed firms induces profit maximization; see Dow (2003, p. 149).

Consider how a membership market affects the club's objectives. Let n^0 be the number of original members and let n be the expanded size of the club. All members of the club would obtain benefits net of shared fixed costs, $u(n) - c/n$. A new member is charged a membership fee equal to the difference between net benefits and the value of the outside opportunity, $u(n) - c/n - v$. Insiders choose the number of total members to maximize their individual benefits plus their share of earnings from selling new memberships,

$$U(n) = u(n) - c/n + (u(n) - c/n - v)(n - n^0)/n^0.$$

Simplifying this expression yields

$$U(n) = (1/n^0)\Pi(n) + v.$$

Maximizing each member's benefits over the size of the club yields the profit-maximizing size, n^F . Therefore, when insiders sell shares in the club to new members, they will choose the profit-maximizing membership.

This provides a separation theorem. The objectives of a club can be separated from those of its members when the club's inside members sell memberships to new members. The inside members are the owners of the firm. When inside members choose to sell shares to outside members, the objectives of the club become separate from their consumption objectives and the club becomes a firm.

When consumers form a club, the gains from trade in direct exchange equal

$$g = n^0 u(n^0) - c - n^0 v.$$

Establishing a firm that intermediates exchange yields gains from trade equal to

$$G = n^0 U(n^F) - n^0 v = n^F u(n^F) - c - n^F v.$$

Profit maximization implies that gains from trade increase when a firm intermediates exchange,

$$G - g = \Pi(n^F) - \Pi(n^0) > 0.$$

Because the firm's objectives are separate from the consumption objectives of its owners, establishing a firm increases gains from trade.

1.2 Consumer Characteristics

The general theory of the firm begins with consumers. There is a set of consumers in the economy, $N = \{1, 2, \dots, n\}$. Denote by i an individual consumer. In the theory of the firm, the preferences and endowments of consumers generally are taken as givens.⁷ The preferences and endowments of consumers are the foundations of the theory of the firm.

The discussion follows standard usage in designating all individuals as “consumers.” In addition to being buyers of goods and services, consumers also play important economic roles as workers, managers, suppliers of resources, owners of assets, and entrepreneurs. The number of consumers n will turn out to have significant effects on transaction benefits and costs. This is perhaps the critical determinant of whether or not firms will be established in the economy.

1.2.1 Consumer Endowments

In describing consumer endowments, the theory of the firm incorporates the neo-classical theory of the consumer, but with some notable extensions and additions. Consumer endowments of commodities can be represented as a vector ω^i in the consumption set X^i in the product space R^L , where i is an element of N and where L is the number of products. The number of goods in the economy can be taken as given or the number of goods can be determined endogenously through product innovation.

Consumer endowments can be final consumption goods. Consumer endowments also can be factors of production such as land and natural resources or commodities that are used as both productive inputs and consumption goods. Consumers can have an initial endowment of time that can be used for leisure, labor, and other time-consuming tasks. The consumer can also have an endowment of fiat money or a numeraire good, if these exist. These possibilities conform to the standard neoclassical model.

Consumers in the theory of the firm own production technologies. In the neoclassical framework, consumers own production technologies through their ownership of firms. In the present framework, consumers can operate production technologies without the need for firms. Consumer i owns a production technology Y^i in the product space R^L , where i is an element of the set of consumers N . A

7 Some models consider consumers in a transferable utility setting. Other models apply the neoclassical theory of the consumer, in which each consumer i has ordinal preferences over consumption bundles in their consumption set within a product space. The preferences of the consumer are generalized in standard ways. In particular, consumers can value time spent on leisure. Consumers can evaluate consumption bundles that have goods that are consumed at different points in time. They can apply subjective rates of discount to future consumption. Some models examined in later chapters also consider consumers with von Neumann-Morgenstern utility functions exhibiting risk aversion. As in agency models, consumers can experience disutility of effort.

consumer can operate a production technology either individually or in concert with other consumers. In the neoclassical setting, firms are defined as production technologies and only firms can be producers. Here, firms are institutions rather than production technologies.

Consumers in a general theory of the firm possess practically any knowledge that people have in practice. The theoretical analysis should not artificially limit consumer knowledge while *a priori* assigning that knowledge to firms. Consumers possess information, ideas, and other types of intellectual property. Consumers have scientific and engineering information and many other forms of knowledge as part of their endowments. Consumers can have private information about their own preferences and endowments that is not observable by others. Consumers, in some cases, can have information about the preferences and endowments of other consumers.

The technologies and intellectual property owned by consumers can be established exogenously. Alternatively, the production technologies owned by consumers can be determined endogenously through process innovation. Consumers can transfer technologies to other consumers or to firms, including the sale of intellectual property.

