
A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
FELIX CHRISTIAN HERBERT GOTTSCHALK

Master of Arts UZH in Wirtschaftswissenschaften,
Volkswirtschaftslehre
born on 18 October 1986
citizen of Germany

accepted on the recommendation of
Prof. Dr. Wanda Mimra, examiner
Prof. Dr. Massimo Filippini, co-examiner

2017
Abstract

This thesis analyses problems in credence goods markets with respect to their extent, their determinants and possible solutions to them. When purchasing credence goods, consumers have to trust in the accurateness of information provided by sellers. Sellers of credence goods may use this information asymmetry to their own advantage by, for instance, selling more or different services than in the best interest of the consumer. This market failure harms consumers and induces economic costs to the society. Economically significant examples of credence goods markets are health care markets and markets for financial advice – which this thesis is applied to. Because of their relevance, and because the typical problems of credence goods are frequently observed in these markets, they are at the centre of political discussion. Researching them contributes to a better understanding of the necessities and consequences of political intervention and, eventually, to better inform policy makers. Research on credence goods markets is thus highly relevant.

This is a cumulative dissertation with four chapters in which different research methods are used. In chapter 1, the research literature on credence goods is categorized and critically discussed with respect to the definition of credence goods. Chapter 2 presents a field experiment in the dental care market, providing a unique micro-level data set for the analysis of the extent of overtreatment and its determinants in a credence goods health market in an industrialized country. In chapter 3, a theoretical work on the connection between the problems in credence goods health care markets and the prevention decisions of patients is presented. Chapter 4 presents a laboratory experiment in which a new theory about markets for financial advice and the problem of biased advice is tested for the first time and in which the effects of different policies are compared.
Kurzbeschreibung


Acknowledgments

Conducting the research presented in this thesis at ETH Zurich in the last four years was a challenging, but also a rewarding experience, or with Charles Dickens: “It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness”.

I want to thank the people who have contributed to this thesis in many different ways:

Wanda Mimra: I thank her for giving me the chance to start my PhD at her chair – and the corresponding life-long honour of being her first PhD student; for the freedom to pursue my own research interests and for support and guidance in every direction I took; for her constant availability and the detailed feedback on my work; and finally, for support in my experimental project.

Christian Waibel: for introducing me to the research area of credence goods; for providing countless helpful hints and suggestions; for patiently answering the thousands of questions I posed to him. I further thank him for the health economics reading group and for his positive attitude to life.

Both Wanda Mimra and Christian Waibel for their collaboration in the “dentist field experiment” - a memorable project in many respects! If this thesis needed a one-word-summary, it had to be: “dentists”.

Prof. Massimo Filippini for co-supervising this thesis, Prof. Hans Gersbach for chairing the examination committee and for supervising my master’s thesis back in 2012 which stimulated my interest in research.

The people at the chair: Pierre Fleckinger for helping me to start my first project; my office-mate Anastasia Sycheva who took care of my plants; Sabine Keller; our student assistants; and of course our test patient.

I thank my friends and PhD confederates who shared their insights on doing a PhD, economics, life in general and in the specific – in the café bqm and elsewhere. I thank the Swiss electorate for the ETH and the administration of the ETH for its efficiency.

Last, I want to thank my family: my parents who have always supported me during my studies and long before them; my sister Katharina and my brother-in-law Thomas who have visited us in Zurich frequently in the last years; and I thank my fiancée Stefanie – for contributing in so many ways to my life and thereby to my dissertation (and to a considerable amount of distraction in my first year at ETH)!
Contents

Introduction VII

1 What characterizes Credence Goods? A critical look at the literature. 1
  1.1 Introduction 1
  1.1.1 The diversity of credence goods 2
  1.1.2 Credence characteristics versus credence goods 5
  1.1.3 Credence goods problems 8
  1.2 Different characterizations of credence goods 10
    1.2.1 A classification of characterisations in the literature 10
    1.2.2 Consequences of different credence goods characterisations 13
  1.3 Conclusion 15

2 Health Services as Credence Goods - A Field Experiment. 17
  2.1 Introduction 17
  2.2 Related literature 22
  2.3 The market for dental care in the canton of Zurich 25
  2.4 The field experiment 26
    2.4.1 Case & test patient 26
    2.4.2 Experimental design 28
    2.4.3 Experimental procedure 30
  2.5 Comprehensive data on market, practices and dentists 32
    2.5.1 Data sources and variables of interest 32
    2.5.2 Descriptives 35
  2.6 Results 37
    2.6.1 Treatment recommendations 37
    2.6.2 Diagnoses 47
  2.7 Conclusion 52
  2.8 Appendix 54
    2.8.1 Additional figures and photos 54
    2.8.2 Population of dentists 56
2.8.3 Detailed information on the recruitment and training process of the test patient ........................................... 56
2.8.4 Covariates and model fit .............................................. 57

3 Why prevent when it does not pay? Prevention when health services have credence qualities. 59
3.1 Introduction ............................................................... 59
  3.1.1 Motivation ............................................................ 59
  3.1.2 Related literature ..................................................... 62
3.2 Model .......................................................... 65
  3.2.1 Agents, actions and information ................................ 65
  3.2.2 Timing and solution concept ................................... 67
3.3 Benchmark I: no market for treatment (nM) ......................... 68
3.4 Benchmark II: no asymmetric information (nA) ...................... 69
3.5 The credence goods market ........................................... 73
  3.5.1 Institutional settings and supply-side moral hazard ............ 73
  3.5.2 Patient behaviour with supply-side moral hazard ............. 77
  3.5.3 The physician’s price choice ..................................... 81
3.6 Welfare & policy ......................................................... 84
  3.6.1 Welfare ............................................................... 85
  3.6.2 Policy ............................................................... 86
  3.6.3 Social health insurance ............................................ 89
3.7 Discussion ........................................................................ 91
3.8 Conclusion ....................................................................... 93
3.9 Appendix ........................................................................ 95
  3.9.1 Prevention and physician demand ................................ 95
  3.9.2 The physician’s price choice ...................................... 96
  3.9.3 Comparison of prices between settings ......................... 98
  3.9.4 Price regulation ....................................................... 99
  3.9.5 Social insurance ...................................................... 103
  3.9.6 Price policy vs. social insurance in (L/noV) .................. 110

4 Advice markets. An experimental investigation. 113
4.1 Introduction ............................................................... 113
4.2 Related literature ........................................................ 117
4.3 Theoretical framework .................................................. 120
  4.3.1 The setup ............................................................ 120
  4.3.2 Equilibria in the experimentally tested market settings .... 125
4.4 Experiment ................................................. 130
  4.4.1 Experimental design .......................... 130
  4.4.2 Parametrization & predictions .............. 134
  4.4.3 Experimental protocol ....................... 136
4.5 Results .................................................... 138
  4.5.1 Experimental conditions ..................... 141
  4.5.2 Advisor and consumer behaviour in detail .... 146
  4.5.3 Market welfare & agents’ payoffs .......... 150
  4.5.4 Developments over time ..................... 152
  4.5.5 Strategic sophistication ...................... 152
4.6 Discussion & Conclusion ............................ 153
4.7 Appendix I: additional figures .................. 157
4.8 Appendix II: regression tables ................... 161
4.9 Appendix III: theoretical framework ............ 169
  4.9.1 Undisclosed commissions with naive consumers 169
  4.9.2 Undisclosed commissions with wary consumers 170
  4.9.3 Disclosure ...................................... 171
4.10 Appendix IV: instructions .......................... 173

5 Bibliography ................................................. i

List of Tables ............................................. xvii

List of Figures ............................................ xix

Curriculum Vitae ......................................... xxi
Introduction

When making purchase decisions in credence goods markets, consumers have to rely on information provided by experts. Credence goods markets can be found all over the economy: health care, repair services, legal and financial advice and management consulting are among the most significant examples. All of these markets have in common that – in many situations – consumers are not able to evaluate whether information provided by the expert is accurate or not, even after purchase.\(^1\) In health care, patients may not be able to judge whether a hip replacement was necessary or not. Likewise firms who engage management consultants may be unable to evaluate whether a consulting project was necessary or not. When purchasing investment products, consumers may not be able to evaluate whether the bad performance of a product has to be attributed to bad luck or to a wrong recommendation. Likewise, professional investors may find it hard to judge whether a credit rating provided by a rating agency reflected the true default risk of a company. Last, consumers often cannot judge whether information provided on consumer products – for instance fair-trade labels on food – convey the truth or not. Credence goods markets are of high economic importance. Health care markets alone produce more than 10% percent of GDP in most industrialized countries (OECD, 2016). Their share of GDP has been growing for decades and is likely to continue to grow (Chernew and Newhouse, 2012). The fraction of GDP produced by financial markets in the US has increased from around 3% in 1945 to 9% in the 2000s (Philippon, 2015), an observation that holds for most industrialized countries (Kerschbamer and Sutter, 2017).

The general problem in credence goods markets through all sectors is that experts may face incentives to reveal information which is not in the best interest of the consumer, leading to consumer harm and to reduced welfare through misallocation of resources. Such provider fraud is widely recognized.\(^2\) The financial crisis 2007/08

\(^1\)Credence: Belief in or acceptance of something as true (Oxford Living Dictionaries).
\(^2\)Provider fraud in credence goods markets occurs in different forms, depending on the incentives of the seller and the institutional framework of the market. The most common problems are selling more services than needed (overtreatment), charging for more services than provided and selling less or less fitting services than needed (undertreatment/biased advice)
serves as a recent example. With respect to health care, it is estimated for the US that up to 10% of all health care expenditures—a sum of about $180B per year—are due to supplier fraud (FBI, 2011). Not surprisingly, credence goods markets like health care and financial advice markets are frequently subject to discussions about appropriate regulation. Yet, despite problems in credence goods markets are recognized, it is not easy to solve them, because research on credence goods markets faces severe methodological challenges.

Theoretical research on credence goods markets was initiated by Darby and Karni (1973). Following theoretical work helped to understand the different sources and natures of problems in credence goods markets and to identify factors which may impact market outcomes (Pitchik and Schotter, 1987; Emons, 1997; Dulleck and Kerschbamer, 2006). However, theoretical predictions from many of these models did not match the widely held perception across industries and countries that expert fraud in credence goods markets is a persistent problem.3

The empirical analysis of the problems of credence goods markets poses severe challenges to researchers, however. In many cases it is almost impossible to determine the extent and the determining factors of seller fraud using aggregate naturally-occurring data. One reason is that data in many credence goods markets is scarce, for instance in health care. But even when data is available, the asymmetric information which characterises the expert-consumer relationship in credence goods markets—the expert knows whether he provides the correct information while the consumer does not—typically persists in the data-researcher relationship: researchers are not able to infer the accurateness of the information provided by experts for single cases from aggregate data, for instance whether a hip replacement was indicated in a specific case or not. This poses a major challenge to the identification of the extent and determining factors of expert fraud in credence goods markets.

As a result, authors have begun to generate their own data by the means of experiments.4 Laboratory experiments (Dulleck et al., 2011; Mimra et al., 2016b) allow researchers to make causal statements and to observe the behaviour of real people in the desired context. To observe the behaviour of real people is particularly important in credence goods markets where unethical behaviour and trust are crucial elements, because psychological factors and non-standard preferences are likely to be important in such environments (Kerschbamer et al., 2017). The lack of naturally-occurring data for the analysis of credence goods markets has further been

---

3For instance, Dulleck and Kerschbamer (2006) find that under fairly general assumptions, credence goods markets should be able to solve the suspected problems endogenously.

4See List (2007) for a review on experimental methods in economics.
Introduction

approached by the conduction of (natural) field experiments. Field experiments are experimental studies conducted in real world environments and therefore “combine the most attractive elements of the lab and naturally-occurring data: randomization and realism.” (List, 2007). Field experiments in credence goods markets have provided insights about the extent of expert fraud and possible influencing factors where before there had been only anecdotal evidence. In these field experiments, expert fraud can be identified in every single data point, contrasting aggregate data for which this is not possible.

This thesis reflects the approach taken by the literature on credence goods. It analyses problems and their determinant in credence goods markets by employing a broad tool set of research methods. I present a natural field experiment (chapter 2), theoretical work (chapter 3) and a laboratory experiment (chapter 4). In chapter 1, I further provide a critical overview of the distinct lines of literature which refer to credence goods. The work in this thesis is applied to health care and financial advice, two of the most economically significant credence goods markets.

In the first chapter, I present a categorization of distinct lines of literature referring to credence goods into markets with expert providers, expert advisors and information experts. The classical credence goods literature has focussed on expert providers who provide both diagnoses and services (Dulleck and Kerschbamer, 2006). Important examples are repair services (cars, computers) and health care. More recently, a line of literature which focusses on expert advisors has evolved (Inderst and Ottaviani, 2012a). Such experts advice consumers about which product(s) provided by third-parties best fits their needs. The main example for such markets are markets for financial advice and health care markets in which commission payments reach considerable sizes. The chapters in this thesis consider expert providers (chapters 2 and 3) and expert advisors (chapter 4). This thesis hence combines two lines of literature which are typically analysed in isolation.

In chapter 2, me and my co-authors Wanda Mimra and Christian Waibel present the first field experiment of the credence goods literature conducted in an industrialized health care market, the Swiss market for dental care. We report dentist

---

5See Schneider (2012) for car repairs, Balafoutas et al. (2013) for taxi rides, Kerschbamer et al. (2016) for computer repairs.
6Therefore, field experiments in credence goods markets are part of the recently termed meta-field of “forensic economics” (Zitzewitz, 2012).
7Examples for the third category – goods sold by information experts – are good containing non-verifiable product information such as fair-trade fruit or low-emission electricity. I do not explicitly analyse markets with information experts in this thesis.
8I presented this paper at: SSPH+ Science Flash Talk 2017 (Lugano), Spring Meeting of Young Economists 2017 (Halle), dggoe conference 2017 (Basel), International Health Policy Conference
behaviour under varying patient characteristics and under control of a large number of market and dentist characteristics. The data was collected in 180 dentist visits of the same undercover test-patient, yielding a unique micro-level data set. Our main findings are that overtreatment occurred in more than 25 percent of the visits and that low short-term demand, indicating excess capacities, leads to significantly more overtreatment recommendations. The chapter furthermore adds insights with respect to the role of socio-economic characteristics in the patient-physician relationship and patient information in the age of the internet. The paper contributes to both the general credence goods literature and the more health related literature on physician-induced-demand (PID). It stands out in the PID-literature by employing a micro-level data set with a clear definition of the appropriate treatment for every data point, contrasting many studies which analyse aggregated data in which it is challenging to identify whether treatment decisions were appropriate or not (see Leonard et al. (2009)).

In chapter 3, I present a theoretical study on health care credence goods markets with prevention. The interaction of prevention and the credence goods information asymmetry have not yet been analysed, although prevention activities are a characteristic of many health care markets. This is even more surprising, because physician supply-side moral hazard in the form of overtreatment or overcharging should reduce prevention incentives of patients by making physician visits in the absence of severe health problems more expensive. I use the framework of Dulleck and Kerschbamer (2006) to show that markets with typical credence goods problems are characterised by inefficiencies with respect to treatment and prevention decisions. Patients with high health risks prevent inefficiently little, while patients with lower health risks prevent inefficiently much. These results provide a new point of view for explaining why inefficient levels of prevention are observed in many real-world health markets. Moreover, I present a comparison of markets in different institutional settings in which supply-side moral hazard appears either as overtreatment (when the institution verifiability holds) or as overcharging (when the institution liability holds, but verifiability does not). I show that both markets are characterised
by different monopoly prices and correspondingly by different levels of prevention in the society. I show exemplary for the two policies price regulation and compulsory health insurance, that institutional settings potentially have important impacts on the potential of policies and on the question which policy is suited best to improve market outcomes.

In chapter 4, I present a laboratory experiment on credence goods markets with expert advisors including third-party providers. The credence good characteristic in these markets concern the consumers’ inability to tell whether a product recommendation was in their best interest or not, because even if the recommendation turns out wrong, it may still be attributed to bad luck (for instance unforeseeable market developments). Advisors in real world markets often face conflicts of interest as providers of investment products pay commissions to financial advisors, and pharmaceutical companies make payments to physicians and pharmacists. Hence, advisors may be inclined to give advice in favour of products with high commissions instead of acting in the best interest of the consumers (biased advice). With the financial crisis in 2007/08, markets for financial advice have become the focus of regulation. A highly debated policy issue regards the disclosure of conflicts of interest. The topic of disclosure is also highly debated in health care markets. My experiment represents the first market experiment to test the theory of Inderst and Ottaviani (2012a) who provide a model for markets with advice, third-party payments and credence goods. In this setting, the more cost-efficient of two providers has incentives to pay higher commissions than the less cost-efficient provider, inclining the advisor to give biased advice. I compare the policy of disclosure with a theoretically equivalent alternative, advisor-fines for wrong advice. The experimental results yield insights into how unregulated and regulated experimental markets for advice work compared to theory. A main finding is that disclosure of commissions does not work in the predicted way through lower commissions, but because consumers accept more recommendation with disclosure than without. Moreover, consumers are better off than predicted even when they do not expect biased advice in the first place. These and other results show many indications for psychological traits which are not considered by theory. The study contributes to a highly policy relevant area and hopefully will be able to encourage more detailed research.

11 As an illustration for the importance of such third-party payments, pharmaceutical companies in the USA have paid USD 8.18B to 631,000 physicians and 1,146 hospitals in 2016 (CMS).

I present an overview of the examples for credence goods used in different lines of literature. I categorize the examples into markets with expert providers, expert advisors, information experts and no expertise. I discuss the classic characterization by Darby and Karni (1973) (DK73) and how different modelling approaches used in the literature relate to it. I show that recently authors have characterized credence goods in different ways, some of which are inconsistent with the characterization in DK73. I discuss the associated problems.

1.1 Introduction

Every good has diverse characteristics – also called aspects or qualities – which may be important to a consumer. Characteristics of a bottle of wine are its design, its taste and its processing attributes. Characteristics of a dental tooth filling are the location of the practice, the kindliness of the staff, the comfort during the surgery and the quality of both the diagnosis and the filling. The characteristics of goods can be categorized with respect to the time when consumers can evaluate them. Search characteristics are evaluable before the purchase and experience characteristics are evaluable after purchase - this distinction dates back to Nelson (1970). Credence characteristics – added by Darby and Karni (1973) (DK73) – are not evaluable even after purchase. The original characterisation by DK73 states:

Search qualities are those that can be ascertained in the search process prior to purchase and experience qualities are those that can be discovered only after purchase as the product is used ... Credence qualities are those which, although worthwhile, cannot be evaluated in normal use.

(Darby and Karni, 1973) (pp.68-69)
Chapter 1

Some remarks are in place:

- The characterisation simply demands that the consumer is not able to evaluate credence qualities.
- “Qualities” refers to characteristics, elements or aspects of a good which are of interest to the consumer. This interest stems from a consumer’s “want” (DK73, p.67) routed in a preference or in a need to maintain a condition (e.g., a healthy condition).
- I use the term characteristics instead of qualities in order to separate the term from what is used to describe how well a good is crafted.
- DK73 speak of credence characteristics rather than credence goods. The term credence good appears for the first time only on the fifteenth page of DK73 (p.81). The term credence good is used as an illustrative reference to goods with credence characteristics when these characteristics are central to our analytical interest in the good.\(^1\) I will use the term credence good in this sense throughout this thesis.
- The characterisation provides a clear-cut good categorization of characteristics into either search, experience or credence characteristics.

In the decades following DK73, different lines of literature referring to credence goods have evolved. In this short article, I will categorize them and discuss recent modelling approaches in light of the characterisation by DK73.

1.1.1 The diversity of credence goods

**Four categories of credence goods.** I classify credence goods into four categories as presented in table 1.1. The first three categories are markets in which expert sellers typically possess an informational advantage over consumers. These three categories have in common that the consumer may not be able to verify information provided by the seller and hence has to be believe that the provided information is true. In the fourth category both sellers and consumers are not able to evaluate certain product benefits and hence both have to believe in a presumption about these benefits. The term “credence” originates from these observations – credence: “belief in or acceptance of something as true” (Oxford Living Dictionaries).

First, there are goods provided by expert providers who combine the diagnosis of the consumer’s need with the provision of services which can satisfy this need.\(^2\)

\(^1\)A related thought was expressed by Andersen and Philipsen (1998): “Credence goods are goods for which the buyer’s decision-making is dominated by concerns about credence characteristics.”

\(^2\)“Diagnosis and treatment” in the wording Dulleck and Kerschbamer (2006).
1.1. Introduction

TABLE 1.1
A categorization of credence goods.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>PROVIDER SELLS</th>
<th>SELECTED EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Expert providers</td>
<td>Diagnosis and service</td>
<td>· Repair services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Health services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Management consulting</td>
</tr>
<tr>
<td>(2) Expert advisors</td>
<td>Diagnosis and 3rd-party products</td>
<td>· Advice (financial, health)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Rating agencies</td>
</tr>
<tr>
<td>(1) &amp; (2): Consumers cannot evaluate the seller’s representation of the consumers need.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Information experts</td>
<td>Information and product</td>
<td>· Products with fair-trade labels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Organic food</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Journalism</td>
</tr>
<tr>
<td>(3): Consumers cannot evaluate the seller’s representation of product characteristics.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) No expertise</td>
<td>Products/services only</td>
<td>· Preventive care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Medication during pregnancy</td>
</tr>
<tr>
<td>(3): product performance cannot be evaluated.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The two major examples in this category are repair services and health services. As noted by Mimra et al. (2016a), many of the goods in the first category involve the so-called “professional services’ (or ‘liberal professions’)). These markets have been the focus of a number of important theoretical contributions (Pitchik and Schotter, 1987; Wolinsky, 1993; Taylor, 1995; Emons, 1997; Pesendorfer and Wolinsky, 2003; Fong, 2005). Dulleck and Kerschbamer (2006) provide a survey and a unifying modelling framework for market settings with expert providers to that date. More recently, the literature was expanded by a number of experimental contributions (Dulleck et al., 2011; Currie et al., 2011; Schneider, 2012; Balafoutas et al., 2013; Mimra et al., 2016a,b; Kerschbamer et al., 2016; Balafoutas et al., 2017; Rasch and Waibel, 2017, forthcoming). Chapters 2 and 3 of this thesis are concerned with markets with expert providers.

Second, there are services of expert advisors who only advice a consumer about which product or service best fits his or her need, whereas the products or ser-

---

3Newer theoretical contributions include Liu (2011); Fong et al. (2014); Bester and Dahm (2017) and applied work as referenced in table 1.2.
vices traded in these markets are provided by third-parties. This literature has evolved more recently in the wake of the financial crisis 2007/08 and focusses on financial investment advisors who advice consumers about which third-party investment product(s) they should purchase (Inderst and Ottaviani, 2012a,b,c; Inderst, 2015). Another example are rating agencies (see, for instance, Becker and Milbourn (2011)), which have received considerable attention since the financial crisis. The category also applies to many health care markets in which physicians and pharmacists advice patients about which pharmaceutical product best fits their needs. Chapter 4 of this thesis falls into this category.

The credence characteristic in markets with expert providers and expert advisors is that consumers do not know whether the expert correctly stated their needs or not, i.e., consumers may not be able to evaluate the “seller’s representation of the buyers condition” (Hubbard, 1998), even after purchase. This in turn can have several sources. For instance, when a patient has an aching stomach and his appendix gets removed: after the surgery, a patient may not be able to tell if the appendix needed removal or not, because this need does not influence his or her experience (DK73). Another source is a stochastic outcome. For instance, a consumer may not be able to distinguish whether the bad performance of an investment product is due to bad advice or bad market developments.

The third category consists of goods which are produced by information experts who provide non-verifiable information to consumers, often in the form of labels. This category includes consumer products for which consumers have certain desires about the hidden processing characteristics, for instance with respect to truthful reporting (journalism, science), environmental balance, worker compensation (fair-trade labels) or animal breeding conditions. This literature has grown in the last two decades (Andersen, 1994; Feddersen and Gilligan, 2001; Grolleau and BenAbid, 2001; Vetter and Karantininis, 2002; Baksi and Bose, 2007; Roe and Sheldon, 2007; Bonroy and Constantatos, 2008; Gabszewicz and Resende, 2012; Bonroy et al., 2013; Baksi et al., 2016). In this category, consumers know what they need, but cannot evaluate whether they received the (hidden) characteristics they desire.

The fourth category presented in table 1.1 has very few references: goods for which neither sellers nor consumers can evaluate certain characteristics. There is

---

4 The line between expert providers and expert advisors may not always be sharp in reality. For instance, physicians may be both expert providers and expert advisors depending on the case.

5 Depending on the integration of the market, seller and producer can be the same or two distinct identities.

6 Also in markets with expert providers, consumers may not be able to verify whether they have received a product or service with certain characteristics (Dulleck and Kerschbamer (2006) call this “no verifiability”). Such goods thus involve multiple credence characteristics.
1.1. Introduction

no expertise for goods in this category. Such goods may exist when consumers and sellers have (unproven) positive expectations about the performance of a product, e.g., new medications in health care. Crow et al. (2002) state with respect to credence goods that “their characteristics may not be fully or reliably apparent” and refer to preventive care. While the authors do not go into further detail, I may carry their example on by thinking of preventive care or medical procedures with suspected long-term benefits which (as yet) lack scientific evidence. Another example is pregnancy, where scientific evidence on the effects of medication is lacking, because ethical reasons prohibit research. In the words of Eddy (1984): “there is no way to shorten the time needed to observe ten-year survival rates, and there is no way to increase the frequency of rare diseases”. Certainly there is a broad middle-ground between the two extremes of no evidence and the other extreme of full evidence on the other side. The credence characteristic in the forth category is simply that the benefits of such goods are not known. Goods in this group are not traded with information asymmetries between sellers and buyers - typical credence goods problems are therefore presumably not as important as in the other categories. Further research in the area is needed to test this presumption.