In addition to production technologies, consumers can own transaction technologies. Such a transaction method t^i , where i is a member of the set of consumers, can be used to represent a vast array of possible ways of interaction between consumers. They can include search, communication, negotiation, payment methods, and contract formation. These types of transaction methods are referred to as direct exchange or decentralized exchange to indicate that they operate without firms. These types of transaction technologies do not operate through intermediaries such as firms.

A consumer i also can own a transaction method T^i that represents centralized exchange through firms, where i is an element of the set of consumers N . These types of transaction technologies include centralized market mechanisms such as posted prices, auctions, and matchmaking services.

Consumer-entrepreneurs may establish firms and provide the firms with their transaction technology. The transaction technology T^i can be viewed as the intellectual property of the consumer. The transaction technology also can be interpreted as the result of invention of a new business method. Spulber (2002c) introduces the concept of a “transaction innovation.”

1.2.2 Consumer Activities

Consumers play many economic roles. They can engage in a wide variety of activities, some of which are consistent with the neoclassical model and some of which clearly are not. As in the traditional neoclassical setting, consumers purchase final consumption goods and supply primary goods to other consumers and to firms. Consumers provide labor services, acting as consumer-workers.

Consumers can operate production technology, acting as consumer-producers. As already noted, this departs from the neoclassical setting in which only firms can be producers. Consumers are skilled craftsmen, manufacturers, and farmers. Consumers are merchants and intermediaries between other consumers. They provide services as physicians, attorneys, and accountants. Consumers create works of art as musicians, composers, dancers, authors, painters, and sculptors.

Also, consumers are inventors, devising new products or new production processes. They generate new information by imagination, observation, calculation, and reasoning. They can create scientific research and are owners of intellectual property. Consumer-inventors can employ inventions in production or they can sell the intellectual property to other consumers.

Consumers also participate in financial markets by borrowing, saving, and investing in firms. Once firms are established, consumers then transact with firms as customers buying the firms' output. Consumers also supply the firm with capital, resources, production technology, and transaction methods. Consumers are the managers and employees of the firm.

Perhaps most critically for the theory of the firm, consumers establish firms by acting as entrepreneurs. Having established a firm, a consumer-entrepreneur becomes an owner of the firm. This will be spelled out more fully in Chapters 4 and 5.

1.3 Consumer Cooperation and Transaction Benefits

Establishing firms can improve the efficiency of transactions relative to exchange between consumers without firms. To understand transactions, it is necessary to know the purpose of transactions. Because transactions are voluntary, they must yield benefits for each of the parties. Consumers transact with other consumers to obtain those benefits. Identifying potential gains from trade allows a classification of transactions. Such a classification suggests what types of costs are required to accomplish the transaction. Such a classification also shows the purpose of firms, because their objective is to increase transaction benefits net of transaction costs.

The Pareto ranking of economic allocations, which is at the heart of neoclassical economics, welfare economics, and international trade, tells us much about transaction benefits. Neoclassical analysis considers economic equilibria that are attained without transaction costs. Direct exchange between two consumers results in an allocation of goods on the contract curve. If there are no costs of establishing and operating a price system with a Walrasian auctioneer, then the two theorems of welfare economics hold. Any market equilibrium will be an allocation on the contract curve, and any allocation on the contract curve can be supported by a price system with a suitable redistribution of initial endowments. More generally, any market equilibrium is a Pareto optimal allocation and any Pareto optimal allocation can be supported by a market equilibrium given a suitable redistribution of initial endowments.

Representing transaction benefits using neoclassical efficiency conditions does not suggest that the economy reaches a Pareto optimum. Pareto improvement is a necessary but not a sufficient condition for a transaction. The choice of transactions for individuals and in the aggregate will depend on transaction costs and the institutions of exchange.

Although neoclassical economics has tended to neglect the institutions of exchange, it provides a solid foundation for understanding the benefits of exchange. The literature on transaction costs, in rejecting the frictionless equilibrium of neoclassical economics, tends to bypass many of the major contributions of classical and neoclassical economics. But these are the rules of economics – the returns to specialization and division of labor that have been studied at least since Adam Smith. A comprehensive understanding of the institutions of exchange necessarily incorporates the benefits of exchange.

Transaction benefits explain the incentives of individuals to participate in transactions. They illustrate the types of trading partners and the types of exchanges that would be preferred in a frictionless world. The complexity of transaction benefits helps to suggest the diverse types of transactions that would be needed to achieve some of these benefits. Some transaction benefits are more subtle and involve some types of cooperation that do not necessarily correspond to the neoclassical efficiency conditions. Of course, the types of trades that will occur in the economy cannot be determined without introducing transaction costs, which are the subject of the next section.

In what follows, many of the potential gains from trade result from pure exchange and from cooperation in production. The discussion is presented in terms of the benefits that consumers receive when they engage in direct exchange with each other. This includes exchange between consumers who own endowments of goods and services or consumers who own and operate production technology. This helps to distinguish gains from trade from institutions of exchange.

1.3.1 Marginal Rates of Substitution

Consider a pure exchange economy in which each consumer has an initial endowment of goods. The slope of a consumer's indifference curve is the marginal rate of substitution. For a given allocation of endowments of goods, marginal rates of substitution differ across consumers. By trading endowments, two consumers can reduce the differences between their marginal rates of substitution. Each consumer will receive an allocation of goods that they prefer to their initial endowment.

Consider a standard pure exchange economy with two consumers and two goods. Each consumer i has an initial endowment of the two goods, $\omega^i = (\omega_1^i, \omega_2^i)$, $i = 1, 2$. The consumers have preferences represented by continuously differentiable and concave utility functions $U^i(q_1^i, q_2^i)$, $i = 1, 2$. If the two consumers cannot meet to trade for some reason, then they are restricted to consumption of their initial endowments.

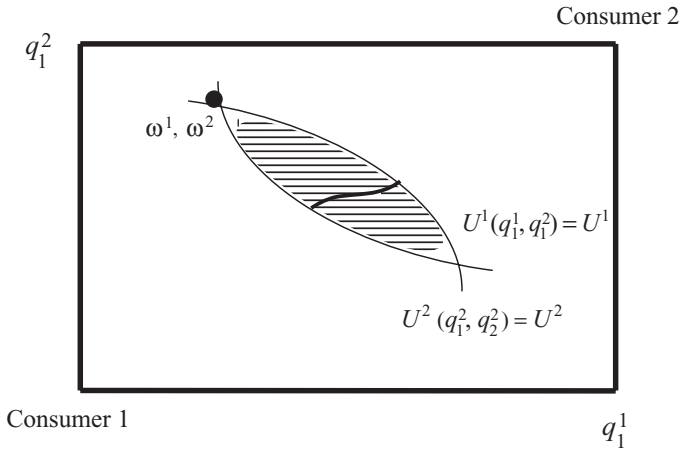


Figure 1.2. Potential gains from trade and the contract curve.

Gains from trade refer to the improvement of consumer benefits relative to what they obtain by consuming their initial endowments. Figure 1.2 depicts the classic Edgeworth-Bowley box. The set of allocations that make both consumers at least as well off and one consumer better off than they are at their initial endowments is represented by the lens-shaped shaded area in Figure 1.2. The contract curve is the set of Pareto-efficient allocations. Along the contract curve, the two consumers' marginal rates of substitution are equalized. The Core is the portion of the contract curve within the lens, which is shown as a bold line. The Core is the set of allocations that are Pareto efficient and Pareto superior to the initial endowment point.

These gains from trade cannot be attained unless the two consumers are able to carry out a transaction. Those transactions outside the lens are Pareto inferior to the allocation of initial endowments, because both consumers are worse off. Those transactions within the lens are Pareto superior to the allocation of initial endowments, while those allocations on the contract curve are Pareto superior to all other allocations. Each consumer is made better off without making the other consumer worse off.

In the neoclassical model of the pure exchange economy, consumers engage in direct exchange without transaction costs. Consumer exchange is mediated through markets that are established exogenously and the Walrasian auctioneer exogenously adjusts prices. Consumers maximize preferences subject to budget constraints where consumer wealth is the market value of initial endowments, evaluated at market-clearing prices. Consumers take prices as given exogenously. At the market equilibrium, each consumer's market basket is chosen independent of the choices of other consumers, although the preferences of all consumers affect aggregate demand and thus determine equilibrium prices along with the supply of initial endowments. In the neoclassical framework, therefore, prices separate consumer decisions in comparison with bilateral exchange through barter. When

there is a complete set of markets, that is, a market is established corresponding to each good, the economy attains a Pareto optimum.

If transactions are costly, gains from trade must be greater than or equal to transaction costs for trade to occur. Economic agents will choose those transactions that produce the greatest net gains from trade, that is, with the greatest transaction benefits net of transaction costs. The resulting transactions need not be those with the greatest gains from trade, because other transactions may be more convenient or efficient. With many goods and many consumers, reducing differences in marginal rates of substitution across the economy is likely to require many bilateral trades.

1.3.2 Marginal Rates of Technical Substitution

The neoclassical production technology when there are multiple inputs has isoquants that represent combinations of inputs necessary to produce a given output level. The slope of the isoquant is the marginal rate of technical substitution. Suppose that producers are consumers who own and operate production technologies. For a given allocation of input endowments, marginal rates of technical substitution differ across producers. By trading endowments of inputs, two producers can reduce the differences between their marginal rates of technical substitution. Each producer will receive an allocation of inputs that lets them increase output relative to the output they could produce with their initial endowment of inputs.

With many inputs and many producers, reducing differences in marginal rates of technical substitution is likely to require many bilateral trades. [Chapter 2](#) further examines the economy with production.