The four categories of credence goods are handled by distinct lines of literature, although all these goods are correctly termed credence goods in the sense of DK73 as they involve characteristics which cannot be evaluated after purchase.

In table 1.2, I present a detailed list of examples for goods with important credence characteristics, credence goods, which have appeared in the literature.

1.1.2 Credence characteristics versus credence goods

The characterisation by DK73 refers to characteristics instead of goods. Yet most authors – including DK73 and me in this article – use the illustrative term “credence good” to refer to goods with credence characteristics when these characteristics are at the centre of our analytical interest. Yet most, if not all goods in the world exhibit search, experience and credence characteristics simultaneously. Furthermore, the category of a particular characteristic can change over time and over situations. This section illustrates these points by examples.

Goods have search, experience and credence characteristics. Goods typically combine search, experience and credence characteristics. A bottle of wine has a design, a taste and several processing-attributes. The design of a bottle can be evaluated before purchase and therefore represents a search characteristic. The

---

7 This is reflected in the discussion on evidence-based medicine; see for instance Smith (1991).
## Table 1.2
Examples of credence goods in the literature.

<table>
<thead>
<tr>
<th>Group</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Expert providers    | *Repair services:* car repairs\(^1\), computer repairs\(^2\), plumbing, roofing work, appliance services, home improvement contractors, television replacements  
                      *Health services:* dentistry\(^3\), veterinary services\(^4\), the removal of an appendix\(^5\)  
                      *Other services:* management consulting; legal advice auditing\(^7\); funeral services\(^8\), taxi rides in unknown cities\(^11\), budget allocation by politicians\(^9\), infrastructure projects\(^9\) |
| Expert advisors     | Financial advice markets\(^6\), financial rating agencies\(^6\), pharmaceutical advice by physicians and pharmacists\(^6\), scientific consulting, real estate agency, computer equipment consulting |
| Information experts | *Process-attributes of food products:* (organically produced food, dolphin-safe tuna, free-range poultry, genetically modified organism (GMO)-free food, irradiated food, organic products, fair trade products, types of goods, use of pesticides)\(^12,13\), possibility of recycling\(^14\), restaurant hygiene\(^18\)  
                      *Process-attributes of non-food products:* journalism\(^15\), clothes produced by well-compensated workers\(^16\), products claiming better environmental performance (e.g., low-emissions electricity) |
| No expertise        | *Products with unknown benefits:* Long-term preventive care\(^17\); new medication for rare diseases |


taste of the wine is an experience characteristic, because it can only be evaluated after consumption. The information on the label stating that the wine has been processed organically is, however, not easily verifiable and therefore constitutes a credence characteristic. Therefore a statement like “Experience goods (like wine)” (Dulleck et al. (2011), p.530) is possibly misleading when a reader associates search or credence characteristics with “wine”.

Another example is a dental tooth filling (see chapter 2 of this thesis). The location of the practice and the politeness of the staff are search characteristics. The fit and comfort of the filling is an experience characteristic. Finally, whether the filling
1.1. Introduction

was necessary at all or not typically remains unknown and therefore is a credence characteristic. A researcher who cares about whether patients are treated appropriately, or possibly overtreated, focusses on the credence characteristic of dental tooth fillings. Nevertheless, people may care about the non-credence characteristics of the service (“the doctor was so polite”) and it hence would be misleading to make the general statement that dental care is a pure credence good.

**Characteristics may differ with consumer information.** Take the ordinary consumer Peter who brings his car to a repair shop, because it makes an unfamiliar noise. When Peter picks up his car a day later, the mechanic tells him that some parts (which Peter has never heard of) had to be replaced. The noise is gone, but Peter cannot be sure whether it was necessary to replace all the parts, and the repair therefore has credence characteristics. Now imagine the retired car mechanic Katie brings her car into the same shop, because he has heard the same noise as Peter. But Katie knows that the noise indicates that a specific part needs to be replaced. Not credence but experience characteristics shape this case, as Katie will be able to evaluate whether the repair was according to his needs nor not. Peter may be the standard consumer in the car repair market, but depending on the market there may be more or less Katies around.

**Characteristics may change over time.** Credence characteristics may become experience characteristics when these characteristics become evaluable after some time (DK73, Zweifel and Eichenberger (1992)). An example for this is given by an auditing case described in Causholli et al. (2013): a firm is uncertain about his/her auditing needs. The auditor subtracts substantial rents by underauditing the client. As long as there are no contrary indications, stock owners cannot tell whether the firm was appropriately audited or not. Underauditing can remain undetected for years – until it is discovered, for instance by unexpected bankruptcy. With the bankruptcy of the firm, the former credence characteristics of auditing have become experience characteristics. Another possible channel through which credence characteristics can become experience characteristics exists when a stochastic process is the reason why consumers cannot evaluate a service. For instance in financial advice consumers cannot tell whether a bad outcome is due to biased advice or bad luck. Sampling from repeated purchases may enable consumers to evaluate advice after some time. The possibilities for such sampling on the individual level are limited in many real-world situations, however, as many products, e.g., retirement saving plans, are purchased infrequently (Zweifel and Eichenberger, 1992).

The logic of these credence-to-experience transformation carries on to experience-to-search transformations, for instance, when the taste of a particu-
lar wine has become familiar to consumers after repeated purchases of the same wine.

1.1.3 Credence goods problems

What are credence goods problems? Problems associated with the credence characteristics of a good are called *credence goods problems*. Likewise, problems based on the experience characteristics of a good are *experience goods problems*.

When a seller of wines makes a false claim about the wine’s taste, the consumer will find out shortly after the purchase and we hence have an experience goods problem. When the seller makes a false claim about the processing of the wine (fair-trade), this is a credence goods problem as the consumer cannot verify the information provided. Similarly, a bad quality dental filling is an experience goods problem as the patient will detect it breaks when chewing on hard food. Placing a filling when no filling is needed, however, is a credence goods problem as the patient possibly will never know whether the filling was necessary or not.

The pre-condition for different credence goods problems which have been analysed in the literature is the information asymmetry between seller and consumer. When the interests of sellers and consumers are not aligned, sellers may misrepresent the consumers’ needs (in markets with expert providers and expert advisors in table 1.1) or the true attributes of purchased products (in markets with information experts in table 1.1), because credence characteristics prevent the consumer from noticing such fraud.

Credence goods problems and their modelling. The main categories of problems described in the literature are presented in table 1.3. My categorization is based on Dulleck and Kerschbamer (2006), but adds some elements.

These problems are not necessarily limited to goods with credence characteristics. But for them to be problems stemming from credence characteristics, i.e., “credence goods problems”, the consumer must not be able to evaluate product characteristics, for instance the seller’s representation of the consumer’s need or the underlying attributes of a product.

---

8Sometimes it is argued that the causality is that information asymmetries lead to credence characteristics. I argue that the causality is rather reversed. As already noted by DK73, credence characteristics are the reason why consumers have a willingness to pay for expertise in the first place. In principle, consumers could decide to become experts as well (e.g., by studying), but the costs associated with this are often too high.

9The term “fraud” was already used by DK73 and is used widely in the literature.
1.1. Introduction

<table>
<thead>
<tr>
<th>Problem</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overtreatment</td>
<td>Consumer receives more or more expensive products or services than needed, but enough to solve a potential problem.</td>
<td>Repair services, health services, management consulting, legal advice</td>
</tr>
<tr>
<td>Overcharging</td>
<td>Consumer is charged more than he/she would be willing to pay could he/she verify which service or product he/she received.</td>
<td>Repair services, health services, goods with hidden characteristics</td>
</tr>
<tr>
<td>Undertreatment/Biased advice</td>
<td>Consumer receives products or services which are insufficient given his/her need.</td>
<td>Financial advice, auditing, rating agencies, health advice</td>
</tr>
</tbody>
</table>

**Overtreatment:** When consumers are overtreated, i.e., purchase more or more expensive goods than needed, they notice that a potential problem (for instance, being sick) is absent, but cannot evaluate whether the service or product they received was necessary or whether a cheaper service or product would have sufficed. Experts have incentives to overtreat, when overtreating is associated with greater marginal profits than providing an appropriate treatment.\(^{10}\) Overtreatment is typically modelled such that the consumer receives the same utility (ignoring prices) from overtreatment as from the more efficient appropriate treatment.

**Overcharging:** Consumers may sometimes be charged more than they would be willing to pay if they knew which service or product they have actually purchased. A pre-requisite for overcharging is that consumers are not able to verify what product or service they have received. As an example from car repair, consumers may not be able to verify whether the parts the mechanic claims to have changed really have been changed. Also false product labelling of products (goods in the second category in table 1.1) is a variant of overcharging: Consumers have a higher willingness to pay for products with a fair-trade label and may be charged more than they would be willing to pay if they knew that the label was a fake. As in the case with overtreatment, the inability to evaluate overcharging is typically modelled such that consumers receive the same utility (ignoring prices) from consumption, with or without overcharging.

**Undertreatment/biased advice:** Being undertreated means that the consumer obtains a service or product which is insufficient to satisfy his/her need. Incentives for undertreatment may occur, for instance, when the capacities of an expert are constrained (Emons, 1997) or when third-party commission payments lead to incen-

\(^{10}\) These incentives can have exogenous or endogenous sources.
Chapter 1

tives for biased advice (Inderst and Ottaviani, 2012a). Undertreatment is a problem of credence characteristics in accordance with the characterisation by DK73 only if the consumers’ is not able to notice seller fraud. Hence a different modelling approach than with overtreatment is needed. In the literature, this is for instance implemented by assuming that the success of the outcome is a stochastic function of the seller’s recommendation or service (Emons, 1997; Becker and Milbourn, 2011; Inderst and Ottaviani, 2012a). In these models, the consumer may realize that he was undertreated, but cannot evaluate the accurateness of advice, because the bad outcome can be due to both (bad) luck or seller fraud. If consumers could sample a large number of recommendations they could infer that seller’s defrauded on average, but arguably many services with credence characteristics are purchased too infrequently to allow for such statistical identifiability at least on the personal level (Zweifel and Eichenberger, 1992). Hence, the stochastic element prohibits that the consumer could prove expert fraud in court and hold experts liable. A related way to ensure that undertreatment can be called a credence goods problem could be provided by assuming that the expert’s (unobservable) diagnosis effort determines the probability that the expert can make the correct decision. Thereby it is necessary that even with the greatest effort, the expert does not obtain a perfect signal about the consumer’s need. In the credence goods literature, however, it is usually assumed that the highest diagnosis effort leads to a perfect signal (Dulleck and Kerschbamer, 2009; Bester and Dahm, 2017), allowing consumers to infer that undertreatment is due to low diagnostic effort of the expert.

1.2 Different characterizations of credence goods

In this section, I survey and categorize the characterizations of credence goods which authors have provided over the years. I will describe the problems associated with different characterizations.

1.2.1 A classification of characterisations in the literature

In table 1.4, I present a (not mutually exclusive) categorization of how authors have characterized credence goods since Darby and Karni (1973).\footnote{I also include some papers who do not directly analyse credence goods, but refer to them in a direct way like Zweifel and Eichenberger (1992); Bolton et al. (2007); Huck et al. (2016b) and some wo do not even refer to the term like Jin and Leslie (2003) and Becker and Milbourn (2011).} It shows that many authors have followed the characterisation by DK73 (category (1)). Since
1.2. Different characterizations of credence goods

the 2000s, some authors have characterized credence goods less strictly than DK73 (category (2)). Further, some authors have begun to characterize credence goods with a focus on markets with expert providers, i.e., the traditional repair services (category (3)).

The first category (1) assembles papers that follow the characterisation by DK73 as they require for a good to be a credence good, that consumers are not able to evaluate product characteristics even after purchase. An example is:

“Since from ex post observations the buyer can never be certain of the quality of the services he has purchased, such services have been termed credence goods (Darby and Kami, 1973).”

(Emons, 1997) (p.107)

While Emons writes “can never be certain”, other authors in this line use similar formulations like “never able” (Bolton et al., 2007) or “is ex-post nonverifiable” (Dulleck et al., 2015). Other characterisations are less general, but nevertheless much in the spirit of DK73:

“We assume that the consumer can detect only whether the problem still exists. If the problem no longer exists, the type of repair actually performed is unknown. Thus, the repair is a credence good (see Michael Darby and Edi Karni, 1973).”

(Pitchik and Schotter, 1987) (p.1033)

Although Pitchik and Schotter focus on a repair problem, they attribute the credence characteristic in the repair to the fact that the consumer cannot evaluate whether the seller correctly represented the consumers’ condition.

The second category (2) assembles papers with less strict characterizations of credence goods. These characterizations depart from DK73 by not strictly requiring that consumers be unable to evaluate characteristics of a good. The authors typically only demand that this holds “often” (Fong, 2005), “typically” (Balafoutas et al., 2013) or “in many cases” (Beck et al., 2014).12

In the third category (3) in table 1.4, I collect works which share the feature that, for a good the a credence good, they require an information asymmetry between seller and consumers with respect to the consumers’ needs. In the characterisation of DK73, information asymmetries between seller and consumer may occur, but it holds with or without it. Characterizations in (3) therefore refer to a subset of

---

12Formulations which refer to the costs of a possible evaluation – like “not easily” and “difficult” (Mattila and Wirtz, 2002) – are categorized into (1), because they express that any consumer could theoretically decide to become an expert herself. However, the costs associated with acquiring expert knowledge are usually high.
Chapter 1

credence goods, mostly repair services for which this kind of information asymmetry prevails. Two examples from category (3):

“The key feature of credence goods is that consumers do not know which quality of a good or service they need.”

(Dulleck and Kerschbamer, 2006) (p.7)

“In the case of taxi rides in an unknown city, the service traded on the market is a credence good (Darby and Karni, 1973), meaning that an expert seller possesses superior information about the needs of the consumer.”

(Balafoutas et al., 2017) (p.2)

<table>
<thead>
<tr>
<th>(1) Characterisations stressing consumers’ inability of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitchik and Schotter (1987); Zweifel and Eichenberger (1992); Ekelund et al. (1995); Taylor (1995); Emons (1997); Richardson (1999); Emons (2001); Crow et al. (2002); Mattila and Wirtz (2002); Alger and Salanie (2006); Baksi and Bose (2007); Bolton et al. (2007); Iizuka (2007); Baron (2011); Dulleck et al. (2011); Liu (2011); Inderst and Ottaviani (2012a); Causholli et al. (2013); Emons (2013); Dulleck et al. (2015); Dulleck and Wigger (2015); Baksi et al. (2016); Huck et al. (2016a,b); Kerschbamer et al. (2016); Schneider et al. (2016); Bester and Dahm (2017)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(2) Characterisations not requiring the inability of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fong (2005); Suelze and Wambach (2005); Dulleck and Kerschbamer (2006); Balafoutas et al. (2013); Beck et al. (2013); Fong et al. (2014); Mimra et al. (2016a,b); Hilger (2016); Bester and Dahm (2017)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(3) Characterisations with focus on expert providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fong (2005); Alger and Salanie (2006); Dulleck and Kerschbamer (2009); Dulleck et al. (2012); Beck et al. (2014); Balafoutas et al. (2017)</td>
</tr>
</tbody>
</table>

No category: Wolinsky (1993); Pesendorfer and Wolinsky (2003); Hyndman and Ozerturk (2011); Brown and Minor (2012); Balafoutas et al. (2013); Bonroy et al. (2013); Das et al. (2016); Kerschbamer et al. (2017); Rasch and Wasbel (2017, forthcoming)

13 The quote by Balafoutas et al. also supports my notion that the conflicts between new characterisations and DK73 went largely unnoticed in the literature, as Balafoutas et al. refer to DK73 in the same sentence in which they require asymmetric information, although this is not required by DK73.
1.2.2 Consequences of different credence goods characterisations

Due to its great influence, I use Dulleck and Kerschbamer (2006) (DK06) to illustrate how some credence goods characterizations can lead to conflicts with the original characterisation by DK73. DK06 provide a widely-cited survey of the credence goods literature and a unifying model for markets with expert providers which has been adopted by other authors since. Bester and Dahm (2017) call the setup by DK06 “the now standard credence goods problem”. There is a considerable overlap of the characterisation by DK73 and the characterization by DK06. Nevertheless, the characterization by DK06, and those by authors influenced by their work, excludes several of the examples for credence goods presented in table 1.1; and it includes some examples not included in table 1.2, because these examples would be termed experience goods rather than credence goods, when judging by DK73.

1.2.2.1 Experience goods problems vs. credence goods problems

In the setup of DK06, a seller of credence goods has two treatment options: a “cheap” and an “expensive” one, where the latter is associated with higher costs. The consumer is in one of two states and either requires the cheap or the expensive treatment in order to avoid a loss. As long as the loss is avoided his utility from the treatment is \( v > 0 \) and zero if the loss occurs. In this setup the consumer is overtreated if he/she requires the cheap, but receives the expensive treatment and is overcharged if he/she is charged the expensive treatment, but (secretly) only receives the cheap treatment. Both cases are characterized by the inability of the consumer to evaluate his/her state, i.e., whether he/she needed the cheap or the expensive treatment – his/her utility is always \( v \). Hence, these situations are in line with the credence goods characterization of DK73.

Undertreatment is the third possible problem in DK06. A consumer is undertreated when she requires the expensive treatment, but only receives the cheap treatment. Yet, when undertreated, the consumer directly learns that she was undertreated by observing his utility of 0. Because the expert can diagnose the consumer’s need without error, the consumer can directly infer that the expert must have misrepresented his needs. The information provided by the expert therefore has experience characteristics rather than credence characteristics. The consumer could prove his case in court or contract on the fulfilment of his need in the first place. Undertreatment in DK06 is therefore only possible by assuming the absence of liability. It is, however, not clear why the absence of liability should not lead
to undertreatment in markets for experience goods as well.\footnote{The lemons-problem in Akerlof (1970) may be interpreted in this way.} When an experience characteristic is at the centre of our analytical interest, the analysed problem would better be called a an experience, and not a credence goods problem. Other authors (Hilger, 2016; Mimra et al., 2016a; Bester and Dahm, 2017) have recently followed this approach and study goods with both credence and experience characteristics under the term credence goods.

1.2.2.2 Credence goods and asymmetric information

Authors who require the informational asymmetry between seller and consumer with respect to the consumer’s need (category (3) in table 1.4) tend to exclude many examples of credence goods listed in table 1.1 from their characterization. Such a characterization excludes markets with information experts: goods for which consumers know what they need, but not what they get. This exclusion contrasts the fact that these goods are credence goods in the sense of DK73. Furthermore, they are concerned with a very similar problem as the model by DK06, namely the problem of overcharging which occurs in DK06 when consumers cannot verify what service or treatment they have received.

The setup of DK06 does also not consider markets with expert advisers in which the seller recommends products provided by third parties to consumers (Inderst and Ottaviani, 2012a). A consumer may realize that he/she has purchased a non-optimal good, but he/she is unable to judge the quality of advice received, because diagnosis and outcome involve a stochastic element.\footnote{Inderst and Ottaviani (2009) comment that the credence goods literature puts much focus on what I call expert providers. Already DK73 stated that “much of our discussion focuses on the key problem of the joint provision of diagnosis and services – such as the choice and execution of an automobile repair or taxicab route.” (p.67), where they used the word “much” instead of “all”.} Last, markets without expertise (the last category in table 1.1), where the true benefits are unknown, are not considered by credence goods characterizations which require information asymmetries. For these goods, expertise does not exist per definition.

1.2.2.3 The distinction between experience goods and credence goods

It is simple to distinguish between search, experience and credence characteristics, this distinction cannot simply be carried on to distinguish search, experience and credence goods. Recently, the focus on credence goods instead of credence characteristics has led to greater divergence from the categorization of DK73. A
possible reason is that when authors do not require that good characteristics are non-evaluable after purchase, they naturally break with the goods categorization of DK73.\footnote{The literature has not yet noted this as only a few authors who conflict with DK73’s characterisation have commented on the difference between experience goods and credence goods in the first place.} One of the few comments on the difference between credence and experience goods reflecting this problem is provided by Dulleck et al. (2011):

“Experience goods differ from credence goods in several important dimensions. For example, (i) while the valuation of a consumer is strictly increasing in quality with experience goods, it is constant whenever the quality is sufficient with credence goods; (ii) for given prices a consumer can tell exactly which quality he prefers in the case of experience goods, but he does not know it with credence goods; (iii) whereas the quality of the good is unobservable ex ante but perfectly observable ex post with experience goods, it may be observable either ex ante, or ex post, or neither ex ante nor ex post with credence goods.”

(Dulleck et al., 2011) (p.533, footnote)

Arguably, this characterization is not straight-forward. With respect to (i) I may counter that it is not generally the case that valuation increases with quantity with experience goods. Take the example of a beer drinker with who despises wine, i.e., whose utility is independent of the sort (quality) of wine she consumes, although wine has experience characteristics (the authors themselves call wine an experience good (p.530)). (ii) seems to state a natural consequence of the credence goods problems of overtreatment and overcharging, but it does not hold when undertreatment occurs in the model of Dulleck et al. (2011), because then the consumer knows he would have needed the expensive treatment/quality. Finally, (iii) may be interpreted as a consequence of the authors reference to goods instead of characteristics.

1.3 Conclusion

Goods as diverse as health care, financial advice, food labels and journalism have frequently been termed credence goods, because important characteristics of these products cannot be evaluated by consumers. I have categorized the different lines of research which refer to credence goods into markets with expert providers, expert advisors, information experts and no expertise. The focus of the literature has been on expert providers for a long time, but the two other categories have developed more recently, particularly the literature on markets for advice in the
wake of the financial crisis. I have shown that, mostly in the literature focussing on expert providers, the characterisation of credence goods has diverged from the classic characterisation by DK73. I argue that this should be avoided in order to make sure that first, authors only refer to credence goods when they analyse credence characteristics of a good; and second, that authors do not formulate general characterisations of credence goods when they actually only give a characterisation for a subgroup of credence goods, e.g., those sold in markets with expert providers. I propose to use the classic characterisation by DK73 as it is clear and can successfully combine closely related subfields under one roof.

Clarity with respect to the characterisation of credence goods is important, because it raises awareness for similar problems in differently appearing markets. This may serve to improve research within the field, the communication of research findings to researchers from other fields, to the public and to politics.
2 Health Services as Credence Goods - A Field Experiment.

Information problems are a defining characteristic of health care markets. Despite their importance, there is little direct evidence of the impact of information problems between patients and physicians on the quality of treatment. Furthermore, the role of market conditions is not well understood. In this paper, we present the results from a field experiment in the market for dental care. A test patient who does not need treatment is sent to 180 dentists to receive treatment recommendations. In the experiment, we vary two factors: First, the information that the patient signals to the dentist. Second, we vary the perceived socioeconomic status (SES) of the test patient. Furthermore, we collected data to construct several measures of short- and long-term demand and competition as well as dentist and practice characteristics. We find that the patient receives an overtreatment recommendation in more than every fourth visit. A low short-term demand, indicating excess capacities, leads to significantly more overtreatment recommendations. Physician density and their price level, however, do not have a significant effect on overtreatment. Furthermore, we observe significantly less overtreatment recommendations for the patient with higher SES compared to lower SES under standard information. More signalled information however does not significantly reduce overtreatment.

2.1 Introduction

Information problems are ubiquitous in health care markets. In the patient-physician relationship, physicians often have an informational advantage vis-à-vis their patients: Whereas patients neither know exactly which disease they suffer from nor which health care service best treats their health problem, physicians are able to diagnose patients and judge which treatment is appropriate.\(^1\) Patients thus

\(^1\)Physicians also typically have better information on other dimensions than appropriateness of treatment, such as costs, compensation and their own skills.
have to rely on the physician to recommend and provide the appropriate treatment. Often, even after receiving treatment, patients cannot judge whether the treatment was appropriate or whether, for instance, a less invasive or less expensive treatment might as well have solved their health problem. Therefore, health care services have credence qualities and are called credence goods. Put more generally, it is difficult to assess quality objectively in health care markets.

This questions whether standard market incentive systems such as competition lead to efficient outcomes. For instance, depending on financial incentives, a physician might exploit his informational advantage by overtreating. This overtreatment cannot be detected, and it is a priori not clear whether more competition in the form of a higher physician density helps to alleviate this problem. On the contrary, in a version of the physician-induced demand hypothesis the lower demand for an individual physician with higher physician density would be compensated by inducing demand, for instance by overtreating. Now in theory, price competition could lead to equal mark-up prices across treatments such that there are neither under- nor overtreatment incentives, however equal mark-ups across all treatments are an appealing theoretical but not empirically implementable concept.

Even though there is some empirical evidence indicating that physicians react to financial incentives in treatment decisions, this evidence is typically indirect and based on highly aggregated data such as administrative data from hospitals. This has the major drawback that the physician-patient interaction, including patient demand effects, communication and the actual asymmetry of information, cannot be controlled for. Then, although it might be concluded that physicians react to financial incentives, it cannot be concluded that the provision of health care services was actually inappropriate or inefficient due to supply side reasons.

In this paper, we address this fundamental problem and provide direct evidence of physicians’ treatment decisions and their determinants with a focus on the role of the market environment as well as patient characteristics and information. To do so, we conducted a natural field experiment in the Swiss dental market based on individual physician visits. In our study, a single test patient visits 180 randomly

---

2The categorization of good qualities into search, experience and credence qualities was proposed by Darby and Karni (1973). Goods for which credence qualities are important are usually called credence goods, although the same good can generally have qualities of all three categories. See Dulleck and Kerschbamer (2006) for an overview on the theoretical credence goods literature and Kerschbamer and Sutter (2017) for an overview on credence goods lab and field experiments.

3For an early discussion, see Evans (1974).