1.3.3 Relative Factor Productivities

In the classic Ricardian gains from trade, two countries have different relative factor productivities. For a given allocation of input endowments, and fixed coefficients technology, each of the two countries can change their output mix by altering the underlying allocation of inputs. By trading outputs, the two countries each can increase available outputs relative to their initial output mix. The two countries can be viewed as individual consumers who have endowments of inputs and each own and operate different production technologies. Such gains from trade apply to many other situations. For example, they extend to two producers who each potentially produce the same set of goods. Gains from trade are achieved by each producer specializing in different subsets of goods. With many producers and many goods, taking full advantage of potential gains from trade likely will require many bilateral trades.

1.3.4 Marginal Rates of Transformation

Suppose that a producer is a consumer who owns and operates a multiproduct technology. A producer with a given multiproduct technology and a given endowment

of inputs will have a production possibilities frontier that represents the combinations of products that can be produced. The marginal rate of transformation is the slope of the production possibilities frontier. Two producers with production possibilities frontiers can have different marginal rates of product transformation. By trading outputs, two producers can obtain classic gains from trade. The gains from trade result from moving along the production possibilities frontiers such that the producers reduce the differences between their marginal rates of transformation. The potential gains from trade reflect differences in underlying endowments of productive resources or differences in technologies. For example, the classic Heckscher-Ohlin theorem in international trade states that each country exports the good that makes relatively more use of the input that it has in greater abundance. Technology differences also lead to specialization. With many goods and many producers, more complex patterns of trade are likely to be necessary to take full advantage of gains from trade.

1.3.5 Producer Marginal Rates of Transformation and Consumer Marginal Rates of Substitution

In the present framework, consumers can own and operate multiproduct production technologies. Consumers also have endowments of final goods. When producer marginal rates of transformation differ from consumer marginal rates of substitution, there are opportunities to change the mix of products produced and to exchange those products with other consumer-producers and with other consumers. Gains from trade are realized when producer marginal rates of transformation move closer to consumer marginal rates of substitution. These type of transactions require coordination of both consumption and production decisions. As before, when there are many producers and many goods, taking full advantage of these types of potential gains from trade likely will require many bilateral trades.

1.3.6 Economies of Scale and Scope

Suppose that consumers have access to a common production technology that exhibits economies of scale. Then consumers will benefit from cooperation to jointly produce the good. This will require coordination across multiple consumers to take advantage of joint production economies. There will be a need to coordinate production and consumption decisions. Gains from trade are realized as a result of lower total costs of production when output production is combined. Additional gains from trade are obtained from increased consumption that can be achieved from lower costs. In addition, lower production costs due to economies of scale permit consumers to produce a greater variety of goods.⁸ A related type of gains

8 This type of gains from trade is observed in models of international trade, such as Krugman (1979).

from trade results from economies of scope that would result from joint production of different products. Gains from trade are realized from lower total costs when products are produced together. Further gains from trade from economies of scope are realized from increased consumption and greater product variety. Coordination of multiple consumers to take advantage of economies of scale and scope will require multilateral trades or a coordinated set of exchanges.

1.3.7 Risk Aversion

Consumers who are risk averse can obtain gains from trade by sharing risk through insurance and other types of risk-sharing contracts. Consumers who face different types of risk can benefit from pooling their risks. Consumers with different marginal rates of substitution across states of the world can benefit by reallocating resources across states through risk-sharing agreements. Gains from trade are realized by narrowing the differences between their marginal rates of substitution across states of the world. More consumers with diverse risks widen the pool and increase gains from trade. Taking advantage of risk pooling requires multiple trades.

1.3.8 Rates of Time Preference

Consumers can obtain gains from trade by exchanging goods over time. Consumers may have different marginal rates of substitution across time periods. These differences may be due to different rates of time preference or to the allocation of endowments. Consumers realize benefits by transferring goods over time and narrowing differences in their marginal rates of substitution. Consumers also realize gains from trade through borrowing and lending. Realizing the benefits from differences in rates of time preference could require multiple trades because consumers will benefit from debts and loans of varying size and duration.

1.3.9 Delegation

A consumer acting as a principal can benefit from delegating authority to another consumer acting as an agent. There are gains from trade for the principal and agent if the agent has some particular skill or expertise that the principal does not have. Total productivity is increased by delegating a productive task to an agent. There are also gains from trade if the principal has a different opportunity cost of time than the agent. The agent can act as the principal's representative in dealing with third parties.

1.3.10 Cooperative Production

Consumer-workers may have complementary skills that allow them to produce efficiently in a team, producing more together than they could produce separately.

Consumer-workers who initially have similar skills may obtain gains from trade from a division of labor and from the expertise that comes from specialization. This is one aspect of Adam Smith's key insight that specialization and division of labor depend on the extent of the market and was exemplified by Smith's classic example of the pin factory. To obtain these benefits of joint production requires consumer-workers to participate in transactions that support joint production. Some types of benefits from specialization and division of labor can be realized simply by trading the output produced by individuals and realizing gains from trade. Other types of benefits of specialization require working together directly to manufacture goods and services.

1.3.11 Public Goods

Public goods have the property that consumption is not competitive, so that multiple consumers can benefit from consuming the same good. For example, multiple consumers can benefit from sharing information. Multiple consumers can enjoy illumination from the same street light. There are also local public goods for which access can be excluded. Public goods provide potential gains from trade because consumers can share the costs of jointly producing a public good instead of each consumer bearing the full costs of separately producing the same public good. For example, it is efficient to produce a movie that is seen by many people, rather than producing a different movie for each person. Obtaining the benefits of a public good requires coordinating the production and joint consumption of the public good so that consumers can share the cost savings. The cost savings from joint production of public goods generate additional gains from trade because they permit increases in the quantity, quality, and variety of the public goods that are provided.

1.3.12 Externalities

When there are externalities there are potential gains from trade. This is the situation described by Coase (1960), who considered bargaining over the extent of externalities. Suppose that there are two consumers, and one creates pollution that harms the other. The pollution provides benefits to the polluting consumer because it is a byproduct of consumption. For example, suppose that the pollution consists of the noise from a loud radio. Suppose that there are no legal restrictions on pollution. Then there are potential gains from trade that can be achieved if the consumer creating the pollution reduces the pollution in return for compensation from the consumer being harmed by the pollution. Alternatively, suppose that there are legal restrictions on pollution if another is harmed. Then there are potential gains from trade that can be achieved if the consumer creating the pollution is able to continue some pollution by paying compensation to the consumer being harmed by the pollution.

1.4 Consumer Coordination and Transaction Costs

Consumer coordination to achieve transaction benefits entails transaction costs. Neoclassical economics generally assumes frictionless transactions. Markets establish prices and allocate goods and services without consuming resources. When transactions are costly, however, the economy can only move toward outcomes that satisfy the traditional neoclassical efficiency conditions. Consumers can achieve some but not all potential transaction benefits. The neoclassical efficiency conditions must be modified to take transaction costs into account.

Consumers expend time, effort, and resources in arranging transactions. The great diversity of institutions of exchange reflects the potential benefits and the relative costs of alternative transactions. Transaction costs are economic forces that deter or prevent exchange. The many type of exchanges are designed to overcome such frictions and obtain gains from trade. The many types of gains from trade require different types of coordination between consumers. Thus, the structure of transactions depends on the underlying transaction benefits and transaction costs. This section examines some of the major categories of transaction costs.

1.4.1 Communication and Information Processing

Consumers must communicate with one another to arrange trades or to form consumer organizations. Even the most basic bilateral trade requires communication and information processing. Forming consumer organizations requires more complex communication and information processing. Consumers must make and receive offers and evaluate economic alternatives. Consumers can experience cognitive limits in processing market information and attempting to make rational decisions based on that information. Simon (1976, pp. 81–84) cautions that there are limits to rationality: incomplete knowledge, imperfect anticipation of future events, and difficulty in identifying all possible alternatives.⁹ Williamson (1985, p. 45) observes that “[b]ounded rationality is the cognitive assumption on which transaction cost economics relies.”

For consumers to achieve transaction benefits often requires either simultaneous multilateral transactions or at least a series of bilateral transactions. In the context of mitigating externalities, Coase (1960) emphasized that the more parties are involved, the greater will be the costs of bargaining. This fundamental insight can be applied to the theory of the firm. The more consumers are involved, the greater will be the costs of arranging bilateral transactions.

With multiple goods and multiple consumers, achieving gains from trade requires multilateral trades or a complex series of trades. Moving toward equality

⁹ Rubinstein (1998) examines diverse attempts at economic modeling of bounded rationality. He considers models that address procedural decision making, defining knowledge, limited memory, choosing what to know, and limits on strategic decision making in games.

of marginal rates of substitution entails a reallocation of initial endowments. Also, consumers prefer greater variety, due to convex preferences. If consumers have sufficiently different endowments, multiple trades realize higher gains from trade.

For many types of transaction benefits, the more consumers that are involved the better. For example, the benefits of risk sharing are enhanced when the number of consumers that pool risks is greater. Multiple consumers are needed to realize the benefits that result from economies of scale. Realizing the benefits of joint production of public goods requires many consumers.

Multiple consumer-workers must be involved for the benefits of joint production to be attained. Consumer-workers can offer complementary skills or realize the benefits of specialization and division of labor. Multiple consumer-workers are needed to achieve the benefits of delegation of authority. Achieving the benefits of joint production is likely to require a complex set of transactions to coordinate the consumer-workers who will cooperate in production.