4See Clemens and Gottlieb (2014) for an overview.

selected dentists for a checkup. At each dentist visit, the test patient asked for a diagnosis—based on an examination and the same x-ray photograph—, a treatment recommendation and a cost estimate. The test patient had a superficial caries lesion which should not be treated with an invasive treatment—such as a filling—according to the Swiss Dental Guidelines and four cooperating reference dentists. Thus, we focus our analysis with the case at hand at the credence goods problem of overtreatment which wastes resources and may spur adverse health effects in the long run.\textsuperscript{6}

To analyze whether and how market characteristics affect overtreatment, we collected data from several sources to construct different measures of short- and long-term demand and market competition. This is complemented by data on dentist and practice characteristics. We consider five measures of the market environment. First, we collected two measures of the demand at an individual dentist: (i) short-term demand, measured by the patients’ waiting time for the next possible appointment, and (ii) medium-term demand measured by whether the physician has an informative web page. The logic behind this latter measure is that physicians who do not have a full patient book are more likely to try to attract patients with a well-done and informative web page. Second, we have two measures of the competitive behavior of an individual physician: (iii) the physician’s price level\textsuperscript{7}, and (iv) whether the physician clearly displays the price level in the practice. Displaying the price level in the practice is required by regulation, however 60\% of practices in our sample failed to do so\textsuperscript{8}. Last, we measured (v) long-term market competition by physician density.

Furthermore, we apply a 2x2 design for our experimental variations: First, we vary the information that the patient signals to the dentist. In particular, the patient either goes as a standard patient (who simply asks for a diagnosis) or the patient informs the dentist that he has, out of curiosity, uploaded his x-ray the day before to an internet dentist platform which provides information and diagnoses,

\textsuperscript{6}In our case, an invasive treatment such as a filling can lead to a higher subsequent caries risk. Another fundamental problem in credence goods markets is undertreatment, which can occur for several reasons, e.g., when liability does not hold. Although this is a serious problem, we cannot address undertreatment in a similar field experiment due to ethical considerations as it would imply denying a necessary treatment during the time of the study.

\textsuperscript{7}Dentists in Switzerland can decide, to some extent, on their overall price level. A detailed description will be provided in Section 2.3.

\textsuperscript{8}At a first glance, the rate of physicians not displaying the price level in their practice may seem high but is in line with other studies. For instance the Swiss magazine K-Tipp reported in 2004 that 27 out of 35 sampled dentists in large Swiss cities did not display the price level (K-Tipp No. 13, 8/2004). The consumer organization FRC sampled 170 dentists in the French speaking part of Switzerland in 2005 and found an adherence rate of 59\% (reported in Schweizer Monatsschrift für Zahnmedizin Vol 115, 10/2005).
and that he is curious about what both recommendations will be. Thus, the patient signals that he will likely receive another diagnosis from an internet platform.\textsuperscript{9} For simplicity of terminology, we will refer to the patient as either the \textit{standard} patient or the \textit{informed} test patient and the experimental conditions\textsuperscript{10} as ST vs. INFO.

The second variation is whether the patient is perceived as a patient with a lower or a relatively higher socio-economic status (LS vs. HS). This variation is implemented by modifying the physical appearance of the patient in terms of clothing and accessories. Table 2.1 summarizes the experimental design.

\begin{table}[h]
\centering
\caption{Experimental conditions.}
\begin{tabular}{lll}
\hline
\multicolumn{1}{l}{Information} & Standard & Informed \\
\hline
SES & Low & ST-LS & INFO-LS \\
     & High & ST-HS & INFO-HS \\
\hline
\end{tabular}
\end{table}

Our study has two main contributions: First, we investigate physicians’ provision of health care services on the level of individual patient-physician interactions. The design allows us to observe for each physician whether she/he provides the appropriate treatment recommendation or an overtreatment recommendation instead of observing only \textit{aggregate} provision rates. Thus, we can provide direct evidence of overtreatment. Our micro approach allows us to not only observe the overtreatment behavior but also to control for the covariates on the individual level. With the experimental variation, we furthermore provide results on the role of patient SES and signalled information on overtreatment. Second, we provide results on the role of the market environment and physicians’ characteristics on the level of overtreatment. Understanding how demand and competitive conditions affect physicians’ treatment decision under asymmetric information is an important prerequisite for market design and regulation in these markets and thus far empirical evidence is scarce.

Our central result is an overtreatment recommendation rate of 28\% (50/180). Conditional on an overtreatment recommendation, mean overtreatment costs taken from the collected cost estimates amount to CHF 535 (about $550), the median costs.

\textsuperscript{9}The patient clearly communicates that he has not yet received information from the platform. We choose this design in order not to anchor the dentist’s treatment recommendation on an already received diagnosis, but to signal that the patient is likely to receive another diagnosis.

\textsuperscript{10}Note that we refer to experimental \textit{conditions} instead of \textit{treatments} to distinguish between dentists’ treatments and experimental conditions.
being lower at CHF 444 (about $455). Regarding the treatment—the test patient has a superficial interproximal caries lesion that should not be treated by an invasive treatment such as a filling—, the suggested number of fillings at a dentist ranges from 1 to 6. Furthermore, we observe that 13 different teeth to be treated with a filling appear across all cost estimates.\footnote{This constitutes a lower bound, as in several cost estimates it is not indicated which particular tooth—each tooth has a number—will be treated.} Thus, besides our finding of a considerable overtreatment recommendation rate, we also observe a striking heterogeneity in the treatment recommendations.

Looking into the role of market characteristics, our second main result is that lower short-term demand leads to a significantly higher level of overtreatment recommendations. Every additional day of waiting time reduces the likelihood of receiving an overtreatment recommendation by more than a percentage point. This result suggests that physicians with free capacity are more likely to provide treatments that are not necessary. Physician density does not have a significant influence on overtreatment recommendations, even if we do not control for the other competition measures such as short-term demand. We consider this result to be important in light of the results of a vast body of empirical work that approaches the physician-induced demand hypothesis by relating treatment volumes per capita to physician density based on aggregate data.

Furthermore, the physician’s price level per se does not have a significant influence on overtreatment, however, we find significantly less overtreatment recommendations at physician visits for which the price level was clearly displayed in the practice. The first result regarding the price level is surprising given that the price level determines the magnitude of the financial incentive to overtreat. However, it has to be noted that the difference in final price across different price levels for the treatment considered is rather small, which can explain the result. One interpretation of the second result—that there are significantly less overtreatment recommendations in physician practices that implement the regulation and display the price level in the practice—is that there might be different types of physicians: Those that abide by regulation and (treatment) guidelines, and those that are less prone to do so. A caveat when interpreting our demand and competition measures is that we do not have a separate measure of the perceived quality of the dentists. Although rating websites start to be in place for dentists, we do not have enough data to construct a meaningful measure of perceived quality and thus an important demand component cannot be directly controlled for.
Regarding our experimental treatment variations, we observe the counterintuitive result of significantly less overtreatment recommendations when the test patient is a high- rather than a low socio-economic status (SES) patient, given that the patient is a standard patient. When the patient is informed, the differences diminish. These results suggest a complex role of patients’ SES as well as interactions with signalled information. With respect to the information variation itself, we do not find a significant change in the overtreatment recommendation rate. We observe a reduction in the overtreatment rate for a patient with lower SES when going from standard to more information, however, this difference is not significant. It is generally assumed that information provision and diagnostics from the internet increase patient information and quality of care. Our results show the limits of this argument when the case at hand has credence goods characteristics and is complex, as is the case for most health care services.

The remainder of the paper is as follows: In the next section, we discuss the related literature. Section 2.3 provides a description of the market for dental care in Switzerland. In Section 2.4, we present the field experiment, and describe our data in detail in Section 2.5. Results are presented and discussed in Section 2.6. Section 2.7 concludes.

2.2 Related literature

There are several strands of literature relating to our field experiment: we first present a brief overview of the empirical literature on physician-induced demand (PID) and physician behavior as well as recent audit studies in health care. We then turn to field experiments and empirical results on other credence goods markets.

Physician-induced demand and financial incentives

Early empirical studies find that an increase in the number of surgeons is associated with an increase in the number of surgeries (Fuchs, 1978; Cromwell and Mitchell, 1986). Grytten et al. (1990) show a similar effect for dentists. Grytten and Soerensen (2001), however, find that a higher physician density is not correlated with more treatments, regardless of the remuneration method. In a meta-study, Leonard et al. (2009) provide an overview and comparison of a vast literature on PID in which many studies approach the topic by analyzing the correlation between physician density and a measure of health care utilization such as annual number of procedures per general practitioner. The idea behind is that a higher density implies
lower demand per physician, which is compensated by physician-induced demand. This literature generally finds a significant association between physician density and health care consumption.

Gruber and Owings (1996) report that the decline in fertility rates in the US in the 1970s was partly compensated by a substitution from normal childbirth to the more profitable cesarean delivery. Using data from the Japanese prescription drug market, Iizuka (2007) finds that the mark-up on drugs influences prescription decisions, yet physicians are also found to be willing to forgo profits in order to reduce costs for patients. Clemens and Gottlieb (2014) analyze area specific price shocks following a Medicare consolidation reform and find that areas with higher payment shocks experience significant increases in health care supply. The empirical literature thus indicates that physicians react to financial incentives in their medical decisions. However, the importance and extent of physician-induced demand as well as its determinants are still an open question.

Audit studies in health care

A small number of recent audit studies uses data from direct observation of a physician-patient interaction as in our study. In Currie et al. (2011) and Currie et al. (2014), the authors sent students, trained as test patients, with identical verbally communicated flu-like complaints to physicians in Chinese hospitals. The institution setting is such that physicians prescribing medication receive kickbacks on medication bought at the hospital pharmacy. The authors analyze whether patients signaling that they are informed about inappropriate antibiotics use are prescribed less antibiotics than other patients. Currie et al. (2011, 2014) find that the signal reduces the probability of receiving an antibiotic prescription by 25 percentage points, from 64% to 39%, and that the signal also reduces drug expenditures. Lu (2014) also investigates physicians’ prescribing behavior in Chinese hospitals. The variations are whether patients are insured and whether they indicate to purchase the prescribed drug at the hospital. Lu (2014) finds that physicians write prescriptions that are significantly more costly for insured than uninsured patients, but only if physicians receive kickbacks.

In a large audit study in India, Das et al. (2016) compare physician effort and treatment between private and public health care providers. Test patients are sent to physicians and communicate symptoms of one out of three predefined diseases. Das et al. (2016) find that physicians’ diagnosis and treatment quality do not vary between public and private providers although private providers lacked medical
qualifications. Private providers balance their worse qualification by a significantly higher effort level.\textsuperscript{12} The level of unnecessary treatments is high under both public and private health care provision. In fact, 70\% of the providers provide an unnecessary treatment. Das et al. (2016) also compare the behavior of physicians with both public and private practices and find that all quality metrics are higher in their private clinics.

In contrast to the above studies, in our experimental design the diagnosis is not based on communicated symptoms (of different patients), but on examination of the patient and an x-ray. Our case is thus more strongly patient-specific reflecting true credence goods characteristics. This difference is important since patient-independent information such as inappropriate antibiotics use for a flu cannot easily improve outcomes in our case, and credence goods markets in general. In terms of focus of the analysis, we concentrate on the impact of the market environment by collecting several measures of demand and competition on physician overtreatment decisions, which has not been analyzed in the previous health care audit studies.

**Credence goods field experiments**

There are four field experiments in a credence goods markets framing that relate to our paper:\textsuperscript{13} Schneider (2012); Balafoutas et al. (2013, 2017); and Kerschbamer et al. (2016). Schneider (2012) finds overtreatment in an undercover experiment in the market for car repairs. Balafoutas et al. (2013) and Balafoutas et al. (2017) perform field experiments in the Greek market for taxi rides. Their results in the market for taxi rides contrasts ours from health care: While they find that customers with less information are overtreated (taken on detours) more often than well informed customers, they find no significant differences in overtreatment across customers’ income levels. Balafoutas et al. (2017) report evidence for second-degree moral hazard: Customers who indicate that their expenses are reimbursed by their

\textsuperscript{12}In a preceding study, Das and Hammer (2007) investigate differences in quality of health care provision in India using vignettes as well as accompanying physicians during one day in their practice. They find as well that physicians in public hospitals exert significantly less effort than physicians in private hospitals however physicians in public hospitals exert more effort than physicians in small public clinics. In this study, a physician’s effort and his competence are positively correlated: the more competent a physician is, the more effort he exerts.

\textsuperscript{13}Due to the challenges of designing and performing field experiments in credence goods markets, the number of laboratory studies has grown in the past years. Dulleck et al. (2011)’s seminal paper investigates the impact of market institutions on expert behavior under endogeneous prices. Mimra et al. (2016a) shows that price competition may inhibit quality competition and thus lead to more inefficient market outcomes than regulated prices. Mimra et al. (2016b) show that second opinions may be an effective instrument to reduce the level of overtreatment. In contrast, Huck et al. (2016a) find that insurance increases the level of overtreatment.
employer are overtreated more often than those customers that do not. By varying insurance coverage, Kerschbamer et al. (2016) confirm the importance of second-degree moral hazard in the market for computer repairs. We add to this literature on credence goods with a field study in the health care market, the economically most significant credence goods market.\footnote{An interesting difference compared to other credence goods markets is that social preferences such as altruism and ethical norms are presumably playing a more important role in expert behavior.}

\section*{2.3 The market for dental care in the canton of Zurich}

The canton of Zurich is the most populous among the 26 cantons of the Swiss Confederation. In December 2015, the canton had 1.46 million inhabitants (Switzerland: 8.31 million). The two largest cities in the canton are Zurich and Winterthur with 396'000 and 108'000 inhabitants, respectively. In 2014, there were 4,217 registered dentists (57 per 100’000 population) in Switzerland among which 823 (51 per 100’000 population) were working in the canton of Zurich (BfS). Updating the population of dentists leads to (see Appendix 2.8.2 for the construction of the dentist population) 865 dentists in the canton of Zurich among which 402 (46%) were located in the city of Zurich, 70 (8.1%) in the city of Winterthur and 393 (45.4%) in the other municipalities of the canton.

The market for dental care in the canton of Zuerich is characterized by a dual system of providers: dentists and dental hygiene practices. Whereas dentists provide dental care in the classic sense, dental hygiene practices focus on preventive measures. As a part of these, dental hygiene practices may perform x-rays, for instance to check for interproximal caries. It is not uncommon that patients bring recent x-rays taken at other providers to dentists for diagnosis, which we will use in our experimental design. Dentists are mostly self-employed and are thus residual claimants of their provided services.

The Swiss association of dentists publishes binding treatment guidelines for common cases (Schweizerische Zahnärzte-Gesellschaft, 2015). Adult patients are essentially not insured for dental care, the Swiss Dental Association estimates that 85-90 % of the expenses are paid out-of-pocket by the patients.\footnote{Imfeld (2008) reports that even 94% of the dental care expenditures are paid out of pocket in Switzerland.} Only if patients

\pagebreak
Chapter 2

suffer from an accident or a severe underlying disease, health insurance covers dental expenses (Schweizerische Zahnaerzte-Gesellschaft, 2015).

The prices for dental care are regulated according to the Swiss Dental Tariff (Schweizerische Zahnaerzte-Gesellschaft, 2017) in the following way. The total price is a combination of two components: the number of points multiplied by the point value (PV).\textsuperscript{16} The number of points is regulated for each treatment, i.e. it is specified in the Swiss Dental Tariff how many points can be attached to a particular treatment. To be more precise, there is a small interval of points that can be assigned to each treatment. The upper bound of this interval is 33 to 35 percent above the lower bound. Based on the difficulty of treatment dentists may choose to apply slightly more or less points. As an example, the point range in the Swiss Dental Tariff for a standard consultation for diagnosis including anamnesis, checking for caries, inspection of the oral cavity etc. and information and discussion with the patient is 18-24, with 21 being the indicated standard rate. The point value is the multiplier of the points and is chosen individually by each practice. By law, each practice has to publish the point value that the practice applies (Staatssekretariat für Wirtschaft, 2004). The regulator sets an upper limit of CHF 5.8 for the choice of the point value. As an example, if the dentist’s practice has chosen a point value of CHF 4, a standard diagnosis consultation of 21 points is billed with $21 \times \text{CHF } 4 = \text{CHF } 84$ (about $\$86$).

2.4 The field experiment

2.4.1 Case & test patient

A field experiment in the health care markets is delimited by several strong requirements\textsuperscript{17} with respect to the test patient’s case: Physicians have to be able to diagnose the patient’s condition without side effects. Furthermore, the condition must not change during the time of the experiment to ensure that all diagnoses are made under the same prerequisites. The diagnosis itself must be based on identical information for each of the visits. Additionally, treatment guidelines have to give a clear recommendation of what is the appropriate treatment and what is considered to be overtreatment. The condition of our test patient, a minor superficial

\textsuperscript{16}The German expressions used in official documents are “Taxpunkte” (tax points) and “Taxpunktwert” (tax point value). For clarity we will from now on refer to “points” and “point value”, respectively.

\textsuperscript{17}See List et al. (2010); List (2011) for requirements for field experiments.
caries lesion between two teeth (interproximal), satisfies all of these criteria. One of our reference dentists took an x-ray picture at an initial visit that displayed the superficial caries lesion. Based on the x-ray and an inspection of the patient, all four reference dentists stated independently that the superficial caries lesion had not yet progressed to the dentin and should hence not be treated according to the treatment guidelines. Instead, the patient should be told to continue to well brush his teeth and that the lesion should be checked at a visit in a year’s time.\textsuperscript{18} It has to be noted that the test patient additionally suffered from two even more minor caries lesions and one shadow underneath an existing filling. All four reference dentists agreed that these even more minor caries lesions clearly do not require treatment and that the shadow is clearly recognizable as not being a caries due to its sharp borders and does not require any treatment.\textsuperscript{19}

In the experiment, the test patient presented the same x-ray at each visit and reported to have been shortly before at a dental hygiene practice that performed the x-ray and gave the recommendation to visit a dentist for a checkup. Due to the duality in providers (see section 2.3) and the fact that patients pay expenses out-of-pocket, it is not uncommon that patients present x-rays to dentists even at a first visit. Providing the x-ray ensured that all dentists would base their diagnosis on the same information.\textsuperscript{20} Our four reference dentists further agreed that diagnosing the interdental caries is almost effortless and hardly allows for diagnostic error. A second x-ray after the study shows that the caries lesion of our test patient did not advance during the study.\textsuperscript{21}

**Overtreatment** Based on the above, we define a recommendation that includes at least one filling to be an overtreatment recommendation. We furthermore use the number of suggested fillings as well as the billed amount as indicators for the extent of overtreatment.

Not only are the requirements high for the patient’s condition, several criteria also have to be met by the patient: Besides having the ”correct disease”, the minor superficial interdental caries lesion, the test patient had to be available for a long enough time span and be willing to conduct the large number of 180 dentist visits. Reliability was another key characteristic that our test person had to fulfill.

\textsuperscript{18}It is important to note that the check-up indicated is a year later, showing that nothing needs to be done during a considerable timeframe.
\textsuperscript{19}In terms of design, it would certainly have been preferable to have a test patient without these minor additional imperfections. However, from discussions with one of the reference dentists, it appears virtually impossible to find such a case.
\textsuperscript{20}Furthermore, additional x-rays, given their side effects, are precluded for ethical reasons.
\textsuperscript{21}Note that this second x-ray was, of course, also medically indicated as the yearly check-up.
To find candidates that would meet all the criteria, we sent out more than 6'000 emails to the University of Zürich psychology department’s subject pool and placed an advertisement online at the universities’ platform *Marktplatz*. The text in the email and advertisement was phrased very generally: A person which might have a teeth-related health problem such as a beginning caries was sought for a research project. 49 persons replied to our search. We sent out a detailed questionnaire to 44 candidates who were then interviewed on the phone. The most promising candidates were invited for an interview and a visit to one of our reference dentists. The next step was an assessment of the candidates’ cognitive skills and reliability. We found a male person in his twenties that fulfilled all criteria. All four reference dentists checked the test patient and independently agreed that the test patient was suffering from a minor interproximal superficial caries lesion and that no treatment was indicated.

The recruitment process was followed by an intense training of the test patient. In cooperation with the reference dentists, we worked out a detailed script that was tailored to the case and our test patient’s characteristics. We took pictures of the test persons’ outfits under the different roles (see Figure 2.13 in Appendix 2.8.1.2) in order to ensure an identical outfit on all visits. Once the role scripts were trained and the outfit was defined, the test patient was sent to five different dentists as real-life training sessions. The training was completed by another visit of one of the reference dentists where the test person’s appearance and script was once more evaluated.

### 2.4.2 Experimental design

We vary two test patient characteristics between the visits: the information that the test patient signals to the dentist and whether the test patient is perceived as a patient with a lower or higher socio-economic status. *Table 2.2* summarizes the conditions and provides the number of observations per condition in parentheses.

To indicate a *higher socio-economic status*, the test patient wore a high quality suit and high-end accessories such as an expensive watch, a car key and a new and expensive mobile phone. The test patient specified his occupation as a translator at a bank when asked to fill out the patient form. In the *lower socio-economic status* role, the test patient wore cheap unbranded clothes, an old backpack and had

---

22Marktplatz is an online trading platform provided by the University of Zürich and the Swiss Federal Institute of Technology and can be accessed here: http://www.marktplatz.uzh.ch/ (accessed on July 11th, 2017).
TABLE 2.2
Experimental conditions and number of observations.

<table>
<thead>
<tr>
<th>Information</th>
<th>Standard</th>
<th>Informed</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>ST-LS (45)</td>
<td>INFO-LS (45)</td>
</tr>
<tr>
<td>High</td>
<td>ST-HS (45)</td>
<td>INFO-HS (45)</td>
</tr>
</tbody>
</table>

no accessories. The test patient declared to be a student of translation doing an internship.\(^{23}\) Note importantly that in both socio-economic status-roles the patient signals to have studied. Hence, the level of education was kept constant across the two conditions.

We call the condition *informed* when the test patient indicates to the dentist that he has uploaded his x-ray—out of curiosity—to an internet platform where dentists offer free advice.\(^{24}\) The test patient further states that he had not yet received a reply to his x-ray upload in order to signal that he was expecting to get another diagnosis based on the x-ray but had yet not received one.\(^{25}\) In the *standard* patient role, there was no additional text to the script such that the patient went to the visits as for standard doctor visits.

Experimental hypotheses

We derive the following two hypotheses with respect to our experimental variations: The first hypothesis concerns the impact of the patient’s SES on the likelihood to receive an overtreatment recommendation. Dentists may perceive patients with a higher SES as wealthier and less price-sensitive than patients with a lower SES. Dentists might thus expect patients with a higher SES to be more likely to accept higher treatment bills. Hence, we hypothesize:

**Hypothesis 2.1** (Socio-economic status). The Patient with the higher SES receive an overtreatment recommendation more often than the patient with the lower SES.

Our second hypothesis refers to the information variation. Dentists may perceive patients that signal to get another diagnosis from an internet platform to be

\(^{23}\)See *Figure 2.13* in *Appendix 2.8.1.2* for a photograph of the test patient’s outfits and accessory.

\(^{24}\) An example for such a platform is [www.zahnforum.org](http://www.zahnforum.org) (accessed on July 13th, 2017). If dentists asked the patient on which platform he uploaded his x-ray to, the test patient referred to this platform.

\(^{25}\) Note that dentists consulted via internet only compete with practices in terms of diagnosis but not treatment. The platform is intended as a platform for information and diagnosis, but not to channel patients to dentists for treatment.
more informed than patients that do not and might thus expect the informed patient to be less likely to return for treatment than the standard patient. If dentists intend to keep the patient and have reputational concerns, this should lead to less overtreatment recommendations. Based on this we hypothesize:

**Hypothesis 2.2** (Patient information). *The patient in the informed condition receives an overtreatment recommendation less often than the patient in the standard condition.*

### 2.4.3 Experimental procedure

#### Random draw of dentist sample

Our database listed 865 practicing dentists in the canton of Zurich. We randomly drew 180 dentists from this dentist population. Each of these 180 dentists were visited by our test patient. All visits were conducted in 2016. *Figure 2.1* illustrates the location of the visits. Among the visits, 78 dentists (43.33%) were located in the city of Zurich, 15 (8.33%) in the city of Winterthur and 87 (48.33%) in other municipalities. These shares reflect the population of dentists well among which 402 (46%) is located in the city of Zurich, 70 (8.1%) in the city of Winterthur and 393 (45.4%) in other municipalities of the canton.

*Table 2.3* presents means and standard deviations of the covariates for each of our four experimental conditions. The covariates will be presented in detail in the next section. The ANOVA analysis reveals that the randomization worked well as there are no significant differences between any pairs of means of the same covariate (*Table 2.3*). At a first glance, the differences in waiting time and price level displayed seems to be large. Pairwise tests, however, confirm that these differences are not significant (see *Appendix 2.8.4* in *Tables 2.9* and 2.10).

#### Dentist visits

The 180 visits were conducted as follows: The test patient called the randomly selected dental practices in a randomly determined order. At each call, the test patient asked for the earliest available check-up appointment. After arranging the

---

26 When drawing two dentists from the same practice, we randomly replaced one of the dentists by a randomly drawn dentist from the population.

27 The patient was trained to ask this question in a way that did not signal that seeing a dentist was urgent. When asked about the reason for his visit, he referred to the story in the script. In most cases he did not choose the earliest offered date for the actual appointment.
2.4. The field experiment

FIGURE 2.1
Map of the canton of Zurich, Switzerland. Observations per district and population densities.

appointment, the test person visited the respective dentist. Visits were conducted based on the script. The script indicates that the test patient provides the dentist with the digital x-ray and tells the dentist that he has recently been at a practice for dental hygiene where the x-ray picture had been taken. The patient furthermore says that the dental hygiene assistant recommended seeing a dentist which is why he was here for a check-up. If the dentist proposes a treatment, the test patient is instructed to ask for a cost estimate. After each visit, the test patient completed a detailed protocol about the visit in order to document the communication with the dentist as well as a set of dentists’ and practices’ characteristics.