Further transactions are then needed for the group of consumer-workers engaged in productive activities to trade with other consumers or other groups of consumer-workers who will provide financing, materials and productive inputs. Also, the group of consumer-workers engaged in joint production may need to sell or trade its output with other consumers or other groups of consumer-workers to achieve gains from trade.

The transaction costs of carrying out and coordinating multiple transactions are likely to increase substantially the transaction costs identified for bilateral exchange. The costs of searching for multiple trading partners are likely to increase more than proportionately. Compare the cost of arranging a meeting between two people with the costs of arranging a meeting with more than two people. Scheduling conflicts are likely to multiply.

Consider the potential number of interactions between n people. There are $\frac{1}{2}n(n-1)$ potential contacts, so that the number of potential contacts increases with the square of the number of people. Often this point illustrates the benefits of belonging to a network, because having more members means more potential contacts. However, when contacts are costly, the need to make more potential contacts implies that costs tend to increase with the square of the number of people involved. Thus, communication can become more difficult with a larger group of consumers.

The costs of assembling and managing a team for joint production can be substantial. Finding the members of the team involves the costs of search and negotiation required for group transactions. The costs of putting a group together are likely to increase more than proportionally with the size of the group.

1.4.2 Search and Matching

When there are few consumers, transaction costs tend to be small and consumers can achieve efficient allocations through barter and bargaining. They can transact within families and social groups. They can form cooperatives to coordinate efforts

and share the fruits of productive activities. In contrast, when there are many consumers, the transaction costs of coordination will tend to be large. With many consumers, transaction costs are likely to increase disproportionately, sometimes outweighing the benefits of greater opportunities for trade.

Search costs involve time, effort, and resources. There are costs to finding the right type of trading partner when there are diverse buyers and diverse sellers. Consumers must cope with the effects of asymmetric information that make the characteristics of potential trading partners costly to observe. As a means of dealing with search costs, consumers can incur the costs of forming various types of social and business networks. In addition, consumers may bear the costs of establishing travel and communications networks. The transaction costs of finding trading partners involve both the direct costs of search and the indirect costs of imperfect matches. Product differentiation increases the costs of finding a suitable trading partner. For example, search costs are greater in real estate markets because houses are differentiated products.

The problem of “the absence of a double coincidence of wants” refers to the situation in which such costs are prohibitive and lead to autarky. A consumer with apples who is seeking oranges must find a trading partner with oranges who is seeking apples. Additionally, the consumer must find such a trading partner at the desired time and place. Generally, the benefits from exchange must exceed transaction costs to avoid the problem of the absence of a double coincidence of wants.

1.4.3 Bargaining

Bargaining is necessary for bilateral exchange between consumers. Consumers also need to bargain with each other to establish and manage an organization. Bargaining requires time and effort. Similarly to the labor-leisure choice, individuals will value the time spent on transactions in comparison to their opportunity costs of time. Consumers determine the benefits from goods and services taking into account the convenience of the transactions needed to acquire them. Because bargaining takes time, there is a transaction cost from the delay in consumption. The present value of benefits derived from goods and services will reflect a consumer’s rate of time preference and the time needed to complete the transaction. The cost of bargaining also depends on consumers’ disutility of effort spent on negotiation. Some consumers may enjoy the social interaction in market exchange more than others. Consumers may differ in the effectiveness or productivity of effort devoted to transactions. Some consumers are more skilled at search or bargaining.

Bargaining costs have efficiency implications. Consumers will forego some transactions if bargaining costs exceed gains from trade. For more complex negotiations, bargaining costs can result in incomplete contracts that imperfectly share risks. Costly bargaining can lead to mistakes in transactions. The costs of bargaining may prevent the formation of some types of consumer organizations or limit the size of such organizations.

1.4.4 Moral Hazard

Consumers engaged in bilateral exchange encounter transaction costs due to moral hazard. Consider a consumer who contracts with another consumer to produce a good. Suppose that the effort of the producer is unobservable. The transaction may be inefficient relative to the outcome under full information. For example, in the moral hazard model of agency, the principal pays the agent to carry out a designated task and the agent decides how much effort to devote to the task. The principal hires an agent because she does not have the time or capability to do the work herself. The principal's problem is to design an incentive schedule that motivates the agent to choose a desired level of effort.

The principal must rely on performance-based rewards such as bonuses and commissions to induce the agent to work. The contract could potentially induce the agent to devote an efficient level of effort by allowing the agent to keep all the returns to his effort. Yet such a performance-based rewards system has the significant drawback that it shifts risk to the agent. Suppose that the principal is risk-neutral but the agent is risk-averse. The principal needs to compensate the agent for the cost of risk-bearing to induce him to enter the relationship. Accordingly, to reduce the risk borne by the agent, thus reducing the cost of compensating the agent for that risk, optimal contracts consist generally of a fixed payment plus some performance-based rewards. Sharing output between the principal and the agent necessarily results in some shirking because the agent's rewards do not capture the full effects of his effort.