\footnote{For the time of the study, the test patient received a dental hygiene treatment in regular intervals to support the story. The date indicated on digital x-ray was regularly updated.}


\[ TABLE \, 2.3 \]
Balance of covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ST-LS</td>
<td>INFO-LS</td>
</tr>
<tr>
<td>Waiting time for appointment (days)</td>
<td>8.756</td>
<td>8.267</td>
</tr>
<tr>
<td></td>
<td>(9.815)</td>
<td>(8.874)</td>
</tr>
<tr>
<td>Informative webpage</td>
<td>0.578</td>
<td>0.733</td>
</tr>
<tr>
<td></td>
<td>(0.499)</td>
<td>(0.447)</td>
</tr>
<tr>
<td>Competition density</td>
<td>0.561</td>
<td>0.554</td>
</tr>
<tr>
<td></td>
<td>(0.565)</td>
<td>(0.582)</td>
</tr>
<tr>
<td>Practice price level (PV)</td>
<td>3.840</td>
<td>3.859</td>
</tr>
<tr>
<td></td>
<td>(0.353)</td>
<td>(0.294)</td>
</tr>
<tr>
<td>Price level (PV) displayed</td>
<td>0.422</td>
<td>0.467</td>
</tr>
<tr>
<td></td>
<td>(0.499)</td>
<td>(0.505)</td>
</tr>
<tr>
<td>Swiss licence age (years)</td>
<td>19.22</td>
<td>19.47</td>
</tr>
<tr>
<td></td>
<td>(10.90)</td>
<td>(10.27)</td>
</tr>
<tr>
<td>Practice owner</td>
<td>0.822</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>(0.387)</td>
<td>(0.405)</td>
</tr>
<tr>
<td>Median income in area (cont.)</td>
<td>53.09</td>
<td>52.40</td>
</tr>
</tbody>
</table>

Note: the table reports means and standard deviations (in brackets) for covariates over conditions.

2.5 Comprehensive data on market, practices and dentists

2.5.1 Data sources and variables of interest

We complement the experimental data by a unique dataset that combines information about the market, practices and dentists. For the dentist data we reverted to the Swiss Medical Register (MedReg, Bundesamt für Gesundheit (2015)) and updated the register for recent changes. The Swiss Medical Register provides information on dentists’ gender, nationality, education, licensing, and specialization. We complemented this registry data by information on the practices. We collected the practice information from the dentists’ web-pages and during the visits. Practice
2.5. Comprehensive data on market, practices and dentists

characteristics include the practice age as well as information about the practice owner.

Besides the practice and dentist characteristics, we are particularly interested in variables that reflect the market environment. Therefore, we construct several measures of demand and competition from the data available. In the following, we explain the measurement of all variables in detail. Table 2.4 provides an overview on the variables’ descriptions and measurements.

Market: demand and competition measures

The demand and competition measures are intended to capture the short- and medium-term demand at a dentist, the long-term competitive environment and comprise two price indicators as well. As a measure of the short-term demand at a dentist, we use the waiting time for the next available appointment that was collected at the call made by the test patient. The rationale is that a short waiting time indicates unused capacity in the short run and thus a low short-term demand at an individual dentist. As a medium-term demand measure, we propose to use whether or not a dentist has an informative web page. The logic behind is that dentists whose patient books are not completely full are more likely to put effort into an informative web page to attract more patients. A web page is thereby defined as informative if it provides information on the services offered in the practice or the dentist’s biography.

We use the number of dentist practices in the vicinity adjusted by population size as a measure of long-term competition. In particular, we take the number of other dentists’ practices within a 500 meters distance and adjust it by the number of inhabitants and workers to account for different population densities. This measure, physician density, has been used previously in a large body of work on physician-induced-demand.

Regarding pricing, we have two measures: First, the point value chosen by a dentist’s practice gives the practice’s price level. Second, as a measure of ex ante price transparency, we use whether a not the point value is displayed in the practice. Note that while transparency about the point value is required by regulation, the majority of practices do not display the point value in their practice.

Practice and dentist characteristics & other variables

Our variables of interest with respect to the practice and dentist characteristics are whether or not a dentist owns the practice and since how many years the dentist
## TABLE 2.4
Description of variables on dentists, visits, and the market.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demand &amp; competition measures</strong></td>
<td></td>
</tr>
<tr>
<td>Waiting time for appointment <em>(short term demand)</em></td>
<td>Number of work days between phone call and earliest offered appointment (excluding dentist vacations). Source: phone call protocols.</td>
</tr>
<tr>
<td>Informative webpage <em>(medium term demand)</em></td>
<td>Indicator variable that the webpage of the practice contains information either on the biography of the dentists working in the practice or on the offered spectrum of services. Incorporates responsive design. Binary. Source: own research.</td>
</tr>
<tr>
<td>Competition (adjusted) <em>(long term competition)</em></td>
<td>Dentist density per 1’000 workforce-adjusted inhabitants. Formula: ( A/(B+C^2) \times 1000), where A: number of dentists within a 500m circle around the practice; B: inhabitants in the municipality/city district (Zurich and Winterthur); C: Full-time equivalent work-force in the municipality (rural areas) or city district (Zurich and Winterthur). Source: own research.</td>
</tr>
<tr>
<td>Practice price level (PV)</td>
<td>Value in CHF that a practice attaches to every point it charges. The number of points that can be attached to a service is defined in the Swiss Dental Tariff. The practice has to use the same PV (point value) for all patients across all services with some exceptions for state welfare recipients. Source: diagnosis bills.</td>
</tr>
<tr>
<td>Price level (PV) displayed</td>
<td>Indicator whether the visited practice complies with the obligation to display its price level (PV) in the practice well visible. Source: test patient protocols.</td>
</tr>
<tr>
<td><strong>Dentist &amp; practice variables</strong></td>
<td></td>
</tr>
<tr>
<td>Swiss licence age</td>
<td>Number of years the dentist has been possessing his/her licence for working in Switzerland (in 2016). Source: MedReg</td>
</tr>
<tr>
<td>Practice owner</td>
<td>Indicator whether the treating dentist is the owner or one of the owners of the practice. Sources: own research of webpages and phone books, telephone calls with practice assistants, MedReg.</td>
</tr>
<tr>
<td><strong>Other variables</strong></td>
<td></td>
</tr>
<tr>
<td>Median income in area</td>
<td>Median income per year (in k CHF, 2013) in the municipality (and on district level in the city of Zurich) of the practice. Source: Statistical Office of the Canton of Zurich.</td>
</tr>
</tbody>
</table>
2.5. Comprehensive data on market, practices and dentists

is licensed to practice in Switzerland. Besides reflecting the age of the dentist, the license age also gives an indication whether dentists still repay loans for initial investments when setting up practice, or whether these are already paid off. We furthermore look at the interaction of having an informative web page and the license age. The rationale behind that is that dentists with a higher license age are more likely to make their website informative because they plan to sell the practice, rather than to attract patients. Hence, we expect an informative web page to be an indicator of medium term demand measure and, as a result, to have an impact on overtreatment recommendations, only for younger dentists and lower license ages.

As an additional variable, income within the vicinity of a practice is included in the analysis, as a higher local income may imply different practice and treatment recommendation styles.

2.5.2 Descriptives

We first describe the data on dentists before turning to the visit and market characteristics. The visited dentists were females in 32.2% (58/180) and males in 67.8% (122/180) of the cases. Dentists had been possessing their approbation for on average 21.7 years (sd: 9.28), with a minimum of 3 and a maximum of 49 years. The average license age was 19.41 years (sd: 10.31) (see Figure 2.3a). In total, 47.8% (86/180) of the dentist were working in a single practice and 52.2% (94/180) dentists were working in a group practice with at least one other dentist. The dentists visited obtained their diploma in Switzerland in 73.9% (133/180), in Germany in 15.6% (28/180) and in other countries in 10.6% (19/180) of the cases.

Our test patient visited the dentists on average 7.59 work days (sd: 9.71) after the date of the first possible appointment that was offered. When entering the practice, the test patient waited on average 6 minutes and 33 seconds (sd: 7.36, min: 0, max: 35) until he was asked to follow to the examination room. On average, the test patient spent 19 minutes and 40 seconds in the examination room (sd: 7.57, min: 5, max: 50).

Regarding the market variables, we find that the waiting time for the next appointment ranges from zero to 43 work days. On average, the next possible appointment was offered between eight and nine work days from the initial phone call (mean: 8.78 work days, sd: 8.90). The distribution of the waiting time is illustrated in Figure 2.3b.

The descriptives show that 66% of the dentists had an informative web-page at the time when the test patient’s visits started. The number of competitors within
FIGURE 2.2
Distribution of key variables.

(a) Distribution of dentists’ license (approbation) age.

(b) Distribution of the waiting time for the next possible appointment.

(c) Distribution of number of dentists in 500m perimeter around practice.

A radius of 500m ranged from 0 to 61 (see Figure 2.3c). Adjusting the number of competitors for the number of inhabitants and workers in the municipality, we find our long-term competition measure physician density to range from 0 to 2.33 with an average of 0.65 (sd: 0.58). The price level of practices ranged from factor 2.8 to 4.85 (average: 3.88, sd: 0.49). Only 39% of the practices displayed the price level in the practice. Table 2.5 summarizes the descriptives.
2.6. Results

2.6.1 Treatment recommendations

Our test patient received an overtreatment recommendation in more than every fourth visit. More precisely, dentists suggested at least one filling in 27.78% (50/180) of all visits. In Table 2.6, we show the overtreatment rate for each of the four conditions.

Conditional on an overtreatment recommendation, mean overtreatment costs taken from the collected cost estimates amount to CHF 535 (about $550), the median being lower at CHF 444 (about $455). Regarding the treatment, the suggested number of fillings per dentist ranges from 1 to 6. An illustration is provided in Figure 2.4. Furthermore, we observe across all cost estimates that 13 different teeth are to be treated with a filling. Thus, besides our finding of a considerable overtreatment recommendation rate, we also observe a striking heterogeneity in the treatment recommendations.

We perform a parametric analysis on our binary overtreatment variable using a random effects probit regression model. The explanatory variables of interest are the two binary factor dummies, informed and high_ses, the interaction ef-
Chapter 2

### TABLE 2.6
Overtreatment recommendations per conditions. Number of observations in parentheses.

<table>
<thead>
<tr>
<th>Information</th>
<th>Standard</th>
<th>Informed</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES Low</td>
<td>37.78% (17/45)</td>
<td>26.67% (12/45)</td>
<td>32.22% (29/90)</td>
</tr>
<tr>
<td>High</td>
<td>20.00% (9/45)</td>
<td>26.67% (12/45)</td>
<td>23.33% (21/90)</td>
</tr>
<tr>
<td>Average</td>
<td>28.88% (26/90)</td>
<td>26.67% (24/90)</td>
<td>27.78% (50/180)</td>
</tr>
</tbody>
</table>

### FIGURE 2.4
Cost estimates per number of fillings for overtreatment recommendations.

---

The effect²⁹ between the conditions $informed_i \times high_{ses}$, and the different measures for competition $short\_term\_demand$, $medium\_term\_demand$, $long\_term\_competition$, $price\_level$ and $price\_level\_displayed$. $X_i$ reflects a vector of practice and dentist covariates. This vector includes the age of the dentist’s license to practice in Switzerland and an indicator for whether the dentist is the practice owner. Further, we include an interaction term between our measure for mid-term demand, informativeness of the webpage, and the Swiss licence age. Last, we control for the median income in the area of the dentists practice. Hence, our specification is as follows:

$$overtreatment_i = \beta_0 + \beta_1 informed_i + \beta_2 high_{ses_i} + \beta_3 (informed_i \times high_{ses_i}) + \beta_4 short\_term\_demand_i + \beta_5 medium\_term\_demand_i$$

²⁹Note that the interpretation of regression coefficients of interaction terms in non-linear regressions might be misleading (see Ai and Norton (2003)). We therefore also display the results from a OLS regression in model (5) in Table 2.7. Both regressions show consistent results.
2.6. Results

\[ + \beta_6 \text{long term competition}_i + \]
\[ + \beta_7 \text{price level} + \beta_8 \text{price level displayed} \]
\[ + \beta_8 X_i + \beta_9 \text{median income in area} + \epsilon_i \]

Table 2.7 provides the results. To deepen the analysis and account for the extent of overtreatment, we run the same model with two more dependent variables, the number of recommended fillings and the cost estimate. Figure 2.5 shows the distributions for both number of recommended fillings and the cost estimate. The number of recommended fillings represents count data with values between zero and six. The cost estimate size displays a typical pattern for health care costs. While there are more than 70% zero observations, the smallest cost estimate conditional on overtreatment is 56 points. The distribution further displays a long right-tail with only two observations larger than 300 points with values of 390.5 points and 419 points, respectively. We ran a negative binomial regression for the number of recommended fillings. The choice of the negative binomial model over the poisson model is akin to overdispersion of the data due to the large fraction of zeros. This choice is supported by a likelihood-ratio test. For the cost estimate size we present a GLM estimation with gamma distribution and log-link. The choice of the gamma distribution is indicated by a Modified-Park Test. Table 2.8 shows the results. Model (M3) from Table 2.7 is displayed for comparison in column (1).\(^{30}\) In the following, we start with the analysis of the experimental variations. We will present and discuss our results without considering diagnostic errors. The discussion of our results in light of potential diagnostic error is provided separately in subsection 2.6.2.1.

2.6.1.1 Effects of socio-economic status and information

The probit regressions show that the likelihood to receive an overtreatment recommendation is significantly lower for a patient with a high- than a low SES (see Table 2.7) in the standard condition. Being a standard patient with a high SES reduces the likelihood of receiving an overtreatment recommendation by about 17 percentage points compared to a patient with a low SES. In contrast to the standard condition, differences in SES do not translate into different overtreatment

\(^{30}\text{We also considered a hurdle model with model (M3) from Table 2.7 in the first part and estimations of the amount of overtreatment, conditional on receiving an overtreatment recommendation, in the second part. With only 50 observation for the second part of the model, however, the poisson regression did not have the power to identify robust effects on the amount of overtreatment.}\)
### TABLE 2.7

Regressions on a binary measure of overtreatment.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(M1)</th>
<th>(M2)</th>
<th>(M3)</th>
<th>(M4)</th>
<th>(M5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Probit</td>
<td>OLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>-0.102</td>
<td>-0.108</td>
<td>-0.112</td>
<td>-0.120*</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td>(0.214)</td>
<td>(0.174)</td>
<td>(0.134)</td>
<td>(0.099)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>High SES</td>
<td>-0.174**</td>
<td>-0.175**</td>
<td>-0.173**</td>
<td>-0.177**</td>
<td>-0.178*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.023)</td>
<td>(0.018)</td>
<td>(0.012)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Informed x High SES</td>
<td>0.181</td>
<td>0.159</td>
<td>0.143</td>
<td>0.128</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>(0.225)</td>
<td>(0.267)</td>
<td>(0.297)</td>
<td>(0.343)</td>
<td>(0.279)</td>
</tr>
<tr>
<td>Demand &amp; competition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting Time for Appointment (days)</td>
<td>-0.011***</td>
<td>-0.010***</td>
<td>-0.009**</td>
<td>-0.007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.014)</td>
<td>(0.026)</td>
<td>(0.044)</td>
<td></td>
</tr>
<tr>
<td>Informative Webpage</td>
<td>0.093</td>
<td>0.281*</td>
<td>0.236</td>
<td>0.282*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.095)</td>
<td>(0.161)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td>Competition Density</td>
<td>0.008</td>
<td>0.030</td>
<td>0.019</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.893)</td>
<td>(0.590)</td>
<td>(0.732)</td>
<td>(0.873)</td>
<td></td>
</tr>
<tr>
<td>Practice price level (PV)</td>
<td>-0.096</td>
<td>-0.026</td>
<td>0.008</td>
<td>-0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.374)</td>
<td>(0.808)</td>
<td>(0.940)</td>
<td>(0.942)</td>
<td></td>
</tr>
<tr>
<td>Price level (PV) displayed</td>
<td>-0.148***</td>
<td>-0.176***</td>
<td>-0.154***</td>
<td>-0.168**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Practice &amp; dentist variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swiss Licence Age (years)</td>
<td>-0.003</td>
<td>-0.006</td>
<td>-0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.530)</td>
<td>(0.277)</td>
<td>(0.258)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inform. Webpage x Swiss Licence Age</td>
<td>-0.013***</td>
<td>-0.011*</td>
<td>-0.011*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.087)</td>
<td>(0.087)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice Owner</td>
<td>0.164*</td>
<td>0.156*</td>
<td>0.185*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.093)</td>
<td>(0.055)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Income in Area (cont.)</td>
<td>0.007</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.237)</td>
<td>(0.256)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overdiagnosis (conservative definition)</td>
<td>0.210**</td>
<td>0.219**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.728)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.017</td>
<td>0.082</td>
<td>0.160</td>
<td>0.184</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td>0.199</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td></td>
</tr>
</tbody>
</table>

(M1)-(M4): Probit regressions, displaying average marginal effects
(M5): OLS regression
(M1)-(M5): p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
### TABLE 2.8
Regression results on the likelihood (models 1 and 2) and amount (models 3 and 4) of overtreatment.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable</strong></td>
<td><strong>Probit</strong></td>
<td><strong>NBREG</strong></td>
<td><strong>GLM</strong></td>
<td><strong>Probit</strong></td>
</tr>
<tr>
<td>Overtreatment (binary)</td>
<td>-0.112</td>
<td>-0.120*</td>
<td>-0.182</td>
<td>-15.141</td>
</tr>
<tr>
<td>Number of fillings</td>
<td>0.143</td>
<td>0.128</td>
<td>0.221</td>
<td>22.545</td>
</tr>
<tr>
<td>Cost estimate (points)</td>
<td>(0.134)</td>
<td>(0.099)</td>
<td>(0.314)</td>
<td>(0.285)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Treatments</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>-0.173**</td>
<td>-0.177**</td>
<td>-0.444***</td>
<td>-39.463**</td>
</tr>
<tr>
<td>High SES</td>
<td>(0.018)</td>
<td>(0.012)</td>
<td>(0.003)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Informed x High SES</td>
<td>0.143</td>
<td>0.128</td>
<td>0.221</td>
<td>22.545</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.343)</td>
<td>(0.578)</td>
<td>(0.259)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Demand &amp; Competition</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Waiting Time for Appointment (days)</td>
<td>-0.010**</td>
<td>-0.009**</td>
<td>-0.033***</td>
<td>-2.136***</td>
</tr>
<tr>
<td>Informative Webpage</td>
<td>0.281*</td>
<td>0.236</td>
<td>0.759</td>
<td>75.610***</td>
</tr>
<tr>
<td>Competition Density</td>
<td>0.030</td>
<td>0.019</td>
<td>0.158</td>
<td>4.122</td>
</tr>
<tr>
<td>Practice price level (PV)</td>
<td>-0.026</td>
<td>0.008</td>
<td>-0.232</td>
<td>-2.881</td>
</tr>
<tr>
<td>Price level (PV) displayed</td>
<td>-0.170***</td>
<td>-0.154***</td>
<td>-0.301***</td>
<td>-26.788**</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.026)</td>
<td>(0.006)</td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Practice &amp; Dentist Variables</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Swiss Licence Age (years)</td>
<td>-0.003</td>
<td>-0.006</td>
<td>-0.006</td>
<td>0.723</td>
</tr>
<tr>
<td>Inform. Webpage x Swiss Licence Age</td>
<td>-0.013**</td>
<td>-0.011*</td>
<td>-0.035**</td>
<td>-3.494***</td>
</tr>
<tr>
<td>Practice Owner</td>
<td>0.164*</td>
<td>0.156*</td>
<td>0.558*</td>
<td>47.413***</td>
</tr>
<tr>
<td>Other Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Income in Area (cont.)</td>
<td>0.007</td>
<td>0.005</td>
<td>0.026**</td>
<td>1.980**</td>
</tr>
<tr>
<td>Overdiagnosis (conservative definition)</td>
<td>0.007</td>
<td>0.005</td>
<td>0.026**</td>
<td>1.980**</td>
</tr>
</tbody>
</table>

| Pseudo $R^2$ | 0.160 | 0.184 | 0.082 |
| N            | 180   | 180   | 180   |

(1)-(2): Probit models displaying average marginal effects.
(3): Negative binomial regression displaying average marginal effects.
(4): GLM regression with Gamma-distribution and log-link displaying marginal effects at the mean.
(1)-(4): p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
recommendation levels when the patient is informed. Both, under low and high SES, the level of overtreatment recommendations amounts to 26.67%. This result of lower overtreatment with a higher SES as a standard patient is surprising at first thought. In particular, a higher SES—implemented by a more expensive physical attire indicating a higher income while keeping the education level constant—might be interpreted as implying a lower price sensitivity and higher acceptance rate of costly treatment, which increases overtreatment incentives. Thus, one might have expected that the higher SES works in a similar way as the effective consumer price reduction of having (health) insurance. Kerschbamer et al. (2016) for instance show in a field experiment in the market for computer repair that customers’ insurance coverage increases the repair price significantly. Our result suggests that in the patient-physician interaction, another mechanism is at work for the SES variation. We will discuss the potential explanations in turn.

Our finding may be explained by the similarity of patients’ and dentists’ SES. Van Ryn and Burke (2000), e.g., find that patients from the lower socio-economic class are perceived more negatively by physicians than patients from the middle and high economic class. In a field experiment on taxi driver behavior, Balafoutas et al. (2013) for instance find that taxi customers with a lower perceived income — closer
to the low SES taxi driver—are overcharged less often than those with a perceived high income. Note however, that in this case the effect of a similar SES and the economic argument of price sensitivity go in the same direction, contrary to our case.

Another aspect is that the perceived likelihood to return might differ between SES for the standard patient. The importance of reputation-building might play a role.\textsuperscript{32} It could be argued that a physician attempts to build-up reputation by not treating a patient with a minor treatment such as a filling, and that this reputation-building concern is higher for patients with a higher SES due to higher future profits from interaction. However, it is not clear whether reputation-building is actually stronger when not providing a treatment initially than when providing a treatment. Another argument related to the likelihood to return is that overtreatment for the standard patient with a lower SES might be preventive in the sense that the dentists expect this patient to be less likely to return to any dentist for check-ups in the future.\textsuperscript{33} Note however that this explanation is not particularly convincing for the following reason: By going to the dentist with his x-ray from a dental hygiene practice, the patient shows a considerable interest in his dental health.\textsuperscript{34}

A final point relates to an interpretation of a higher SES as implying \textit{better or more information}. In the experimental design, we took care to keep the level of education constant across the variation such that the higher SES goes through higher income but not education level. However, the patient with higher SES might still be perceived as better informed. We will discuss this further below in light of our results from the experimental information variation.

More information as implemented in our experiment does not significantly reduce overtreatment. We observe a considerable drop in the rate of overtreatment recommendations from 37.78\% to 26.67\% between an informed and a standard patient if the SES is low. However, this difference is not statistically significant. We neither find a significant difference between an informed and a standard patient if the SES is high nor if SES treatments are pooled.

Our results suggest that it is crucial to understand and differentiate between the different types of information that are relevant in the physician-patient interaction and in particular how these are perceived by physicians. Signalling information from

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{32}In our experimental variations, due to limitations on the number of visits for our test patient, we could not run a further variation in which the role of reputation-building in repeated physician-patient interactions is analyzed.
\item \textsuperscript{33}This explanation was suggested by one of our reference dentists.
\item \textsuperscript{34}Furthermore, while the patient displays a lower SES compared to the other condition, the patient is by no means a case of very low social status.
\end{itemize}
\end{footnotesize}
the online platform—even when it is on the specific case at hand—does not appear to considerably affect the treatment recommendation of dentists. This may reflect the fact that dentists do not perceive information from the platform as significantly reducing the level of asymmetric information between them and the patient. Furthermore, it might be the case is that dentists assume that they can rationalize their treatment recommendation, independent of differing diagnostics/information from the website. It is generally assumed that information provision and diagnostics from the internet increase patient information and quality of care. Our results show the limits of this argument when the service at hand has credence goods characteristics and is complex, as is the case for most health care services. The importance of the fineprint of what constitutes relevant information and limits of information for credence goods is also apparent when comparing our results to the literature: In Currie et al. (2011) and Currie et al. (2014), the authors sent students, trained as test patients, with identical verbally communicated flu-like complaints and find that patients that signal that they are informed about inappropriate antibiotic use are prescribed less antibiotics than other patients. Observe that this is a case where both diagnostics and information about correct/wrong treatment is simple, unambiguous and not patient-case specific such that the case does not keep credence goods characteristics with simple information provision. There, signalling the corresponding information should reduce wrong prescriptions, as observed empirically. Our results point at the difficulty of information and diagnostics via the internet or other sources to address the problem of wrong treatment—be it due to diagnostic errors or physician-induced demand—for more complex cases of health services.\footnote{Interesting survey evidence is provided by Domenighetti et al. (1993). They find that physicians have a much lower rate of surgeries than regular patients who are not physicians or who do not have physicians (or lawyers) in their families. This evidence is consistent with the information asymmetry hypothesis, however other explanations can also account for this observation, e.g., physicians might have a generically different demand or face different prices.}

2.6.1.2 The role of market and dentist characteristics

Besides the striking main result of an overtreatment rate of 28%, one of our key findings pertains to the role of demand. Model (2) in Table 2.7 shows that a higher short-term demand, measured by a longer waiting time for an appointment, is associated with a lower likelihood to be overtreated. An additional waiting day reduces the probability to be overtreated by, on average, one percentage point. This result is already apparent in the descriptive statistics. We observe that overtreating dentists have an average waiting time for an appointment of 6.16 days whereas
not overtreating dentists have a waiting time of 9.78 days (Mann-Whitney U test, two-tailed (MWU): $p = 0.0008$) (see Figure 2.6).

**FIGURE 2.6**

Waiting time for the next possible appointment (days) by overtreatment.

This result is in line with physicians filling up capacities by overtreating patients, which has been previously been pointed out in theoretical work by Emons (1997, 2013). The low utilization of dentists’ services leads to a lower profit in the short run and thus increases the incentives to provide more services than necessary, which is consistent with profit maximization and the target-income hypothesis. Our result is also consistent with early empirical findings by Marty (1998). Marty analysis data from an insurance company on physicians’ individual treatment decisions and defines upward deviations from the average turnover per patient as overtreatment. He shows that physicians’ idle capacities are in fact positively correlated with above average turnovers.

A possible concern with respect to our analysis may be simultaneous causality. If overtreatment was observed by patients they would possibly switch to a different dentist leading to a lower waiting time at overtreating dentists. However, patients cannot observe whether or not they received the appropriate treatment due to the credence good characteristic. Hence, the measured impact of waiting time on overtreatment does not suffer from reverse causality. With respect to the reason behind short waiting times, it might well be that these dentists are perceived by patients to provide low quality service and therefore only face low demand. Unfortunately, we cannot control separately for perceived quality. Although online rating

---

36For a discussion of the target income hypothesis in the context of physician behavior, see, e.g., McGuire (2000).
websites start to be in place for dentists, we do not have enough data to construct a meaningful measure of perceived quality. However, this does not pose a problem for our result for which we only require waiting time to be a valid proxy for short-term demand, where low demand might be the result of different factors such as low perceived quality or demand shocks.

As a measure of medium-term demand, we use whether or not a dentist has an informative web page. An informative web page indicates effort to provide information to existing but particularly prospective patients. Thus, we consider an informative web page as a means to increase appeal to potential patients, indicating that the dentist wants to increase the customer base to increase demand. The results show that dentists with an informative web page are in fact more likely to overprovide than dentists working under high medium-term demand. The descriptive statistics reveal that the overtreatment rate is 31.10% for dentists with an informative web page and drops by almost ten percentage points to 21.31% for dentists without an informative webpage. The effect is not significant, however (Fisher’s exact test: $p = 0.218$), analogues to not being significant in regression model (M2) in table 2.7 when we do not control for practice and dentist variables as well as area income. The interaction term between informative web page and the license age is negative in (M3) and significant. This result is in line with the argument that an informative web page is a demand indicator for dentists primarily at the beginning of their career when building up the patient base. Among the dentists with a licence age below or at the median licence age of 18 years who have an informative webpage, the overtreatment rate is 43.28% (29 of 67 cases). The rate drops to 15.38% (8 of 52 cases) when the dentists are above the median licence age (Fisher’s exact test: $p = 0.001$).

We do not find a significant effect of our long-term competition measure, dentist density. This result holds even if we do not control for the other competition measures such as short-term demand.\textsuperscript{37} It is an important result in light of the previous literature on physician-induced demand: A vast literature on PID approaches the topic by analyzing the correlation between physician density and a measure of health care utilization such as annual number of procedures per general practitioner. The idea behind is that a higher density implies lower demand per physician, which is compensated by physician-induced demand. This literature generally finds a significant association between physician density and health care consumption.\textsuperscript{38}

\textsuperscript{37}This result is robust to different definitions of the variable such as using an non-population adjusted measure.

\textsuperscript{38}For an overview and comparison of these studies, see Leonard et al. (2009).
problem with the used measures of health care utilization is that the researcher does not know the correct diagnostic and thus cannot correctly classify the provided treatments. Our design allows us to do exactly that, and we do not observe that physician density has a significant effect on overtreatment. It appears that the effect of the PID logic is better captured by our short-term demand measure.

Our results show that the likelihood to receive an overtreatment recommendation decreases by approximately 16 percentage points (see Model (3) in Table 2.7) at a dentists who displays the price level in the practice as required by regulation. However, we do not find evidence that the price level itself affects overtreatment. This latter result seems counterintuitive at first, as the price level directly impacts the financial overtreatment incentives. However, for the case at hand, the difference in final price across different price levels for the treatment considered is rather small, which can explain the result. For treatments with more high powered incentives, this might be different.\textsuperscript{39} A possible interpretation for the first result on price level transparency and overtreatment is that there might be different dentist types in terms of following regulation (display price level in practice) and treatment guidelines (overtreatment). When treated by a practice owner, the probability to be overtreated is increased by more than 16 percentage points. Although only weakly significant, this effect seems intuitive. Practice owner are residual claimants by nature while non-practice owners are often employed on a fixed income basis\textsuperscript{40}.

\subsection*{2.6.2 Diagnoses}

We observe a considerable dispersion in diagnosis fees, illustrated in Figure 2.7.\textsuperscript{41} The lowest fee, charged twice, was CHF 0, the highest fee amounted to CHF 212.65. On average, dentists charged CHF 92.62 (sd: 31.17) for the diagnosis.\textsuperscript{42} Even when correcting for the price level (PV, point value) of the practice, a considerable dispersion persists (right side of Figure 2.7). In the sample, dentists charged between zero and 56 points for the diagnosis with a mean of 23.95 points (sd: 8.02). About half of the sampled dentists charged exactly 21 points for the

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{39}It was our intention to send out a test patient with a case with more high powered incentives alongside the case presented here. However, we could not find a test patient with such a case for which all requirements for conducting the field experiment could be met.
\item \textsuperscript{40}This information was provided to us by our reference dentists.
\item \textsuperscript{41}Our test patient received the bills for the check-up visits on average 14.89 work days after the visit (sd: 34.93). The median is only six days, however. While 34 dentists charged directly on the spot, a procedure not uncommon for first visits, seven dentists sent their bills only after 100 days or more after the visit, the maximum being 235 days.
\item \textsuperscript{42}The total of diagnosis fees paid for our study amounts to CHF 16'671.
\end{itemize}
\end{footnotesize}
diagnosis, which is the specified rate for a standard consultation. The average point
value attached to one point is CHF 3.88 (min: 2.8, max: 4.85, sd: 0.30).

FIGURE 2.7
Distribution of diagnosis fees and points charged per diagnosis (n=180).

FIGURE 2.8
Points charged by SES (left) and patient information (right) (n=180).

With respect to our experimental conditions, diagnosis fees and points charged
per diagnosis do not differ between low and high SES patients (see Figure 2.8 (left)).
However, we observe a (weakly) significant difference with respect to the change in
signalled information: dentists charged weakly more points for the informed patient
than for the standard patient (see Figure 2.8 (right)). The average number of
points charged for the standard patient is 22.91 (sd: 6.81) and 25.00 (sd: 8.99)
2.6. Results

for the informed patient (MWU: p = 0.094). This result holds when considering diagnosis fees without adjusting for point value. The diagnosis fee for an informed test patient is CHF 88.96 (sd: 28.08) for the standard and CHF 96.28 (sd: 33.74) for the informed patient (MWU: p = 0.085).

Looking further into the diagnosis fees, we distinguish between different diagnosis items as classified by our reference dentists: consultation depending on attached points billed, further diagnosis items that are admissible and diagnosis items that constitute overdiagnosis. Overdiagnosis refers to items billed which are not needed for diagnosing the case or providing the necessary information to the patient. An example is the item ‘further information about dental interventions’ with 15 points. In the Swiss Dental Tariff, it is explicitly stated that this item should not be applied for information to patients about routine dental procedures, to which for instance fillings belong. We identified 27 out of 180 visits (15.00%) with overdiagnosis. Figure 2.9 shows the comparison by diagnosis items, in absolute and relative terms, between the standard and the information conditions.

Figure 2.9 shows that the main difference between standard and informed in points charged stems from a difference in overdiagnosis. This reflects the comparison of the number of visits with overdiagnosis across the information conditions: 9 out of 90 in standard compared to 18 out of 90 in informed (Fisher’s exact test: p = 0.094). Note that most overdiagnosis items relate to more time spent at the visit. Looking into the time spent in the treatment room, we indeed observe that it is significantly longer for visits with overdiagnosis (25.04 minutes) than without overdiagnosis (18.60 minutes). Thus, dentists appear to be spending (unnecessarily) more time, which is then billed. Interestingly, when running regressions on overdiagnosis value (CHF), we again find that the coefficient for short-term demand is significant and has the same sign as for overtreatment.

Model (4) in Table 2.7 shows that overdiagnosis is also associated with a higher likelihood to be overtreated. Clearly, endogeneity is a problem at hand, since providing a treatment recommendation for a filling might induce more time spent

---

43When excluding the 17 dentists who only charged 14 points or less—14 points are a natural choice because it is the number attributed to the tariff item for the diagnosis of a ‘recall patient’—the difference of the means becomes more pronounced. The rationale for exclusion of these dentists is that these are likely neither interested in present nor future profits from the test patient. The average number of points charged for the informed test patient is then 23.97 (sd: 6.00) and 26.45 (sd: 8.14) for the standard patient (MWU: p = 0.032).

44The cumulative distribution functions of diagnosis fee and points charged are shown in Figure 2.12 in Appendix 2.8.1.1 separately for informed and standard patient.

45This provides the basis for the classification as overdiagnosis and not as overcharging at diagnosis.

46Our data does not provide sufficient power to identify a significant effect of short-term demand on the likelihood of receiving an overdiagnosis.
Chapter 2

with the patient that is then billed via additional diagnosis items. Indeed, we find a significantly higher time spent in the treatment room for overtreatment than for no overtreatment recommendations (22.12 minutes vs. 18.73 minutes; MWU: p=0.0242).

FIGURE 2.9
Composition of points charged absolute (left) and relative (right) scales.

2.6.2.1 Discussion

Robustness  In further specifications of our regression analysis, we controlled for additional dentist, visit and market characteristics such as questions the dentists asked during the visit, whether the practice was a single practice or shared by more than one dentist, gender and membership in Swiss dental association. However, we do not find a significant impact of these possibly explanatory variables. Our results are not only robust against changes in the specification but also against different regression models, such as the logit, negative binomial or zero-inflated model. We also investigate whether our test person’s behavior changed over time as the test patient became more experienced. However, seasonal dummies do not show to have an impact on the regression results. Also, we make use of several precautionary measures to prevent that the medical condition of our test patient changed over time. First, all diagnoses have been based on the same x-ray which our test patient brought to every visit. Second, the oral condition of our test patient was confirmed after 30, 60, 120 and 180 visits by our reference dentists.
Diagnostic errors  In the design, we took care to select a case for which the scope of diagnostic error should be minimal. However, we cannot fully exclude diagnostic errors. Taking these into account, the interpretation of our results needs to be amended in the following way: First, regarding the information variation, diagnostic errors can explain that we do not find a significant difference between standard and informed. If dentists are not aware that their treatment recommendation is an overtreatment recommendation, an information signal from the patient does not change recommendations. Second, diagnostic errors may also play a role in our result on short-term demand in the following way: If there is a high correlation between dentists who make diagnostic errors and recommend overtreatment and dentists that are perceived as of low quality by the patients and the latter are less frequented by patients, then this can explain why we observe more overtreatment with shorter waiting however. Note, again that patients cannot observe overtreatment. Furthermore, from the credence goods characteristics of health care services, it is difficult for patients to infer true quality and it is not generally clear that perceived quality corresponds to actual quality.\footnote{Fornara et al. (2006) e.g. show that hospital users’ perceived quality of care improves when the humanization degree of the hospital environment increases. For environmental factors, Arneill and Devlin (2002) conducted a study where they showed participants slides of doctors’ waiting rooms and then asked what quality of care participants expected. Arneill and Devlin (2002) find that the perceived quality of care would be significantly higher for waiting rooms that are nicely furnished, light, contain artwork and are warm versus waiting rooms that are dark, have outdated furnishings, contain no artwork or poor quality reproductions and are cold in appearance.} Taken together, we cannot exclude that our results stem, at least in part, from diagnostic errors. This however does not inhibit the main results: We observe a high rate of overtreatment, be it from diagnostic error or physician-induced demand, and overtreatment is associated with a low short-term demand at the dentist.

Overtreatment results and the scope for second opinions  Combining our results on overtreatment and diagnosis costs, we observe that searching for a second opinion from another dentist might be worthwhile. Assume that a patient’s prior for needing a treatment such as a filling is \( \rho \). Denote by \( P_H \) the price that the patient has to pay when being treated and by \( P_L \) when not being treated. Now if physicians give a treatment recommendation with probability \( x \) to a patient who does not need treatment, and the patient’s costs for searching is \( k \), then, under risk neutrality, searching for a second opinion is worthwhile for the patient if

\[ P_H > k + \frac{\rho}{\rho + (1 - \rho)x} P_H + \frac{(1 - \rho)x}{\rho + (1 - \rho)x} (xP_H + (1 - x)P_L). \]
If the customer does not search for a second opinion, she pays the price $P_H$ for sure. If she searches for a second opinion, she incurs search costs of $k$ and again has to pay the high price $P_H$ if she indeed needs treatment (which happens with probability $\frac{\rho}{\rho+(1-\rho)x}$) or receives an overtreatment recommendation again (which happens with probability $\frac{(1-\rho)x^2}{\rho+(1-\rho)x}$). She only pays the lower price $P_L$ if she does not need treatment and does not receive an overtreatment recommendation on her second opinion visit (which happens with probability $\frac{(1-\rho)x(1-x)}{\rho+(1-\rho)x}$). Using average overtreatment costs of CHF 535 for $P_H$, 0 for $P_L$, the overtreatment rate of 28% for $x$ and average diagnosis costs of CHF 93 for $k$ (abstracting thus for a start from other search/opportunity costs), searching for a second opinion in our case is worthwhile if $\rho < 0.47$. Assuming additional search/opportunity costs of CHF 50 on top of the additional diagnosis costs for a second opinion, this reduces $\rho$ to 0.32.

Thus, our case illustrates that as long as both the likelihood of needing a treatment and opportunity costs are not very high, second opinions can be sensible. In many health care markets, health insurers are actually increasingly incentivizing second opinions.\footnote{In Switzerland, e.g., some insurers grant a discount of up to 15% if insurees search for a second opinion before undergoing surgeries such as artificial hip or knee joints or planned caesareans.} Our results show that for cases for which both the likelihood of needing a treatment and opportunity costs are not very high, this might reduce overall costs. Furthermore, incentivizing second opinions might lead to a reduction in overtreatment rates and thus have an additional benefit.\footnote{In a neutrally framed credence goods laboratory experiment, Mimra et al. (2016b) show that the introduction of second opinions significantly reduces overtreatment rates.}

\subsection*{2.7 Conclusion}

We present the results from a field experiment in the market for dental care in Switzerland. Employing a single test patient who undertook 180 dentist visits, we find that overtreatment is an important phenomenon: The test patient receives an overtreatment recommendation on more than every fourth visit. Using a comprehensive set of measures for market conditions, we find that lower short-term demand as measured by a shorter waiting time for the next possible appointment is associated with a significantly higher likelihood of receiving an overtreatment recommendation. In contrast to a large body of literature on physician-induced demand that relates physician density to mostly aggregate measures of health care consumption, we do not find a significant impact of dentist density on overtreatment recommendations. For our experimental condition variations, we observe significantly less overtreatment recommendations for a patient with a higher socioeconomic status (SES) than
2.7. Conclusion

a patient with a lower SES under standard information. Those differences diminish in the condition in which the patient gives a signal of additional information from an online platform to the physician. This suggests that there is a complex role of SES as well as interactions between SES and signalled information that requires further research, in particular to understand the scope and limits of signalling of information in credence goods markets in general and health care markets in particular.
2.8 Appendix

2.8.1 Additional figures and photos

2.8.1.1 Figures

**FIGURE 2.11**
Distribution of point values, diagnosis costs and overtreatment.

**FIGURE 2.12**
Cumulative distribution of diagnosis fees and points charged by information status (n=180).
2.8.1.2 Experimental conditions

**FIGURE 2.13**
Appearance of the test patient and accessory in the SES variation.

**FIGURE 2.14**
Screenshots of the mentioned web forum in the *informed* condition.
2.8.2 Population of dentists

The population list of dentists used in this field experiment comprised 865 entries. The list was based on the publicly available MedReg register issued by the Swiss Federal Office of Public Health. MedReg comprises all dentists in Switzerland with a valid working licence. The MedReg contains information on name, status and age of the working permission and approbation, the permission to sell pharmaceutical products\textsuperscript{50}, and some more characteristics. The register had 1'151 entries for the canton of Zurich as of October 2015. We deleted dentists with specializations such as child care, orthodontia and aesthetic surgery as well as double entries. Moreover, we used information provided on the webpages of practices and in the yellow pages to update the database. Dentists who did not practice or who had retired were deleted and some, mostly young, dentists were added.

2.8.3 Detailed information on the recruitment and training process of the test patient

2.8.3.1 Recruitment

We searched for potential test patients through two channels. First, we advertised on the online-platform 
Marktplatz, run by the University of Zurich and the ETH Zurich. Second, we sent about 6'000 emails using the subject pool of the Department of Psychology at the University of Zurich. Our test patient was eventually recruited via the Marktplatz platform.\textsuperscript{51} After telephone interviews with interested candidates, the most promising candidates were invited to visit one of our expert dentists together with one of the authors of this paper in order to check whether the candidate was suited for the study or not. The recruitment process continued with an assessment of the candidates’ cognitive skills and reliability. Finally, we recruited a male person in his mid-twenties for the study in late 2015.

2.8.3.2 Training

After we had recruited our test patient, a detailed visiting script was developed for the test patient. The script was developed under consideration of the patient’s real characteristics and histories in order to make the implementation of the roles as easy as possible. After the script had been developed, the test patient was

\textsuperscript{50}In Switzerland, physicians can obtain the permission to sell drugs themselves.

\textsuperscript{51}Marktplatz is a trading platform provided by the University of Zürich and the Swiss Federal Institute for Technology and can be reached at: http://www.marktplatz.uzh.ch/
trained accordingly. The dress for both SES roles was protocolled on a photograph to guarantee that they remained identical throughout the experiment. During the experiment, the test patient undertook weekly visits to our office to arrange dentist appointments and hand over the visit protocols to us. These visits were used to keep a check on the test patient’s dresses. The dress for the high SES role mimics a banker’s outfit and has been combined from a sales person in a classy department store in Zurich. At the time of the training sessions we also sent the test patient to five test visits. We did not use these visits for the statistical analysis in this study, but used them to make the test patient familiar with his roles and to test and improve the script. Incoming bills from all visits proof that all visits in the experiment and the medical check-up did indeed take place.

2.8.4 Covariates and model fit

2.8.4.1 Balance of covariates

<table>
<thead>
<tr>
<th>TABLE 2.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-values of pairwise comparisons of means.</td>
</tr>
<tr>
<td>**</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>INFO-LS</td>
</tr>
<tr>
<td>ST-HS</td>
</tr>
<tr>
<td>INFO-HS</td>
</tr>
<tr>
<td><strong>Test:</strong> Fisher’s exact (2-sided)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-values of pairwise comparisons of means.</td>
</tr>
<tr>
<td>**</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>INFO-LS</td>
</tr>
<tr>
<td>ST-HS</td>
</tr>
<tr>
<td>INFO-HS</td>
</tr>
<tr>
<td><strong>Test:</strong> Mann-Whitney (2-sided)</td>
</tr>
</tbody>
</table>
2.8.4.2 Model fit

**FIGURE 2.15**
Model fit.

(a) Comparison of the poisson and negative binomial model predicting the number of recommended fillings.

(b) Comparison of the model fit of OLS and GLM estimations predicting the cost estimate amount.
We analyse patients’ prevention decisions when patients cannot tell when they receive too much treatment or are charged too much, i.e., when health services have credence qualities. A monopolistic physician takes advantage of the situation by overtreating or overcharging patients. It is shown that credence qualities lead to inefficiencies with respect to prevention and visiting decisions, compared to a market without asymmetric information. Notably, patients with high health risks decide not to prevent, because they anticipate to be overtreated or overcharged. Our work adds a novel point of view to the discussion on insufficient prevention and the performance of prevention programmes. We conduct the analysis for institutional settings in which overtreatment is a problem (with verifiability) and in which overcharging is a problem (without verifiability, but with liability). Monopoly prices differ between institutional settings. Finally, price regulation and social health insurance are discussed.

3.1 Introduction

3.1.1 Motivation

Chronic diseases such as heart disease, stroke and type 2 diabetes are the cause of 70% of all deaths in the United States (Department of Health and Human Services (2009)). The WHO claims that 80% of all cases due to these diseases could be prevented by healthier behaviour (World Health Organization (2005)). In 2005, the World Health Organization estimated that preventable chronic diseases were going to cost China’s economy US$ 550 billion in the decade 2005-2014. Reports by international organizations emphasize that the benefits of prevention exceed its
costs. Actual prevention spending is well below the level that would be observed if prevention levels by professional organizations were followed (Kenkel, 2000).

One economic argument for why we observe too little prevention in health care markets is ex-ante moral hazard as a consequence of health insurance (Ehrlich and Becker, 1972). Insurance decreases patients’ health care expenditure and therefore decreases the benefits from prevention. While low prevention seems to be a widespread and persistent problem in health care markets, health insurance is not able to explain many observed prevention patterns. For instance, Baicker et al. (2015) quote a medical study (DiMatteo, 2004) which reports that even high risk patients who highly benefit from prevention prevent insufficiently.\(^1\) Baicker et al. suggest that behavioural hazards may explain health care misuse such as too little prevention. Similar approaches have recently appeared in the health policy literature (Loewenstein et al., 2013). Further explanations from the economic literature include externalities and information problems (Kenkel, 2000) and excessive discounting of future utility (Zweifel et al., 2009).

In this paper, we approach the phenomenon of too little prevention from a different angle: supply-side moral hazard in physician expert markets. We shift focus from the moral hazard problem on the patient side to a moral hazard problem on the supply-side. The physician has an informational advantage about what medical treatment the patients needs. The health service in our model has credence qualities (is a “credence good”) as patients cannot assess whether they have received an appropriate treatment even ex-post consumption (Darby and Karni, 1973)\(^2\). Health care markets are the prime example for credence goods markets. Industrialized countries typically spend more than 10% of their GDP for health care (OECD (2013)).

Credence qualities allow expert that certain types of misbehaviour remain undetected. Depending on the institutional framework and prevailing incentives, the following problems occur. First, physicians may have incentives to provide more treatments than necessary, referred to as overtreatment. Second, physicians may face incentives to charge for unperformed treatments, referred to as overcharging. A third problem is undertreatment, when physicians face incentives to provide less treatment than appropriate, which is not the focus of this paper\(^3\). To illustrate the

---

\(^1\)The example quoted by Baicker et al. (2015) involves diabetes patients. Compared to non-diabetes patients, they face a comparatively high risk for a number of diseases which can be prevented with low costs.

\(^2\)For a comprehensive treatment of the definition of credence goods see chapter 1 of this thesis.

\(^3\)The experience of undertreatment renders the service traded an experience good rather than a credence good if patients can detect that they have been undertreated. See chapter 1 of this thesis.
3.1. Introduction

magnitude of the problem, it is estimated for the USA that up to 10% of all health care expenditures, a sum of more than $200 billion, are due to fraud\textsuperscript{4} (FBI (2011)) such as overtreatment.

We study credence goods markets with patients who differ with respect to their health risk and can engage in preventive behaviour to reduce this risk. Several real world examples illustrate the relevance of these market features. The main example throughout this paper are markets for dental care in which both prevention and natural disposition determine the probability to develop diseases like caries. This example is discussed in greater detail in section 3.6, along with other examples such as cancer screenings and psychological diseases. In the fashion of the dentist example and many other examples from health care markets, we assume that patients either require treatment or no treatment at all. Thereby our study modifies the assumptions about treatment needs from the credence goods model of Dulleck and Kerschbamer (2006), where either a cheap or an expensive treatment, but never no treatment, is needed.

Although prevention has been broadly analysed in health and insurance economics, to our knowledge we are the first to consider prevention in the context of credence goods markets. We show that – compared to the case without asymmetric information – credence goods markets with prevention create inefficiencies with respect to both prevention and treatment. Supply-side moral hazard reduces prevention incentives for patients with relatively high risks, because overtreatment and overcharging reduce the benefits of preventive effort compared to a market without asymmetric information. On the other hand, supply-side moral hazard increases prevention incentives for patients with comparatively low health risks, because through supply-side moral hazard prevention becomes a substitute for physician visits. For the lowest risk patients, supply-side moral hazard makes treatment unattractive enough to make them stay out of the market completely, resulting in untreated health problems. Patients with relatively high risks visit the physician in every state of the world, i.e., too often. The coexistence of these inefficiencies leads to trade-offs in the implementation of policies.

Credence goods market in the real world show a great variety with respect to sectors and institutional characteristics. The institutional setting in credence goods markets has been shown to have a decisive influence on market outcomes (Dulleck and Kerschbamer, 2006; Dulleck et al., 2011), particularly, the two institutions

\textsuperscript{4}Fraud here includes billing for services not rendered, upcoding of services, duplicate claims, unbundling, excessive services or items, medically unnecessary services and kickbacks.
verifiability and liability. We analyse markets with prevention in institutional set-
tings with and without verifiability and liability to reflect the variety of real world credence goods markets. With verifiability, the prominent problem in the market is inefficient overtreatment. In markets without verifiability and with liability the relevant problem is overcharging which is redistributive, but not inefficient per se. This difference results in different monopoly prices and different policy implications.