1.4.5 Adverse Selection

Asymmetric information creates various types of adverse selection problems. These problems affect the efficiency of bilateral exchange between consumers. Akerlof's (1970) model of the "market for lemons" showed that bad cars can drive out good cars. In product markets with unobservable quality, buyers will make their participation decisions on the basis of expected quality.

Asymmetric information can affect the extent of trade. When a buyer and a seller have bilateral asymmetric information, they may decide not to trade even though the buyer's willingness to pay exceeds the seller's cost; see Myerson and Satterthwaite (1983). Asymmetric information affects bilateral exchange between an investor and producer. Diamond (1984) considers a model in which investors make loans to producers who observe the outcome of an uncertain investment project. Producers have an incentive to understate the outcome of the investment project. The optimal contract between an investor and a producer takes the form of debt. The debt contract provides incentives for the producer to report the outcome of the investment project.

Principal-agent transactions are subject to adverse selection problems. The principal designs the menu of contracts so that at a separating equilibrium she will

know with certainty the agent's preferences by observing the agent's choice from the menu. Thus, contracts are self-selecting and the agent will end up revealing his type. Each contract in the menu consists of a payment and an effort level and it is tailored to one of the agents' types. In the case of two types of agents, the contract written for the low-cost agent provides a large payment in exchange for large effort. The contract written for the high-cost agent has lower payment but also requires lower effort. If both contracts gave the same payment, both types of agent would choose the same contract, the one that required less effort. The high-cost agent ends up getting the income that he could get elsewhere in the market, his reservation utility. The low-cost agent obtains informational rents to induce revelation of private information.

1.4.6 Free Riding

The costs of group decision making are likely to be substantially greater than the costs of bilateral negotiation. Achieving a consensus becomes more difficult when many people are involved. As the number of parties to the negotiation increases, different devices such as committees and voting may be required, with all the attendant inefficiencies of collective decision making.

Even after the team is assembled, coordination requires group negotiation costs. In addition, the group will face the costs of selecting and rewarding leaders and the possibility of malfeasance by the leaders. Forming a hierarchy to deal with group coordination entails problems that result from moral hazard, adverse selection, and collusion on the part of subordinates.

An association of consumers is likely to engage in noncooperative behavior that results in free riding. Associations of consumers include buyers' cooperatives, workers' cooperatives, and partnerships. Consumers may offer contributions to the group that fail to maximize joint benefits. They may shirk as a best response to anticipated equilibrium behavior by other group members. Consumers may also misrepresent their preferences as a best response to anticipated equilibrium disclosure of other group members.

An important source of transaction costs for group decision making is due to the budget balancing constraint. An allocation mechanism for a group of consumers must satisfy a balanced budget. In many cases, such break-even mechanisms fail to achieve efficient allocations.

1.4.7 Contracting

Buyers and sellers face a contracting process with several stages, each of which brings its own transaction costs. Transaction costs can introduce inefficiencies at each stage of the process. Moreover, transaction costs in one stage of the contracting process can affect decisions and efficiency at another stage in the contracting process. It is useful to consider three general stages of contracting. First, potential

contracting partners must engage in costly search to find each other. Second, potential contracting partners must negotiate and formalize the terms of the contract, which entails communication costs, the time costs of bargaining, and the costs of writing the contract. A buyer and seller who face bargaining costs may negotiate an incomplete contract that yields lower expected gains from trade than would the complete contingent contract that might be chosen in the absence of transaction costs. Third, after the contract has been formed, the contracting partners must make decisions about performance of the agreed-upon terms, monitoring of performance, settlement of payments, verification of states of the world, and possible contract renegotiation. In the event of a breach of contract, there are legal and transaction costs of contract enforcement.

There has been considerable discussion in economics of the problem of contract hold-up. Although contract law enforces efficient contracts, it is not always possible to write complete contingent contracts. One of the parties may seek to renegotiate the agreements after the other party has made an investment in the relationships. When such investments are specific to the relationship, renegotiating contract terms takes advantage of relationship-specific investment. In making investments that are relationship-specific, the parties will anticipate the possibility that the agreement will be renegotiated. This can lead to inefficient investment.

The standard underinvestment result arises from the inability of the parties to make a binding commitment; see for example Groux (1984). The buyer and seller division of the ex post returns to trade lowers the marginal returns to investment below the ex ante returns to investment. Also, the value of the outside option for the buyer and seller enters into their investment decisions, as shown by Hart and Moore (1990). The return to investment is lowered because it is a weighted average of the outside option and the ex ante return, and the marginal effect of investment on the outside option is lower than that of the ex ante return. Moreover, the externality effects operating through the outside option functions influence the equilibrium investment levels of the buyer and seller.