3.1.2 Related literature

Credence Goods and Physician-Induced-Demand. Darby and Karni (1973) added goods with credence qualities to the goods-categorization into goods with search and experience qualities by Nelson (1970). Credence qualities of a good are those which cannot be evaluated even ex-post purchase, contrasting experience qualities (Nelson, 1970), which can be evaluated after purchase. Goods with important credence qualities, for instance medial advice, are usually termed credence goods.\(^5\) Important models which formalized the ideas by Darby and Karni are Pitchik and Schotter (1987) and Emons (1997). Dulleck and Kerschbamer (2006) provide a unifying framework for a particular setting, now referred to as “the standard credence goods problem” (Bester and Dahm, 2017). In their setting, consumers need one of two possible treatments (cheap or expensive) provided by an expert who can set prices for both treatments. The main result of Dulleck and Kerschbamer is that credence goods markets lead to efficient outcomes if at least on the two institutions verifiability and liability holds, because the expert chooses equal mark-up prices. With equal markup prices, both the appropriate and inappropriate treatment are equally profitable and experts are assumed to behave honestly, even in the presence of market power.

This result is in sharp contrast to empirical and anecdotal evidence, however. Excessive treatment seems to be a persistent problem in credence goods markets. Recently, a number of field experiments have shown that overtreatment and overcharging in different real word credence goods markets persists (see Schneider (2012) for car repairs, Balafoutas et al. (2013, 2015) for taxi rides, Kerschbamer et al. (2016) for computer repairs; Kerschbamer and Sutter (2017) provide a review). In health economics, the literature of physician-induced demand has recognized the incentives of physicians to influence demand at least since Evans (1974) (see McGuire (2000) for a review). The physician-induced demand hypothesis is supported by empirical indications. For instance, Domenighetti et al. (1993) find in a survey-based study

---

\(^5\)See chapter 1 of this thesis for a detailed discussion on the definition of credence goods.
from Switzerland that physicians, lawyers and their families receive the seven most important surgeries 33% less often than the rest of the population, indicating that liability concerns influence supply. Recently, audit studies in the health care sector support the physician-induced demand hypothesis directly (Currie et al. (2011, 2014) and chapter 2 of this thesis). The theoretical literature on credence goods has likewise provided results with overtreatment or overcharging, however, these results are usually sensitive to the assumption in place. Some studies have focussed on settings without verifiability and the corresponding problem of overcharging (Pitchik and Schotter, 1987; Wolinsky, 1993; Fong, 2005; Suelzle and Wambach, 2005; Dulleck and Kerschbamer, 2006), other studies have focussed on the problem of overtreatment in settings with verifiability (Emons, 1997; Dulleck and Kerschbamer, 2006; Alger and Salanie, 2006; Emons, 2013).

We depart from the assumption of Dulleck and Kerschbamer (2006) that there is a cheap and an expensive treatment alternative and instead assume that patients either need treatment or no treatment at all. As a consequence, overcharging remains profitable as long as prices are positive. Moreover, as long as prices reflect some degree of market power, also overtreatment is profitable. With the no-treatment alternative, endogenous price setting does not solve the credence goods problem in these cases. There are many examples of such situations in health care markets. For instance, in chapter 2 of this thesis we present the case of a caries patient who does not need treatment and visits a dentist for a check-up. Other examples include all kinds of situations in which patients face painless symptoms, but require a physician check-up in order to rule out that the symptoms’ are caused by a disease. We consider this a contribution to the modelling of the credence goods problem which could be used by other authors who study credence goods markets in contexts other than prevention.

Different institutional settings with respect to verifiability and liability are a recently discussed topic in the literature on credence goods. Although in Dulleck and Kerschbamer (2006), institutions do no have an effect on outcomes as long as either verifiability or liability holds, an experimental analysis by Dulleck et al. (2011) suggests that the institutional setting in credence goods markets can crucially influence market outcomes. Most of the theoretical credence goods literature has focused on one particular institutional setting (with the notable exceptions of Dulleck and Kerschbamer (2006), Dulleck et al. (2011) and Fong et al. (2014)) – and

---

6 Moreover, anecdotal evidence with respect to overtreatment in health care regularly receive wide attention (Gawande, 2009, 2015).

7 Dulleck et al. (2011) report that verifiability has a bigger impact on market outcomes than verifiability.
due to different model assumptions, the effects of different settings are not easily comparable. In this paper, we analyze the interaction of prevention and supply-side moral hazard in all institutional settings involving either verifiability or liability. We show that physicians set different prices in different institutional settings and that policies have different effects depending on the institutional setting.

The model accounts for heterogeneity among patients with respect to the probability that they need treatment (Dulleck and Kerschbamer, 2006; Dulleck et al., 2014). This heterogeneity may reflect heterogeneous natural dispositions among patients or differences in patients’ current health statuses and helps to see market outcomes in a more nuanced way. In terms of the prevention technology, we follow Zweifel et al. (2009): patients can use prevention to influence the probability distribution of their health state. The prevention technology implies larger benefits of prevention for patients with high probabilities of needing treatment.\(^8\)

**Prevention.** The connection between prevention and medical care is analyzed by Phelps (1978). He argues that prevention and medical care are substitutes in the sense that an increase in the price for care increases the demand for prevention. We show that the substitution between prevention and medical care is more pronounced with supply-side moral hazard than in the case without asymmetric information.

The seminal contribution relating health insurance to possibly averse prevention incentives is Ehrlich and Becker (1972). The authors show in a expected-utility framework with two states of the world that health insurance reduces the difference between the income when healthy and the income when sick and thus reduces the benefits of preventive activities\(^9\). This effect is counteracted when the costs of prevention activities are insured, but insured prevention activities are rather unusual in practice (Breyer et al., 2009). There are more reasons to doubt the importance of the problem. Among these is partial insurance (Harris and Raviv, 1978; Shavell, 1979) which can help to reduce the moral hazard problem, because incomplete coverage provides an incentive to engage in prevention. As noted by Cook and Graham (1977), the fact that health is an irreplaceable commodity might have a similar effect. A point made by Schlesinger and Venezian (1986) is that insurance companies have incentives to invest in preventive activities and can thereby reduce the moral hazard problem\(^10\).

\(^8\)Pitchik and Schotter (1993) consider heterogeneity in search costs, Fong (2005) considers patient heterogeneity with respect to the expected costs of treatment.

\(^9\)Other seminal contributions making this point are Pauly (1974) and Shavell (1979).

\(^10\)A practical example for this may be exercise trails in the city forest of Zurich, Switzerland, which are maintained by the local health insurance company Helsana.
Furthermore, the empirical evidence on ex-ante moral hazard due to insurance is not yet conclusive (Kenkel, 2000; Zweifel and Manning, 2000). This does not mean that the problem does not exist, however, as convincing empirical tests for ex-ante moral hazard due to insurance are hard to design. Recently, studies using exogenous variations in insurance coverage from US health care reforms have found positive (Dave and Kaestner, 2009) and neutral (Barbaresco et al., 2015) effects. A review of the more recent prevention literature in the context of insurance models is provided by Courbage et al. (2013). Other studies report mixed effects (Kelly and Markowitz, 2009). The topic of insurance of prevention activities has been analysed by Barigozzi (2004) and Ellis and Manning (2007).\textsuperscript{11}

Our contribution is to provide a supply-side explanation of prevention patterns as a consequence of asymmetric information in credence goods markets\textsuperscript{12}.

3.2 Model

3.2.1 Agents, actions and information

We consider an expert market with a monopolistic physician and risk-neutral patients. The assumption of a monopolistic physician does not necessarily refer to a single physician, but may represent a medical association that acts as a collective monopoly for its members (Zweifel et al., 2009). The physician only offers his services to the market if his profit is non-negative. We further stipulate that the physician does whatever is most profitable for him (see section 3.7 for a discussion of altruism). Patients are either in a good or a bad health state. Ex-ante, they differ with respect to the probability of being in the bad state. The probability is given by a patient’s type $h$, $h \in [0, 1]$. Correspondingly, the probability for being in the good state is $(1 - h)$. We assume a continuum of patients with mass one, distributed on the interval $[0, 1]$ according to the distribution function $F(h)$ with density $f(h) > 0$.

\textsuperscript{11}Ex-ante moral hazard, which is the concern of our paper, should be viewed separate from a larger literature following Arrow (1963) on ex-post moral hazard and insurance. This literature analyses how health care demand reacts to price changes created by health insurance (Pauly, 1968; Zeckhauser, 1970; Cutler and Zeckhauser, 2000). Contributions in this literature often refer to the general term moral hazard, but specifically analyse ex-post moral hazard. Finkelstein (2014) provides an overview on how the two strands of literature developed since Arrow (1963).

\textsuperscript{12}Schlesinger and Venezian (1986) also provide a theory based on supply-side considerations. However, they consider insurable prevention and focus on the relationship between insurer and consumer when the insurer can influence the loss probabilities of the consumers. There is no asymmetric information in their model.
Chapter 3

A patient in the bad state requires treatment in order to prevent a loss $L > 0$ that occurs if the problem remains untreated. A patient in the good state does not need treatment. Each patient has the choice to engage in a costly prevention activity, in order to reduce the probability $h$. The prevention decision is modelled as a binary choice between prevention and no prevention. Prevention requires an effort with costs $c > 0$, no prevention does not require any effort and is costless.

With prevention, the patient reduces the ex-ante probability for being in the bad health state by the factor $\alpha \in (0, 1)$, from $h$ to $\alpha h$. With this modelling, the absolute reduction of the probability through prevention is increasing in the type $h$. A real life analogy is that adopting a solid dental hygiene with the aim to reduce caries has a greater impact on patients with high ex-ante caries risks than for patients with low ex-ante caries risks\textsuperscript{13}.

There is asymmetric information as patients cannot observe the realization of their health state at the time of the treatment decision, but the physician is able to obtain this information through a (costless) diagnosis. In case a patient decides to visit a physician, the physician diagnoses the patient’s state without diagnostic error. We abstract from costs associated with visiting and diagnosis. The physician either carries out a treatment at cost $\kappa > 0$ or does not provide treatment. The price charged by the physician is denoted $t$, $t > 0$. The physician can generally charge this price although he did not treat the patient, i.e., overcharge.

When the patient receives the treatment she needs, we say she is treated appropriately. The appropriate treatment when the patient requires treatment is to treat her. The appropriate treatment when the patient does not require treatment is not to treat, nor to charge her. Inappropriate treatment can occur as overtreatment or overcharging, respectively.\textsuperscript{14} When a patient gets treated, the utility excluding treatment costs is always 0, regardless whether she requires treatment or not. Overtreatment occurs when a good state patient receives treatment, is charged accordingly and ends up with a utility of $-t$. Overcharging either occurs when a good state patient is not treated but charged and ends up with a utility of $-t$, or when a bad state patient is not treated, but charged and left with a utility of $(L - t)$. The corresponding possible patient utilities are illustrated in table 3.1.

\textsuperscript{13} We could also, without different results, let patients differ in the costs of prevention instead of the benefits. See Hofmann (2007) for an application. This would suggest a socio-economic interpretation of the heterogeneity in our model. People with poor socio-economic characteristics have to work more hours in order to afford the services or goods that achieve a certain prevention outcome than people with higher incomes.

\textsuperscript{14} A third problem is undertreatment which occurs when a bad state patient does not receive treatment and consequently ends up with a utility of $-L$. In this case the medical service has experience instead of credence qualities (see chapter 1 of this thesis).
### TABLE 3.1
Patient utility depending on health state, treatment and charging.

<table>
<thead>
<tr>
<th>Patient state</th>
<th>Not treated, not charged</th>
<th>Treated, &amp; charged</th>
<th>Not treated, &amp; charged</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good</strong></td>
<td>0</td>
<td>0 − t</td>
<td>0 − t</td>
</tr>
<tr>
<td></td>
<td>(appropriate)</td>
<td>(overtreated)</td>
<td>(overcharged)</td>
</tr>
<tr>
<td><strong>Bad</strong></td>
<td>−L</td>
<td>0 − t</td>
<td>−L − t</td>
</tr>
<tr>
<td></td>
<td>(undertreated)</td>
<td>(appropriate)</td>
<td>(overcharged)</td>
</tr>
</tbody>
</table>

Table 3.1 illustrates the credence goods qualities of medical service. First, ex-post, if a patient knows that she has been treated and charged, she cannot infer from her utility whether she was a good or a bad state patient. The utility is −t in both cases. Second, if the patient cannot observe whether she has been treated or not, she cannot know whether she is a treated bad state patient or an overcharged good state patient. Again, the utility is −t in both cases.

#### 3.2.2 Timing and solution concept

The timing of the game is as follows.

1. The physician posts a price for the treatment;
2. The patients observe the price and choose preventive effort;
3. Nature determines each patient’s state (good or bad);
4. Each patient decides whether to visit the physician or not;
5. The physician secretly learns the states of the visiting patients (diagnosis), treats patients or not, and charges the according price.

With the end of the last stage, utilities are realized. The solution concept for the multi-stage game is subgame-perfect equilibrium. We introduce three tie-breaking rules to ease the exposition and make sure that we obtain unique equilibria in pure strategies. First, when a patient is indifferent between visiting and not visiting the physician, she chooses to visit. Second, when a patient is indifferent between preventing and not preventing, she decides to prevent. Third, similar to Dulleck and Kerschbamer (2006), we assume that physicians choose the appropriate

---

This case is probably less common. Examples for such cases are non-verifiable treatments. Such treatments arguably occur more often as parts of more elaborate treatments than on their own. An extreme example are medical treatments performed under general anaesthesia or the use of generics under original-labels.
treatment in the case of indifference and that this is common knowledge. Without this assumption, the fact that the treatment of each patient has no influence on the physician’s profit (we have with a continuum of patients), would be a problem for determining an equilibrium.

Notably, we do not permit patients to reject a treatment recommendation, i.e., patients are committed to the physician’s decision (see the commitment case in Dulleck and Kerschbamer (2006)). This assumption is without loss of generality. The diagnosis is unobserved by the patients and therefore does not reveal any information to them. In equilibrium there will be no reason why a patient should visit the physician without the intention of being treated.

3.3 Benchmark I: no market for treatment \((nM)\)

As a first benchmark, we assume that no treatment possibility exists and denote this regime \((nM)\). The loss \(L\) occurs to a patient with type \(h\) with probability \(h\) if she does not prevent and with probability \(\alpha L\) if she prevents at costs \(c\). A patient takes a preventive effort if his expected utility with prevention, \(EU_p\), exceeds the expected utility without prevention, \(EU_{np}\), i.e., if

\[
EU_p \geq EU_{np}
\]

\[
\iff -\alpha hL - c \geq -hL
\]

\[
\iff (1 - \alpha)hL \geq c
\]

\[
(3.1)
\]

\[
\iff h \geq \frac{c}{(1 - \alpha)L} := \hat{h}
\]

\[
(3.2)
\]

Inequality (3.1) contrasts the costs of prevention on the right hand side with the benefits of prevention on the left hand side. The right-hand side of (3.2) shows the marginal type who is indifferent between prevention and no prevention, denoted by the cutoff \(\hat{h}\). The cutoff positively depends on the costs of prevention, \(c\), and negatively on the efficiency of prevention, \((1 - \alpha)\) and the expected loss \(L\). In order to obtain an interior marginal type, we assume that prevention is efficient for a positive share of patients, i.e., we assume

\[\text{We use the tie-breaking rule from section 3.2.2.}\]
3.4. Benchmark II: no asymmetric information (nA)

Assumption (A1)
\[
\frac{c}{(1-\alpha)L} \in (0, 1)
\]

The benefits of prevention are increasing in the patient type while the costs are equal for all patients. Thus, only types with a sufficiently high type decide to prevent. This result is illustrated in figure 3.1.

**FIGURE 3.1**
Benchmark without market.

Note: Patients’ expected utility depending on type and evolving patient groups.

The share of the population that engages in prevention is denoted \( P_{nM} \) and given by the share of patients above the cutoff \( \hat{h} \).

\[
P_{nM} = \int_{\hat{h}}^{1} f(h)dh = 1 - F(\hat{h}) = 1 - F\left(\frac{c}{(1-\alpha)L}\right).
\]

We summarize the results of this section:

**Result 3.1.** In the absence of a treatment technology, the population is divided into two groups. Types below \( \hat{h} := \frac{c}{(1-\alpha)L} \) do not prevent, types above \( \hat{h} \) prevent.

3.4 Benchmark II: no asymmetric information (nA)

Now we consider a setting with treatment technology, but without asymmetric information, denoted (nA). We assume that treatment is contractible, i.e., there is no scope that physicians could overcharge patients.

The game from the preceding section is changed only at stage 3:

\textsuperscript{17}This assumption is similar to the assumption in Breyer et al. (2009) (p.229). The difference is that we have a model with a continuum of patients while Breyer et al. only consider one patient.
3. Nature determines each patient’s state (good or bad). Patients learn their state.

Patients in the good state learn that they are not in need of treatment and therefore do not visit the physician. Patients in the bad state know that the loss \( L \) will occur with certainty. They demand treatment from the monopolistic physician as long as the treatment price is not larger than \( L \). The physician can treat patients profitably, because the costs of providing treatment are below \( L, \kappa < L \). Thus, at prices \( t \leq L \), all patients in the bad state decide to get treated, while no patient seeks treatment if \( t > L \). The expected utility of a patient with type \( h \) who does not prevent (if \( t \leq L \)) is given by

\[
EU_{np}(t) = -ht. \tag{3.3}
\]

The expected utility of a patient with type \( h \) who prevents (if \( t \leq L \)) is given by

\[
EU_p(t) = -\alpha ht - c. \tag{3.4}
\]

Patients benefit from prevention by saving the treatment costs \( t \) with the probability \((1 - \alpha)h\) while the costs of prevention are \( c \). A patient decides to prevent if

\[
EU_{np}(t) \leq EU_p(t)
\]

\[
\iff -ht \leq -\alpha ht - c
\]

\[
\iff h \geq \frac{c}{(1 - \alpha)t}. \tag{3.5}
\]

The marginal type is denoted \( \hat{h}(t) \):

\[
\hat{h}(t) := \frac{c}{(1 - \alpha)t}. \tag{3.6}
\]

As in the case without market, the population splits into two groups. High risk types prevent, low risk types decide not to prevent. The mass of patients who demand treatment (bad-state patients) depends on the price, because the price influences the prevention decision. Patients who do not prevent, patient types below \( \hat{h}(t) \), visit the physician with probability \( h \). Patients who prevent, types above \( \hat{h}(t) \), visit the physician with probability \( \alpha h \). Demand behaviour is illustrated in figure 3.2. For types with \( h > \hat{h}(t) \) the expected benefit of prevention, \((1 - \alpha)L\), is larger than the
3.4. Benchmark II: no asymmetric information (nA)

**FIGURE 3.2**
Market without asymmetric information.

Note: Market without Asymmetric Information. Patients’ expected utility depending on type and evolving patient groups.

cost of prevention, \(c\). For types with \( h < \hat{h}(t) \) the opposite is true. There is a price below which no patient prevents, because treatment at this price is so cheap that prevention does not even pay off for the types with the highest prevention benefits. This price is denoted \( t_0 \) and obtained by equating \( \hat{h}(t) = 1 \), given by

\[
t_0 = \frac{c}{1 - \alpha}.
\]

(3.7)

We further assume:

**Assumption (A2)** \( \frac{c}{(1-\alpha)\kappa} < 1 \)

Assumption (A2) is a stronger version of assumption (A1). With (A2) we have \( t_0 < \kappa \) and a situation with no prevention will not occur in equilibrium. We denote the share of the population engaging in prevention by \( P_{nA}(t) \), given by

\[
P_{nA}(t) = \begin{cases} 
1 - F(\hat{h}) & \text{if } t < \kappa \\
\int_{\hat{h}(t)}^{1} f(h)dh = 1 - F(\hat{h}(t)) & \text{if } t \in [\kappa, L] \\
1 - F(\hat{h}) & \text{if } t > L.
\end{cases}
\]

(3.8)

At prices below \( \kappa \) there is no supply and the market outcome is the same as in the case without market. The same situation occurs for prices above \( L \) for which there is no demand. For prices \( t \in [\kappa, L] \), high risk types above the cutoff \( \hat{h}(t) \) prevent. In this price domain, prevention is more attractive the higher the treatment price, as a higher price yields higher expected savings through prevention. Therefore, the share of preventing patients is increasing in the price, as can be shown with

\[
\frac{\partial P_{nA}(t)}{\partial t} = f(\cdot) \frac{c}{(1-\alpha)t^2} > 0.
\]

This substitution effect between prevention and treatment.
Chapter 3

resembles the one described by Phelps (1978). As the price increases, types who change their prevention decision substitute treatment with prevention in the sense that they reduce the probability that they need (and receive) treatment by increasing their prevention effort. The share of visiting patients is given by

\[
V_{nA}(t) = \begin{cases} 
0 & \text{if } t < \kappa \\
\int_0^{\tilde{h}(t)} h f(h) dh + \alpha \int_{\tilde{h}(t)}^1 h f(h) dh & \text{if } t \in [\kappa, L] \\
0 & \text{if } t > L.
\end{cases}
\] (3.9)

At prices below \( \kappa \), there is no supply. At prices between \( \kappa \) and \( L \), the share of visiting patients is composed of the mass of not preventing patients (types below the cutoff \( \tilde{h}(t) \)) and the mass of preventing patients (types above the cutoff). The profit of the physician is given by multiplying physician demand and the profit per treatment \( t - \kappa \), i.e.,

\[
\pi_{nA}(t) = \begin{cases} 
0 & \text{if } t < \kappa \\
(t - \kappa) \left( \int_0^{\tilde{h}(t)} h f(h) dh + \alpha \int_{\tilde{h}(t)}^1 h f(h) dh \right) & \text{if } t \in [\kappa, L] \\
0 & \text{if } t > L.
\end{cases}
\] (3.10)

The objective of the monopolistic physician is to maximize \( \pi_{nA}(t) \) with respect to the price \( t \). We do not require a unique solution, but any profit maximizing price \( t^*_{nA} \) lies in the domain \( (\kappa, L] \). We summarize these results:

**Result 3.2.** Consider a monopolistic market without asymmetric information. In equilibrium, patients in the bad state visit the physician and receive appropriate treatment, while patients in the good state do not visit the physician. The physician chooses the price \( t^*_{nA} \in (\kappa, L] \). The population is divided into two groups. Types below \( \tilde{h} = \frac{c}{(1-\alpha)t^*_{nA}} \) prevent, types above \( \tilde{h} \) do not prevent. The proportion of preventing patients is given by \( 1 - F \left( \frac{c}{(1-\alpha)t^*_{nA}} \right) \). All inefficiencies are due to monopolistic pricing.

The monopolistic market outcome is not efficient, because the equilibrium price is too high. Market outcomes are efficient when the price \( t \) equals the marginal costs \( \kappa \). Intuitively, at the efficient prevention level, the expected costs of the marginal patient not preventing, \( \kappa \tilde{h}(t) \), equal the expected costs of the marginal patient preventing, \( -(\alpha \kappa \tilde{h}(t) + c) \). We obtain

\[
-\kappa \tilde{h}(t) = -(\alpha \kappa \tilde{h}(t) + c)
\]
3.5 The credence goods market

\[
\hat{h}(t) \iff \hat{h}(t) = \frac{c}{(1 - \alpha)\kappa} = \hat{h}(t = \kappa).
\]

At prices above \(\kappa\), some patients prevent although the societal costs are lower if they would not prevent. Social welfare is maximized when prices equal marginal treatment costs, because any profits of the physician solely constitute welfare-neutral redistributions from the patients to the physician. Because \(t^*_{nA} > \kappa\), we observe excessive prevention.

3.5 The credence goods market

In the credence goods market, patients do not know their state and the diagnosis only reveals the state to the physician.\(^{18}\) Patients hence have to make their decision whether to visit the physician or not independent of their state and they have no reason not to follow the physician’s treatment recommendation once they have decided to visit, because the diagnosis is not informative. A patient is not able to infer her state from the realized utility even after she has visited the physician in case she gets overtreated or overcharged in the good state (see table 3.1).\(^{19}\) The physician may use this informational advantage to defraud the consumer by overtreating or overcharging, depending on the institutional setting.

3.5.1 Institutional settings and supply-side moral hazard

As Dulleck and Kerschbamer (2006), we consider credence goods markets with four different institutional settings with respect to the existence of the institutions liability (L) and verifiability (V).

\textit{Liability, }L, \textit{ describes institutional features that hold the physician liable for not appropriately treating a patient in need. Liability prohibits undertreatment, but overcharging and overtreatment are still possible}\(^{20}\).

\(^{18}\)In practice, situations like this usually occur when the physician is better trained than the patient with respect to finding and interpreting symptoms or with respect to evaluating results displayed by medical equipment. They can also occur due to procedural practices as illustrated by an example from dentistry: In some cases, after having taken an x-ray from the patient, the dentist leaves the patient in order to analyse the x-ray in private in another room and returns with a diagnosis.

\(^{19}\)To continue the dentist example, imagine the dentist places a filling. The patient may not be able to know whether she needed it (bad state) or not (good state) even after the treatment.

\(^{20}\)The physician’s decision to overcharge always dominates the decision to overtreat for a given price when both are possible, as \(\kappa > 0\).
Verifiability, \( V \), describes the ability of patients to verify the received treatment. Verifiability guarantees that the patient can detect overcharging. Overtreatment, however, is still possible.

First, there is a setting in which neither liability nor verifiability holds, denoted \((\text{noL/noV})\). In this setting, for any price above zero, the physician has an incentive to overcharge every visiting patient, and not to treat all patients in a bad state (undertreatment). We will not further consider this setting in this paper, because a market ceases to exist for any price above zero.\(^{21}\) The market outcome is the same as in the first benchmark (section 3.3). This result is in line with Dulleck and Kerschbamer (2006).

Second, there is a setting in which verifiability holds, but liability does not, denoted \((\text{noL/V})\). Verifiability prevents overcharging. Third, there is a setting in which both verifiability and liability hold, denoted \((L/V)\). We summarize both settings with verifiability, \((\text{noL/V})\) and \((L/V)\), under the notation \((V)\), because they lead to identical results. With verifiability, overcharging is not possible. In both \((\text{noL/V})\) and \((L/V)\), the physician has an incentive to overtreat good-state patients whenever \(t > \kappa\). In \((\text{noL/V})\), the physician is inclined to overtreat rather than undertreat patients as long as prices are above treatment costs.

Last, there is a setting in which liability holds, but verifiability does not, denoted \((L/\text{noV})\). In this setting, the physician has an incentive to overcharge good-state patients when the treatment price is above zero. Undertreatment is not possible because of liability. Tables 3.2 and 3.3 summarize the payoffs in the different institutional settings for the physician and the patients, respectively, assuming supply-side moral hazard inducing prices \(t > 0\) in \((\text{noL/noV})\), \(t > \kappa\) in \((V)\) and \(t > t_\emptyset\) in \((L/\text{noV})\).\(^{22}\)

### 3.5.1.1 Demand and supply in \((V)\) and \((L/\text{noV})\)

At any given price, patient payoffs in \((V)\) equal those in \((L/\text{noV})\) (see table payoffpatient). We can hence analyse patient behaviour for both settings together. Each patient has the choice between the same four strategies: Not visiting and not preventing \((nv,np)\), not visiting and preventing \((nv,p)\), visiting, but not preventing \((v,np)\) and visiting and preventing \((v,p)\). Any patient always chooses \((v,np)\) over

---

\(^{21}\)It is a dominant strategy for all patients to reject any treatment recommendation, independent of their type.

\(^{22}\)We denote by \(t_\emptyset\) the price at which the physician’s profit is just zero in \((L/\text{noV})\).\(^{23}\) Since costs of overcharging are zero, \(t_\emptyset\) is lower than the treatment costs \(\kappa\). At prices below \(t_\emptyset\), the physician does not offer his services in the market.
3.5. The credence goods market

**TABLE 3.2**

Physician payoffs & behaviour
With supply-side moral hazard inducing prices

<table>
<thead>
<tr>
<th>Patient State</th>
<th>Setting</th>
<th>(nA)</th>
<th>(noL/noV)</th>
<th>(V)</th>
<th>(L/noV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Good</td>
<td>0</td>
<td>t</td>
<td>t - κ</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Appropriate</td>
<td>Overcharging</td>
<td>Under treatment</td>
<td>Over treatment</td>
</tr>
<tr>
<td>Bad</td>
<td>t - κ</td>
<td>t</td>
<td>t - κ</td>
<td>t - κ</td>
<td>t - κ</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Appropriate</td>
<td>Overcharging &amp;</td>
<td>Under treatment</td>
<td>Appropriate Treatment</td>
</tr>
</tbody>
</table>

Note: physician payoffs per visiting patient and behaviour in the benchmark without asymmetric information (nA) and the different institutional settings with respect to verifiability (V) and liability (L), for supply-side moral hazard inducing prices \( t > 0 \) in (noL/noV), \( t > \kappa \) in (V) and \( t > t_0 \) in (L/noV).

**TABLE 3.3**

Patient payoffs
With supply-side moral hazard inducing prices

<table>
<thead>
<tr>
<th>Patient State</th>
<th>Setting</th>
<th>(nA)</th>
<th>(noL/noV)</th>
<th>(V)</th>
<th>(L/noV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>0</td>
<td>-t</td>
<td>-t</td>
<td>-t</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Appropriate</td>
<td>Overcharging</td>
<td>Over treatment</td>
<td>Over charging</td>
</tr>
<tr>
<td>Bad</td>
<td>-t</td>
<td>-L - t</td>
<td>-t</td>
<td>-t</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Appropriate</td>
<td>Overcharging &amp;</td>
<td>Under treatment</td>
<td>Appropriate Treatment</td>
</tr>
</tbody>
</table>

Note: patient payoffs in case of a physician visit in the benchmark without asymmetric information (nA) and the different institutional settings with respect to verifiability (V) and liability (L), for supply-side moral hazard inducing prices \( t > 0 \) in (noL/noV), \( t > \kappa \) in (V) and \( t > t_0 \) in (L/noV).

\((v, p)\), because she obtain the same payoff \(-t\) given she visits the physician, regardless of her state. Prevention is of no benefit in this case, but comes with costs \( c > 0 \). Thus, \((v, p)\) will not be considered in the following. Assuming supply-side moral hazard inducing prices, a patient of type \( h \) obtains the following expected utility from the three remaining strategies.

\[
EU_{nv,np}^{CG} = -hL \quad EU_{n, p}^{CG} = -\alpha L h - c \quad EU_{v, np}^{CG} = -t. \tag{3.11}
\]

An important difference to the case without asymmetric information concerns the willingness to pay for visiting the physician.\(^{24}\) Without asymmetric information,

\(^{24}\)Intentionally, we do not say “willingness to pay for treatment”, because with overtreatment or overcharging this is not the same as the willingness to pay for visiting the physician.
all bad-state patients are willing to pay $L$ for medical treatment, all good-state patients are not willing to pay a positive amount. In the credence goods market, patients’ willingness to pay for visiting the physician depends on their type, but not on their state. Supply-side moral hazard leads to a type-independent payoff of $-t$ in case of a physician visit, while the payoff from the alternative strategies is decreasing in the type. The willingness to pay is just as large as the payoff from the best alternative strategy and therefore is increasing in the type. Patients with the highest type, $h = 1$, have the highest willingness to pay which is equal to $\alpha L + c$. This price is lower than the maximum price patients are willing to pay in the market without asymmetric information (from assumption (A1) we get $\alpha L + c < L$). At the price $\alpha L + c$, the highest type with $h = 1$ is just indifferent between the strategies $(v, np)$ and $(nv, p)$. Hence, she weighs her expected payoff from strategy $(v, np)$ without prevention and her expected payoff from strategy $(nv, p)$ with prevention. This is in contrast to the case without asymmetric information, where the decision to visit the physician or not is already made conditional on having realized a bad state state associated with a sure loss $L$, while in the credence goods market, the probability to develop a bad state is decisive for the strategy decision.

Figure 3.3 summarizes the basic of demand and supply behaviour for different price domains in the market without asymmetric information and the credence goods markets with different institutional settings.

![Figure 3.3 Supply and demand](image)

Note: Price domains with positive supply and demand and no demand or supply in the market without asymmetric information and the credence goods markets with different institutional settings.
3.5.2 Patient behaviour with supply-side moral hazard

We first will analyse patient behaviour in the credence goods market in detail and will use the results for the analysis of the physician’s price setting behaviour in the next section. We assume that prices are fixed at some level at which they induce supply-side moral hazard and positive demand, i.e., \( t \in (\kappa, \alpha L + c) \) in \( (V) \) and \( t \in (t_0, \alpha L + c) \) in \( (L/noV) \).

The cutoff between the two strategies without visiting \((nv, np)\), and \((nv, p)\) is the same as in the first benchmark without market:

\[
\hat{h} = \frac{c}{(1 - \alpha)L}.
\]

(3.12)

A patient of type \( h \) chooses strategy \((v, np)\) over strategy \((nv, np)\) if

\[
EU^{CG}_{v,np} \geq EU^{CG}_{nv,np} \iff h \geq \frac{t}{L}.
\]

(3.13)

The cutoff is denoted

\[
\hat{h}(t) := \frac{t}{L}.
\]

(3.14)

A patient of type \( h \) chooses strategy \((v, np)\) over strategy \((nv, p)\) if

\[
EU^{CG}_{v,np} \geq EU^{CG}_{nv,p} \iff h \geq \frac{t - c}{\alpha L}.
\]

(3.15)

The cutoff is denoted

\[
\hat{h}(t) := \frac{t - c}{\alpha L}.
\]

(3.16)

Comparing the two cutoffs leads to

\[
\hat{h}(t) = \frac{t}{L} < \frac{t - c}{\alpha L} = \hat{h}(t) \iff \frac{c}{(1 - \alpha)} < t,
\]

(3.17)

\[\text{We make use of the tie-breaking rules defined in section 3.2.}\]
where $\frac{c}{(1-\alpha)} = t_0$. From assumption (A2), $\frac{c}{(1-\alpha)\kappa} \in (0,1)$ we have $t_0 < \kappa$. Condition (3.17) holds for all prices $t > t_0$ and implies $\hat{h} < \hat{\hat{h}}(t)$. Hence, for $t \in [t_0, \alpha L + c)$ there exists a cutoff $\hat{h}(t)$ between the strategies $(nv, p)$ and $(v, np)$ given by

$$\hat{h}(t) = \frac{t - c}{\alpha L}.$$  

(3.18)

Throughout this paper we assume that the price $t_0$ is lower than the price at which the physician makes zero profits in $(L/noV)$, $t_\emptyset$, i.e., $t_0 < t_\emptyset$.\textsuperscript{26} This assumption guarantees that equilibria with positive demand and supply with zero prevention do not occur.\textsuperscript{27} When (3.17) holds, the patient population splits into three groups in both $(V)$ and $(L/noV)$ with the cutoff $\hat{h}$ between the patient groups who choose the strategies $(nv, np)$ and $(nv, p)$ and the cutoff $\hat{\hat{h}}(t)$ between the patient groups who choose $(nv, p)$ and $(v, np)$. Thus, for all supply-side moral hazard inducing prices with positive supply and demand we observe that the patient population divides into three groups. Figure 3.4 illustrates these results.

**FIGURE 3.4**
The Credence Goods Market.

Note: Credence goods market with three patient groups, assuming $t_0 < t_\emptyset$ and supply-side moral hazard inducing prices. The lines show patients’ expected utility and evolving patient groups depending on type.

Intuitively, high types tend to visit the physician in order to avoid the likely loss $L$. Because of the prospect of being overtreated or overcharged, respectively, there is no incentive to prevent for high types, however, although they would prevent in

\textsuperscript{26}There is no simple formulation for this inequality in terms of the model’s primitives.

\textsuperscript{27}Specifically it implies that condition (3.17) holds in the relevant price range with positive supply and demand in $(V)$, i.e., $\hat{h} < \hat{\hat{h}}(t)$ for $t \in (\kappa, \alpha L + c)$ in $(V)$ and for $t \in [t_\emptyset, \alpha L + c]$ in $(L/noV)$. 

78
the absence of asymmetric information. Types below $\hat{h}(t)$ choose not to visit at all, as visiting is too expensive for them, given that visiting only helps them to avoid the loss $L$ with a comparatively low probability. They would have to pay the treatment costs $t$ for sure when they decide to visit and hence they decide not to visit. Types between $\hat{h}$ and $\hat{h}(t)$ decide to prevent as the benefits of prevention (the saved loss $L$ with additional probability $(1 - \alpha)$) are larger to them than its costs. Types below $\hat{h}$ neither visit nor prevent, because they experience the loss $L$ with only a small probability, such that preventing is more expensive than beneficial.

3.5.2.1 The Inefficiencies of the Credence Goods Market

We can now make statements with respect to the occurring inefficiencies. The only inefficiency in the market without asymmetric information is excessive prevention of types between $\hat{h}(t)$ and $\hat{h}(\kappa)$ due to monopoly prices. This is illustrated in the upper part of figure 3.5. There, patients with types between the cutoffs $\hat{h}(t)$ and $\hat{h}(\kappa)$, partly substitute visiting with prevention, although prevention is not efficient for them. In the credence goods market, three additional inefficiencies occur as illustrated in the lower part of figure 3.5. The figure contains two slightly differing illustrations depending on whether $t \geq \frac{\alpha L c}{(1 - \alpha) \kappa} + c$. In addition to not efficient prevention, we observe insufficient prevention of patient types between $\hat{h}(t)$ and 1 or between $\hat{h}(\kappa)$ and 1, respectively. The reason is that supply-side moral hazard destroys prevention incentives for patients who decide to visit the physician. Further, patients who do not visit the physician due to the high related costs of overtreatment and overcharging, respectively, suffer from untreated health problems in case they realize a bad state. Last, we observe excessive medical treatment in the institutional settings with verifiability $(V)$, because with supply-side moral hazard all visiting patients are overtreated. Importantly, this inefficiency does not exist in the institutional settings with overcharging, because overcharging is not inefficient per se.

The results from this section are summarized in proposition 3.1,28 where we denote $\frac{\alpha L c}{(1 - \alpha) \kappa} + c$ by $t_x$:

**Proposition 3.1.** Consider the credence goods markets $(V)$ and $(L/noV)$ and assume that prices are fixed at a level at which supply-side moral hazard incentives prevail and demand is positive, i.e., $t \in (\kappa, \alpha L + c)$ in $(V)$ and $t \in (t_0, \alpha L + c)$ in $(L/noV)$. Several inefficiencies occur: patient types above $\hat{h}(t)$ (if $t > t_x$) or $\hat{h}(\kappa)$

28For a presentation of the demand function for patient and physician demand we refer to appendix 3.9.1.
FIGURE 3.5
Inefficiencies with fixed prices.

**The Credence Goods Market**

<table>
<thead>
<tr>
<th>No Asymmetric Information</th>
<th>Inefficient prevention</th>
<th>Inefficient treatment (only in (V))</th>
<th>No prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \hat{h} )</td>
<td>( \hat{h}(t) )</td>
<td>( \hat{h}(\kappa) )</td>
</tr>
<tr>
<td>for ( t \in (\kappa, L] )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Untreated health problems</th>
<th>Inefficient prevention</th>
<th>Inefficient treatment (only in (V))</th>
<th>No prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \hat{h} )</td>
<td>( \hat{h}(t) )</td>
<td>( \hat{h}(\kappa) )</td>
</tr>
<tr>
<td>for ( t \in \left( \frac{aLc}{(1-\alpha)\kappa} + c, \alpha L + c \right) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Inefficient prevention**

No Asymmetric Information

- \( \hat{h} \)
- \( \hat{h}(t) \)
- \( \hat{h}(\kappa) \)

for \( t \in (\kappa, L] \)

- Inefficient prevention
- Untreated health problems
- Inefficient treatment (only in (V))
- No prevention

(if \( t < t_x \)) do not prevent; types between \( \hat{h}(t) \) and \( \hat{h}(\kappa) \) (if \( t > t_x \)) or \( \hat{h}(t) \) (if \( t < t_x \)) prevent; types below \( \hat{h}(t) \) remain untreated in the bad state; types above \( \hat{h}(t) \) receive treatment in the good state (only in (V)).

The substitution effects between prevention and treatment are similar to the effect described by Phelps (1978) in the sense that the price for treatment and prevention are positively correlated (see section 3.4). In the credence goods market with supply-side moral hazard inducing prices, the effect is amplified compared to the case without asymmetric information as the substitution effect leads in a
complete stoppage of physician visits, because patients either prevent or visit, but never visit and prevent at the same time.\textsuperscript{29}

### 3.5.3 The physician’s price choice

In this section, we analyze which price the monopolistic physicians chooses, depending on the institutional setting. As mentioned earlier, the assumption of a monopolistic physician does not necessarily reflect single physicians with market power. Instead, we can also think of a medical association which acts as a monopoly for its members (Zweifel et al., 2009).

#### 3.5.3.1 With verifiability ($V$)

With verifiability, the physician overtreats good-state patients when $t > \kappa$, and appropriately treats bad-state patients. Using (3.36) the physician’s profit is given by:

$$
\pi_V(t) = \begin{cases} 
0 & \text{if } t \in (0, \kappa] \\
(1 - F(\hat{h}(t))) \cdot (t - \kappa) & \text{if } t \in (\kappa, \alpha L + c) \\
0 & \text{if } t \geq L \alpha + c,
\end{cases}
$$

(3.19)

where $\hat{h}(t) = \frac{t - c}{\alpha L}$. At prices between the two extremes, profit is always positive, because demand is positive and the price is above costs. In this range, the physician faces a trade-off between a higher profit per visiting patient and lower demand. At $t = \kappa$ there is no profit and no moral-hazard problem, because the physician does not profit from overtreatment and overcharging, respectively. The maximization problem of the physician is given by

$$
\max_t \pi_V(t).
$$

(3.20)

The first order condition of the maximization problem (3.20) for prices $t \in (\kappa, \alpha L + c)$ is given by

$$
\left(1 - F(\hat{h}(t_V^*))\right) = f(\hat{h}(t_V^*)) \hat{h}'(t_V^*)(t_V^* - \kappa).
$$

(3.21)
Chapter 3

The left-hand side represents the marginal profit of a price increase per visiting patient. The right-hand side represents the marginal costs of a price increase, a loss in demand. At the optimal price, denoted $t^*_V$, marginal benefits equal marginal costs.

The maximization problem of the physician yields a unique (interior) solution if the marginal costs of a price increase are increasing in the price, i.e.,

$$\partial \left[ f\left(\hat{h}(t)\right) \hat{h}'(t)(t - \kappa)\right] /\partial t > 0.$$ 

This condition requires that $f(\cdot)$ does not decrease too sharply. It guarantees single-crossing of the marginal profit and the marginal cost function.

**Lemma 3.1.** The maximization problem (3.20) has a unique interior solution, i.e., $t^*_V \in (\kappa, \alpha L + c)$, if

$$f'(\hat{h}(t)) \geq -\frac{\alpha L t - \kappa}{t - \kappa} \forall t \in (\kappa, \alpha L + c)$$ 

with strict inequality for some $t$.

Proof: see appendix 3.9.2.1.

### 3.5.3.2 With liability, without verifiability ($L/noV$)

In ($L/noV$), the physician overcharges patients in the good state. Liability guarantees that bad state patients are treated. As in the institutional settings with verifiability, ($V$), the physician earns $(t - \kappa)$ by treating patients in the bad state. His payoff from overcharging patients in the good state is $t$, however, because overcharging is costless.

In ($L/noV$) price increases are more costly for the physician than in ($V$) at a given price, because the physician overcharges good state patients costlessly in ($L/noV$) while he overtreats them with costs $\kappa$ in ($V$). The physician therefore earns a higher profit per patient in ($L/noV$). The costs of a price increase also differs in another dimension. In ($L/noV$), the lower the type of a patient, the more profitable is to treat her for the physician. The marginal profit from overcharging good-state patients is larger than the marginal profit from appropriately treating bad-state patients, and the lower the type, the higher is the probability that the patient will realize a good state. Hence, when increasing the price the physician looses the his most profitable patients. This contrasts ($V$), where all patients yield the same profit.

The mass of visiting patients who are in a bad state is

$$\int_{\hat{h}(t)}^{1} h f(h) dh.$$

(3.22)
3.5. The credence goods market

The mass of visiting patients who are in a good state is

\[ \int_{\hat{h}(t)}^{1} (1 - h) f(h) dh. \]  \hspace{1cm} (3.23)

Both (3.22) and (3.23) depend on the price, because the price determines the cut-off \( \hat{h}(t) \). The higher the price, the larger is the mass of patients in the bad state and vice versa.

The profit function is given by

\[ \pi_{(L/noV)}(t) = \begin{cases} 
0 & \text{if } t \in (0, t_\emptyset), \\
 t \int_{\hat{h}(t)}^{1} f(h) dh - \kappa \int_{\hat{h}(t)}^{1} h f(h) dh & \text{if } t \in [t_\emptyset, \alpha L + c], \\
0 & \text{if } t > \alpha L + c, 
\end{cases} \] \hspace{1cm} (3.24)

In the price range \([t_\emptyset, \alpha L + c]\), the physician receives \( t \) from every visiting patients, but only bears the costs \( \kappa \) from patients in the bad state, because these patients have to be treated appropriately with liability. Good state patients are overcharged at costs of zero. As mentioned earlier, the price \( t_\emptyset \) denotes the price at which profits are zero and equals average treatment costs per visiting patient. Now, we derive this price formally by setting profits to zero and solving for the price. \( t_\emptyset \) is implicitly given by dividing the total costs by the share of visiting patients:

\[ t_\emptyset = \frac{\kappa \int_{\hat{h}(t_\emptyset)}^{1} h f(h) dh}{\int_{\hat{h}(t_\emptyset)}^{1} f(h) dh}. \] \hspace{1cm} (3.25)

The physician solves the maximization problem

\[ \max_t \pi_{L/noV}(t). \] \hspace{1cm} (3.26)

The physician would not choose a price outside of the price range \( t \in (t_\emptyset, \alpha L + c) \), because profits are strictly positive in that range and zero outside. The first-order condition of (3.26) for \( t \in (t_\emptyset, \alpha L + c) \) is given by

\[ (1 - F(\hat{h}(t_{L/noV}^*))) = \left( f(\hat{h}(t_{L/noV}^*)) \hat{h}'(t_{L/noV}^*) \right) \left( t_{L/noV}^* - \kappa \hat{h}(t_{L/noV}^*) \right). \] \hspace{1cm} (3.27)

The left-hand side of (3.27) represents the marginal profits of an incremental price increase. They are the same as in (V), see (3.21). The right-hand side represents the marginal costs of an incremental price increase. The marginal costs are larger than in (V) as \( \hat{h}(t) < 1 \). The last term in the right-hand side of (3.27) shows that
the costs of a price increase depend on the patient type. Low types are more often good-state patients on which the physician can perform profitable overcharging. The marginal costs of a price increase are the larger, the higher the marginal patient type is, reflected by $\hat{h}(t)' > 0$.

As in the case with verifiability, the maximization problem (3.26) has a unique interior solution when the marginal costs of a price increase are monotonically increasing. This condition requires that $f(\cdot)$ does not decrease too sharply and is similar to, though slightly stronger than, the corresponding condition for $(V)$.

**Lemma 3.2.** The maximization problem (3.20) has a unique interior solution, i.e., $t^*_L/noV \in (\kappa, \alpha L + c)$, if $\frac{f'(\hat{h}(t))}{f(\hat{h}(t))} \geq -\frac{\alpha L}{t - \kappa h(t)} \forall t \in (\kappa, \alpha L + c)$ with strict inequality for some $t$.

Proof: see appendix 3.9.2.2.

### 3.5.3.3 Comparison of prices in different institutional settings

In $(L/noV)$, a price increase results in the loss of the physician’s most profitable patients. In $(V)$, all patients are equally profitable for the physician. Price increases are therefore associated with higher marginal costs when verifiability does not hold in $(L/noV)$ than when verifiability holds $(V)$. Intuition therefore suggests that the optimal price is higher in $(V)$. In the following, we present a proposition stating that the intuition is correct. The proof can be found in appendix 3.9.3.1. It is based on the theory of monotone statics (Milgrom and Shannon, 1994).

**Proposition 3.2.** Consider a monopolistic credence goods market and the equilibrium prices $t^*_V$ and $t^*_{noV}$ in the settings with and without verifiability, respectively. The equilibrium price $t^*_V$ is higher than the equilibrium price $t^*_{L/noV}$, i.e., $t^*_V > t^*_{L/noV}$. It directly follows that more patients prevent in $(V)$ than in $(L/noV)$.

As a direct consequence of the higher price in $(V)$, more patients engage in prevention in $(V)$ than in $(noV)$ and more patients visit the physician in $(L/noV)$ than in $(V)$.

### 3.6 Welfare & policy

The results of the foregoing section are illustrated in figure 3.6 and serve as the basis for the welfare discussion in this section.
We will exemplary discuss how the two policies of price regulation and social insurance, respectively, could be used to improve market outcomes, depending on the institutional setting. Social health insurance refers to an insurance system organized by the government with compulsory enrolment and a uniform product imposed on participants (Zweifel et al., 2009). Both policies can frequently be found in health care markets.

### 3.6.1 Welfare

We consider a utilitarian welfare function. Social welfare is calculated as the sum of payoffs of all agents in the model.\(^{30}\)

**Benchmark I: No Market.** We first consider the benchmark in which no market for medical treatment exists (section 3.3). Social welfare, denoted by \(\omega_{nM}\), is given by

\[
\omega_{nM} = \int_0^h f(h)(-hL)dh + \int_h^1 f(h)(-\alpha hL - c)dh. \tag{3.28}
\]

\(^{30}\)The utility of the monopolistic physician seems to have great weight compared to the utilities of patients. The monopolistic physician, however, should not necessarily be interpreted to represent only one person. By the argumentation of Zweifel et al. (2009) we can instead interpret the monopolistic physician as a medical association that acts a monopoly for its members.
The first term represents the welfare of patients who decide not to prevent, the right term represents the welfare of the patients who prevent. Here, welfare is independent of the treatment price $t$.

**Benchmark II: No Asymmetric Information.** Social welfare is denoted $\omega_{nA}(t)$ and given by

$$\omega_{nA}(t_{nA}) = \int_0^{h(t_{nA})} f(h)(-ht)dh + \int_{h(t_{nA})}^1 f(h)(-\alpha ht - c)dh + \pi_{nA}(t_{nA}).$$

(3.29)

All patients visit the physician in the bad state. The first term in (3.29) represents the utility of patients who do not prevent, the second term represents the utility of patients who prevent and the last term represents the profit of the physician as stated in (3.10). Comparing (3.28) and (3.29) shows that the introduction of the treatment technology does always increase welfare since $t_{nA} \in (\kappa, L]$.

**The Credence Goods Market.** In the credence goods market, social welfare, denoted $\omega_{CG}(t_{CG})$, $CG \in \{(V), (L/noV)\}$, is given by

$$\omega_{CG}(t_{CG}) = \int_0^{\hat{h}} f(h)(-hL)dh + \int_{\hat{h}}^{t_{CG}} f(h)(-\alpha hL - c)dh$$

$$+ \int_{t_{CG}}^1 f(h)(-t)dh + \pi_{CG}(t_{CG}),$$

(3.30)

where $\hat{h} = \frac{c}{(1-\alpha)L}$ and $\hat{h}(t_{CG}) = \frac{t_{CG} - c}{\alpha L}$. The first term in (3.30) represents the utility of patients who neither prevent nor visit. The second term represents the utility of patients who prevent but do not visit. The third term indicates the utility of the patients who visit but not prevent. Finally, $\pi_{CG}(t)$ indicates the profit of the physician, given by (3.19) and (3.24), respectively.