1.5 Consumer Organizations and the Separation Criterion

Consumers form all sorts of transactions, contracts, and organizations that should not be classified as firms. These institutions may perform some functions that are similar to those carried out by firms, including intermediation of transactions. These organizations may be precursors to firms and may evolve into firms. However, many types of consumer organizations make decisions that are often not separable from the interests of consumers who establish, own, manage, or are members of the organization.

An institution satisfies the *separation criterion* if the objectives of the institution can be distinguished from those of its owners. This section applies the separation criterion to demonstrate that consumer transactions and organizations are not

firms. Chapter 2 further develops the theory of the firm based on the separation criterion.

1.5.1 Autarky

Under autarky, the consumer's production and consumption decisions are intertwined. The consumer produces a mix of outputs and combines inputs to maximize his net benefits subject to his production possibilities. Therefore, autarkic production by the consumer is not a firm, based on the separation criterion.

Autarky, as applied in the present discussion, refers to consumers who choose not to exchange goods with others or to participate in consumer organizations. Consumers remain on separate islands and fend for themselves, as did Robinson Crusoe.¹⁰ Consumers can choose to avoid exchange by consuming their initial endowments. Because consumers own and operate production technology, they can be producers as well. In addition, consumers can develop their own production technology. Autarky can refer to production and consumption by families and extended households. Autarky provides a natural benchmark for evaluating gains from trade between consumers, as well as gains from trade once firms are introduced into the economy. Market exchange is voluntary because consumers have the option of autarky.

Although autarky is a useful theoretical construct, it has rarely been observed on a large scale. The most primitive of village communities engage in sharing of goods through gift exchange, feasts, cooperative production, and trade. Throughout recorded history, individuals have traded goods at the local, regional, and even international level, as the discussion of merchants in the next chapter demonstrates.

1.5.2 Exchange Transactions and Contracts

Consumers together can form bilateral or multilateral transactions. Groups of consumers engaged in multilateral exchange, or in multilateral contracts, do not constitute firms. There is no independent economic institution with objectives that are distinct from those of the parties to the transactions. The transactions are simply a group of economic exchanges between consumers.

The parties to a transaction or contract often have conflicting interests. A buyer is maximizing his consumption interest and a seller is maximizing his earnings. Transaction costs aside, the buyer wishes to obtain the maximum net benefit $v - p$ and the seller wishes to obtain the maximum earnings $p - c$. Speaking informally, the buyer and the seller jointly own the transaction. The objective of the transaction is the combination of the objectives of the buyer and the seller and cannot be separated from them.

10 Robinson Crusoe is the character in the classic novel of the same name by Daniel Defoe, who finds himself "some hundreds of leagues out of the ordinary course of the trade of mankind."

Contracts between consumers, that is, transactions with a promise to provide a good in the future, also do not constitute firms. Jensen and Meckling (1976) refer to the firm as a “nexus of contracts.” Although firms are centers of many contracts, not every nexus of contracts is a firm. Consumers form bilateral or multilateral contracts without intermediation by firms. Principals can form contracts with agents. Collections of contracts, no matter how complex they are, do not constitute independent economic actors. They represent the offers and acceptances of the parties to the contracts. In a bilateral contract, the parties to the contract divide the gains created by the exchange. Each of the parties seeks to maximize their share of the value created by the exchange. The contract reflects the objectives of the parties. Some contracting costs can be alleviated by intermediaries that separate the interests of the contracting parties, as the discussion in [Chapter 11](#) emphasizes.

1.5.3 Business Enterprises

Consumers establish business enterprises in any sector of the economy, including manufacturing, agriculture, and commerce. Consumers become entrepreneurs by establishing firms, as will be discussed in later chapters. In the contemporary economy, business enterprises are likely to be organized as firms with objectives separate from those of their owners. A consumer chooses to establish an enterprise as a firm to obtain the benefits from separating the firm’s decisions from his consumption decisions. The advantages of separation are discussed further in the next chapter.

A consumer enterprise that is managed as a combination of consumption interests and commercial activities should not be viewed as a firm. The enterprise may be a family farm, a service provider, a small manufacturing business, or a merchant enterprise. The consumer may adjust his labor effort, his investment, and the output of the enterprise on the basis of consumption needs as well as commercial returns. The enterprise is not a firm when the consumer connects business decisions with consumption decisions.

A family business is not a firm when it does not meet the separation criterion. Historically, most family enterprises were not firms as defined here. The business activities of their owners were closely connected to their consumption activities. Agricultural, artisanal, and merchant enterprises traditionally were intertwined with the personal activities of the merchant. Even highly developed merchant banking enterprises, such as that of the Medicis, were closely tied with the personal finances, consumption, artistic endeavors, political activities, and religious contributions of the merchant’s family.¹¹ The business was part of the family, which supplied labor, management, and capital to the enterprise. The objectives of the family businesses were closely tied to the consumption objectives of the family that operated the business.

11 See Parks (2006).