3.6.2 Policy

3.6.2.1 Price regulation

We assume that a regulating authority sets the price in the market. The participation constraint of the physician is assumed to remain valid, i.e., if the regulator sets prices which lead to negative profits, the physician stays out of the market.

Without asymmetric information, all inefficiencies (too much prevention) are due to the monopoly price (see section 3.4). Consequently, price regulation can restore the first-best market allocation which imposes both efficient treatment and
prevention behaviour by imposing marginal-cost pricing. The welfare maximising price is given by \( t_{\omega^*} = \kappa \).

**With verifiability (V).** In the credence goods markets, the effects of price regulation differ with respect to the institutional setting. With verifiability, the regulator can implement the first-best allocation by setting a price equal to marginal costs, \( t_{\omega^*_V} = t_{\omega^*_{nA}} = \kappa \). With this price, the physician has no incentive to overtreat and we assume that he behaves honestly.\(^{31}\) Despite the first-best can be implemented in both (V) and (nA), there are still differences in the patients’ visiting behaviour. In (V) patients need to visit the physician for a diagnosis, but the physician does not treat good-state patients. In (nA) patients do not need a diagnosis, because they knew their types. There, only bad state patients visit the physician.\(^{32}\)

If the first-best solution cannot be implemented for some reason, the regulator faces the situation illustrated in figure 3.4. By changing the price the regulator faces a trade-off between the inefficiencies illustrated in the upper part of figure 3.6. Say, starting from \( t_{\omega^*_V} \), the regulator reduces the price. This increases the inefficiencies of inefficient overtreatment and not enough prevention, but correspondingly reduces the inefficiency of untreated health problems. In our framework, reducing the inefficiency always outweighs the increased inefficiencies. This can be shown by using the welfare function (3.30), and calculating the derivative, leading to

\[
\frac{\partial \omega}{\partial t} \bigg|_{t^*_V} = f(\hat{h}(t))\hat{h}'(t)(-t + \kappa) < 0 \text{ if } t \in (\kappa, \alpha L + c).
\]  

(3.31)

Details can be found in appendix 3.9.4.

**With liability, without verifiability (L/noV).** In the institutional setting with only liability, (L/noV), the physician overcharges. Overcharging is associated with costs of zero and the incentive for overcharging persists at any price above zero. Therefore, the first-best cannot be implemented with price regulation and we are in a situation as illustrated in figure 3.4. Hence, the regulator is in a similar situation as just described for (V) and faces an inefficiency-trade off. However, in contrast

\(^{31}\) Note that this result only holds because of our assumptions that visiting and diagnosis costs are zero.

\(^{32}\) We can make this result independent of the assumptions adjusting the setup from the case without asymmetric information slightly. We could, without a change of results, assume that patients do not know their type before visiting the physician, but get to know their type in the diagnosis process – while the credence goods market is characterized that only the physician learns the diagnosis outcome. An example for situations in which patients observe the diagnosis outcome are the results of blood tests which are displayed by modern software packages in a way that they can be interpreted by both physician and patient.
to overtreating, overcharging is not inefficient per se. Hence, by reducing the price, say starting from $t^*_V$, the regulator reduces the inefficiency of untreated health problems and only increases the inefficiency of not enough prevention. We have already seen for $(V)$ that the inefficiency reduction outweighs the inefficiency increase when overtreatment is a problem. Therefore, the optimal regulated price in $(L/noV)$ is the lowest possible price, i.e., the price at which the physician makes zero profits. This price is below the costs of treatment, $\kappa$, because the physician overcharges good state patients. The price therefore equals the average costs of treatment, given that only bad state patients are treated, but revenues are generated from both good and bad state patients. Formally, the welfare-optimal price in obtained by dividing total costs by the number of visiting patients:

$$t_{\omega_{L/noV}} = t_0 = \frac{\kappa \int_{\hat{h}(t_0)}^{1} h f(h) dh}{1 - F(\hat{h}(t_0))}. \tag{3.32}$$

In appendix 3.9.4, We derive this result formally. From the welfare function (3.30), We calculate the derivative for the institutional setting $(L/noV)$ and obtain:

$$\frac{\partial \omega_{L/noV}(t)}{\partial t} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L \hat{h}(t) - c + \hat{h}(t) \kappa) < 0 \text{ if } t \in (t_0, \alpha L + c). \tag{3.33}$$

Price regulation achieves a welfare improvement compared to the market with monopoly pricing. Lower prices trade off an increased prevention inefficiency with a decreased welfare loss from untreated health problems for the sake of the latter. Further, the optimal regulated price may be low enough such that additionally the inefficiency of too much prevention is reduced. These dynamics explain the unintuitive finding that welfare can be higher although more patients choose inefficiently low prevention levels. We summarize the results of this section in proposition 3.3.

**Proposition 3.3.** Consider a credence goods market with regulatory price setting. Starting from the monopoly prices, reducing prices is welfare-improving in both $(V)$ and $(L/noV)$. Price reductions lead to higher levels of insufficient prevention and, in $(V)$, to higher levels of inefficient overtreatment, but reduce the inefficiency of untreated health problems and the latter has a larger welfare effect. The welfare maximizing prices equal average treatment costs per visiting patient. In $(V)$, the first-best allocation can be implemented with the price $t_{\omega_{V}} = \kappa$ and market outcomes equal the case without asymmetric information. In $(L/noV)$, the welfare maximising price is $t_{\omega_{L/noV}} = t_0 < \kappa$ and welfare is improved, but not first best.
3.6. Welfare & policy

3.6.3 Social health insurance

Health care in developed countries is usually embedded in social insurance systems. Social health insurance refers to an insurance system organized by the government, with compulsory enrolment and a uniform product imposed to participants (Zweifel and Eisen, 2012). Premiums are usually not based on risk (on the patient’s type in our terminology). As laid out by Zweifel and Eisen (2012), one explanation for the existence of such social insurance systems is that they are correcting market failures. The most frequently discussed market failures in this context are excessive time preference of consumers, altruistic motivation and adverse selection. Potentially, the inefficiencies stemming from supply-side moral hazard in credence goods markets may also be corrected by a social insurance scheme. In this chapter, we introduce such a social health insurance system into the model and analyse how welfare can be increased using the rate of coverage as a policy tool.\footnote{Risk-averse patients are not needed to justify the introduction of social health insurance. We maintain the assumption of risk-neutral patients.}

Switzerland may serve as an example. Prices for ambulant medical treatments are regulated by the Tarmed tariff which is negotiated between insurers and the suppliers of health care. The rate of coverage of the Swiss compulsory health insurance system, however, is set by the regulating authorities\footnote{The current rate of coverage is 90\%, the copayment rate is thus 10\% up to a threshold of CHF 700 per year, from which the copayment is reduced to zero.}.

We consider a risk-neutral monopolistic insurer (the government) who introduces social health insurance with compulsory enrolment and determines the rate of coverage. For simplicity, we assume that patients cannot opt-out of the insurance and therefore abstract from the discussion on whether democratic majorities for such a scheme exist (Zweifel and Eisen, 2012). Claims are financed by unified premiums. We assume that treatment prices in the different markets and settings are fixed and mirror some degree of market power on the supply-side. We further stipulate that prices equal those from sections 3.4 and 3.5 with a monopolistic physician. For the ease of exposition, we do not analyse the price response following the introduction of social health insurance.\footnote{The assumption of fixed prices may well be justified for the middle run, because prices in health care markets are often the result of long negotiations and do not adjust quickly. As an example, the last revision of the Swiss Dental Tariff dates back to 1994.}

The rate of coverage $\delta$ is defined as the share of the treatment price $t$ that is paid by the insurer, $\delta \in [0, 1]$. The corresponding copayment rate is given by $(1 - \delta)$.\footnote{In health economics, the term coinsurance rate is sometimes used synonymously with the term copayment rate.} Insurance is financed by a premium $p$ paid by each individual. We abstract from...
administrative costs and assume zero profits of the insurance system, therefore the sum of premiums is equal to the sum of claims.\footnote{Individual premiums hence depend on the rate of coverage, i.e., \( p = p(\delta) \). The sum of premiums is denoted \( P = P(\delta) \). Since we normalized the population to one, the individual premium is equal to the sum of premiums, i.e., \( p(\delta) = P(\delta) \).}

We can easily discuss this setup with reference to the logic of price regulation. The key difference to price regulation is that with social health insurance, the participation constraint of the supplier holds for any price paid by the patients after insurance. Consequently, prices can even be lower than with price regulation. The premium does not influence patients’ treatment and prevention decisions, because it has to be paid independently of the health state.

Without asymmetric information, social health insurance can implement the first-best allocation, because the only inefficiency – too much prevention – can be corrected by setting a copayment rate of \( (1 - \delta^*_n) \in (0,1) \). On the supply-side, revenues of the physician are not affected by the coinsurance rate. Hence, physician incentives cannot be changed by changing the copayment rate and the first-best is not implementable in both credence goods markets \((V)\) and \((L/noV)\). In the credence goods market, the regulator faces a trade-off between the different inefficiencies illustrated in figure 3.5 similar to the case with price regulation. We have already seen that price reductions increase welfare in the domain bounded from below by the participation constraint of the physician. Now that this constraint is relaxed, we need to analyse what happens when prices are further reduced. Since overcharging is not inefficient in contrast to overtreatment, increasing the copayment rate in \((L/noV)\) is associated with higher welfare costs in \((V)\) than in \((L/noV)\). Intuitively, the welfare-optimal copayment rate should therefore be higher in \((V)\) than in \((L/noV)\). As we show in appendix 3.9.5, this intuition is true. In the institutional settings \((V)\), the welfare-maximizing rate of coverage is given by \( (1 - \delta^*_V)\kappa = \delta^*_V \in (0,1) \) and a positive share of patients prevents. In \((L/noV)\) the welfare-maximizing rate of coverage is given by \( \delta^*_{L/noV} = 1 \), i.e. full insurance with no prevention.

It follows that price regulation is the better policy in \((V)\) while social insurance is the better policy in \((L/noV)\) (see appendix 3.9.6). Although highly stylized–we implicitly assume that a participation constraint of the physician in the analysis of price regulation and assume the absence of participation constraints for patients in the analysis of social insurance–, this result is fairly interesting as it implies that in order to achieve the best possible market allocations in credence goods markets, different policies may be adequate in different institutional settings.
3.7 Discussion

Physicians with social preferences. We have made the extreme assumption that physicians are not altruistic, although physician altruism has been considered in the economic health care literature for some years now (see McGuire (2000)). Further, studies in behavioural economics have pointed out the importance of behavioural traits such as guilt aversion (Charness and Dufwenberg, 2006) or lying aversion (Gneezy, 2005). Physician altruism could be modelled in different ways, but generally be introduced into the model. Assuming that because of altruistic motives, the physician sometimes behaves honestly although incentives incline him to overtreat or overcharge, respectively.\footnote{One way to justify such physician behaviour is to assume that there are many physicians instead of one and that some physicians always behave honestly while prices are still set by a medical association in a monopolistic manner.}

Increased honesty in physician behaviour would drive a wedge between the utility of visiting preventing patients and visiting not preventing patients and therefore make prevention worthwhile for some visiting patients. Moreover, more honest physician behaviour decreases the expected costs of visiting for patients who find visiting too expensive without altruism. This reduces the inefficiencies from untreated health problems.

Supplier-induced prevention. In the institutional setting without verifiability, (noV), the expert earns more with patients in the good state, because these patients are overcharged and produce zero costs of treatment. That is, the physician benefits when more patients prevent. If the physician had efficient tools to promote prevention in the society, he would be willing to invest in their use.\footnote{Again we refer to the idea to view the physician in the model as a medical association which could initiate prevention campaigns (Zweifel et al., 2009).} Whether this conjecture is an artefact of our model or can be transferred to real-world markets seems to be debatable, however.

Competition. Throughout this paper, we have assumed monopolistic price setting. The effects of competition are easy to analyse in the model, however.\footnote{Supply-side investments in prevention have been discussed by Schlesinger and Venezian (1986). In their model, however, asymmetric information is not an issue and the prevention decision is made by an insurer who has the ability to influence the loss-probabilities of the consumers directly.} In the model, Bertrand-competition between two physicians mimics the outcomes of optimal price-setting by a regulator. Price competition drives profits down to zero and leads to pricing at the average costs per visiting patient. That is, with verifiability, (V), prices are driven down to marginal costs \(\kappa\). In the setting without verifiability, \footnote{The effect of competition on market outcomes in markets with asymmetric information has recently been of interest in the experimental literature (see Dulleck et al. (2011), Huck et al. (2016b)) and Mimra et al. (2016a).}
(noV), prices are driven down to $t_0$. In the benchmark without asymmetric information, (nA) and in the institutional setting with verifiability, (V), the first-best allocation is achieved. Here, our results differ from Dulleck and Kerschbamer (2006). In their setting with consumer heterogeneity, competition leads to efficient outcomes in all institutional settings. In our model, efficiency cannot be restored in (noV), because overcharging remains profitable even at low prices.

Examples. The relevance of credence goods markets with prevention can be illustrated with examples from real world markets. The following examples should not be viewed as perfect analogies to the presented model, however, but rather as illustrations of the model’s basic features. The leading example in this paper is dental care. Many dental health services can be considered credence goods. Prevention plays an important role for the probability to get caries and other oral diseases, although natural disposition is important as well (Petersen et al., 2005). Preventive activities involve all aspects of dental hygiene such as brushing teeth, using dental floss and restricting the consumption of sugar-containing food. Overtreatment as a typical credence goods problem is a prevalent problem in dental markets (Mayes (1993); Imfeld (2008) and chapter 2 of this thesis). Marcenes et al. (2013) show that untreated oral diseases are associated with severe health problems. Still, less than half of the US population uses dental services annually, especially the less educated (Meyerhoefer et al., 2014). Colorectal cancer tests are another illustrative example for the main features of the presented model\textsuperscript{42}. According to the National Cancer Institute, a US government organization, patients can reduce the risk of getting colorectal cancer by physical activity, restrictive smoking and low alcohol consumption. The heterogeneity in our model can be justified by the fact that the risk of getting colorectal cancer crucially depends on a patient’s natural disposition, usually assessed by the personal and family history. Hamman and Kapinos (2014) report that only a minority of the population decides to undergo the testing procedure at all. Goodwin et al. (2011) report that a large portion of patients, among those who had already taken a cancer test once, undergoes the screening more frequently than recommended. The interpretation of the test as a treatment is justified by the fact that the test for colorectal cancer is quite involved\textsuperscript{43}.

Another example may be served by common psychological diseases such as stress, burnout and depression. At the diagnosis stage of the disease, the affected

\textsuperscript{42}We thank Albert Ma for suggesting this example.

\textsuperscript{43}Testing for colorectal cancer requires a screening colonoscopy. The procedure itself lasts more than one hour and involves preparations regarding the patient’s diet. According to Ktipp, a Swiss consumerism magazine, such a test costs between 400 and 800 Swiss Francs. Hamman and Kapinos (2014) reports costs between US$ 1’000 and 2’000 for the USA.
person has to consult a psychologist to obtain an opinion on whether treatment is needed or not. In this interpretation, psychological treatments constitute credence goods. Psychological diseases often follow far-reaching and sudden changes in the life of the affected persons. Such events may be job loss, retirement, child birth, death of relatives or career changes. After the occurrence of such an event, an affected person may consult a psychologist in order to obtain a recommendation on whether psychological treatment is needed or not. The affected person is in a “good state” when no psychological treatment is indicated and in a “bad state” when treatment is indicated. Prevention relates to preventive measures that reduce the probability of developing a bad state. In this example, preventive measures are activities in individual stress prevention, e.g., reading books, accessing information provided by public institutions or the participation at seminars. (Maslach et al., 2001). These activities come at a cost as they require investments of time, effort and money. With respect to the heterogeneity in the model people face different probabilities that they will be affected by a psychological disease, depending on personal characteristics and past experiences.

**Institutional Design.** We have assumed that institutions in credence goods markets are exogenously determined. In reality, however, the institutional design of markets is possibly endogenous. Liability can be promoted by laws and a legal system that guarantee victims of undertreatment that they can successfully sue for compensation. Verifiability can be promoted by regulations that require physicians to document treatments. The comparison of different institutional settings may therefore be interpreted in this fashion.

### 3.8 Conclusion

We have shown that it may be reasonable to expect different prevention and treatment patterns in credence goods markets than in markets without asymmetric information, because credence goods markets create inefficiencies that do not occur in markets without asymmetric information. Particularly, high risk patients have no incentive to prevent, because the inevitable experience of too much or too expensive treatment due to supply-side moral hazard in case of visiting the physician destroys the benefits of prevention. These results add a new point of view to the discussion of the problem of insufficient prevention. Moreover, these findings could help to explain why prevention programs do not work (Loewenstein et al., 2013). Demand-side considerations alone may be insufficient to understand observed prevention patterns and design welfare-improving policies. Additionally, supply-side
moral hazard increases the costs of visiting the physician such that low-risk patients refrain from visiting, resulting in untreated health problems. Hence the model also provides a basis for discussion of the problem of inefficient medical treatment in both directions – too much and too little.

Another insight from this model is that equilibrium prices differ with respect to the institutional setting of the market. Different forms of supply-side moral hazard – overtreatment in institutional settings with verifiability and overcharging in institutional settings with only liability – are associated with different costs for the physician. The physician sets lower prices in the institutional settings with only liability, because costs are lower with overcharging. Consequently, when liability holds, but verifiability does not we expect less patients to prevent, but also less untreated health problems.

Last, we have analysed the welfare effects of price changes. The most important insight here is that lower prices in our framework may increase welfare, although they come along with lower levels of efficient prevention. This result is rather counter-intuitive, but can easily be explained by the inefficiency trade-off in the credence goods market. Price reduction increases the inefficiency of not enough prevention (and too much treatment in institutional settings with verifiability), but at the same time lower prices reduce the inefficiency of untreated health problems. This inefficiency does not exist in markets without asymmetric information.
3.9 Appendix

3.9.1 Prevention and physician demand

The share of preventing patients in the institutional settings with verifiability is given by

\[ P_V(t) = \begin{cases} 
1 - F(\hat{h}) & \text{if } t < \kappa \\
1 - F(\hat{h}(t)) & \text{if } t = \kappa \\
F(\hat{h}) - F(\hat{h}(t)) & \text{if } t \in (\kappa, \alpha L + c] \\
1 - F(\hat{h}) & \text{if } t > \alpha L + c. 
\end{cases} \tag{3.34} \]

For prices below \( \kappa \) and above \( \alpha L + c \), the situation from the benchmark without market applies, because there is no supply and no demand, respectively. At \( t = \kappa \), there are no incentives for supply-side moral hazard and the market outcome is an the case without asymmetric information. For prices \( t \in (\kappa, \alpha L + c] \), the share of preventing patients is equal to the share of patients in group 2 who choose strategy \((nv,p)\) as illustrated in figure 3.4. In the setting with only liability, prevention is given by

\[ P_{L/noV}(t) = \begin{cases} 
0 & \text{if } t < t_0 \\
\int_0^{\hat{h}(t)} h f(h) dh + \alpha \int_{\hat{h}(t)}^1 h f(h) dh & \text{if } t \in [t_0, \alpha L + c] \\
1 - F(\hat{h}(t)) & \text{if } t \in (\kappa, \alpha L + c] \\
1 - F(\hat{h}) & \text{if } t > \alpha L + c. 
\end{cases} \tag{3.35} \]

For prices below \( t_0 \) and above \( \alpha L + c \), the situation from the benchmark without market applies, because there is no supply and no demand, respectively. With only liability, the physician has an incentive for overcharging even at low prices. Hence in contrast to \((V)\), there is no situation in which the credence goods problem is solved.

In \((V)\), the share of visiting patients is given by

\[ V_V(t) = \begin{cases} 
0 & \text{if } t < \kappa \\
\int_0^{\hat{h}(t)} h f(h) dh + \alpha \int_{\hat{h}(t)}^1 h f(h) dh & \text{if } t = \kappa \\
1 - F(\hat{h}(t)) & \text{if } t \in (\kappa, \alpha L + c] \\
0 & \text{if } t > \alpha L + c. 
\end{cases} \tag{3.36} \]

The share of visiting patients is zero when there is no supply, at prices below the treatment costs \( \kappa \). When \( t = \kappa \), the credence goods problem is solved, because
the physician behaves honestly and outcomes are as in the case without asymmetric information. At prices \( t \in (\kappa, \alpha L + c] \), supply-side moral hazard prevails (overtreatment of good state patients) and the share of visiting patients equals the share of group 3 as illustrated in figure 3.4. At prices above \( \alpha L + c \), demand is zero.

In \((L/noV)\), the share of visiting patients is given by

\[
V_{L/noV}(t) = \begin{cases} 
0 & \text{if } t < t_\emptyset \\
1 - F(\hat{h}(t)) & \text{if } t \in [t_\emptyset, \alpha L + c] \\
0 & \text{if } t > \alpha L + c.
\end{cases}
\]  

(3.37)

The share of visiting patients is zero when there is no supply, at prices below the zero-profit price \( t_\emptyset \). At prices \( t \in [t_\emptyset, \alpha L + c] \), supply-side moral hazard prevails (overcharging of good state patients) and the share of visiting patients equals the share of group 3 as illustrated in figure 3.4.

### 3.9.2 The physician’s price choice

#### 3.9.2.1 The physician’s price choice in \((V)\)

The physician solves the maximization problem

\[
\max_t \pi_V(t).
\]  

(3.38)

The first order condition of the maximization problem (3.38) for \( t \in (\kappa, \alpha L + c) \) is given by

\[
\left( 1 - F(\hat{h}(t^*_V)) \right) = f(\hat{h}(t^*_V)) \hat{h}'(t^*_V)(t^*_V - \kappa).
\]  

(3.39)

The marginal profit on the left-hand side of (3.39) is strictly monotone decreasing in \( t \) and approaching zero as the price \( t \) approaches the treatment costs \( \kappa \). A higher price means that fewer patients visit. The marginal costs on the right-hand side of (3.39) intuitively increase in the price, because the costs of loosing a patient, \( (t - \kappa) \), are increasing in \( t \). The physician does neither choose the corner solutions \( \kappa \) nor \( \alpha L + c \), because profit is zero at both corners and strictly larger than zero at all prices in between the corners. For the maximization problem to yield a unique interior solution it is thus sufficient that the marginal costs are monotonic increasing in \( t \) on \((\kappa, \alpha L + c)\). This is the case if

\[
\partial \left[ f(\hat{h}(t)) \hat{h}'(t - \kappa) \right] \geq 0
\]  

(3.40)
\[ f'(\hat{h}(t))\hat{h}'(t - \kappa) + f(\hat{h}(t))\hat{h}' \geq 0 \]
\[
\iff 
\frac{f'(\hat{h}(t))}{f(\hat{h}(t))} \geq -\frac{\alpha L}{t - \kappa},
\]

and condition (3.41) holds with strict inequality for some \( t \in (\kappa, \alpha L + c) \). We have used \( \hat{h}' = \frac{1}{\alpha L} \) to derive the last inequality. Condition (3.41) requires that \( f() \) must not decrease too sharply, i.e., that \( f'(()) \) does not become too negative, particularly when the price \( t \) approaches \( \alpha L + c \) such that the right-hand side (3.41) is comparatively large.

Note that (3.41) implies concavity of the profit function. To see this note that the derivative of the profit function for \( t \in (\kappa, \alpha L + c) \) is given by
\[
\frac{\partial^2 \pi}{\partial t^2} = -2f'(\hat{h}(t))\hat{h}'(t) - (t - \kappa)f'(\hat{h}'(t))\hat{h}'^2(t)
\]
\[
\iff 
\frac{1}{(\alpha L)} \left( -2f'(\hat{h}(t)) - \frac{1}{(\alpha L)}f'(\hat{h}(t))(t - \kappa) \right).
\]
The expression in (3.43) is strictly smaller than zero if
\[
\frac{f'(\hat{h}(t))}{f(\hat{h}(t))} > -\frac{2\alpha L}{t - \kappa},
\]
which is implied by (3.41).

3.9.2.2 The physician’s price choice in \((noV)\)

The physician solves the maximization problem
\[
\max_{t} \pi_{noV}(t).
\]
The physician chooses the price \( t_{noV}^* \) such that
\[
t_{noV}^* = \arg \max_{t} \pi_{noV}(t).
\]
First, we take the derivative of the profit function (3.24) for \( t \in (t_{\text{emptyset}}, \alpha L + c) \) with respect to \( t \) and obtain the first-order condition.

\[
\frac{\partial}{\partial t} \left[ t \int_{h(t)}^{1} f(h)dh - \kappa \int_{h(t)}^{1} hf(h)dh \right] = (1 - F(\hat{h}(t)) - tf(\hat{h}(t)\hat{h}'(t)) + \kappa f(\hat{h}(t))\hat{h}'(t))\hat{h}(t) \\
\Rightarrow (1 - F(\hat{h}(t_{noV}^*)) = \left( f(\hat{h}(t_{noV}^*)\hat{h}'(t_{noV}^*)) \right) (t_{noV}^* - \kappa \hat{h}(t_{noV}^*)) \tag{3.47}
\]

We can ignore corner solutions as profits are always positive between the corners and zero at the corners. Similar to the case with verifiability, we require the right-hand side of the first-order condition, the marginal costs of a price increase, be monotonic in the price \( t \), i.e., we require

\[
\left[ \left( f(\hat{h}(t_{noV}^*)\hat{h}'(t_{noV}^*)) \right) (t - \kappa \hat{h}(t_{noV}^*)) \right]' \geq 0 \tag{3.48}
\]

\[
\Leftrightarrow \frac{f'(\hat{h}(t))}{f(\hat{h}(t))} \geq - \frac{\alpha L}{t - \kappa \hat{h}(t)} \tag{3.49}
\]

for all \( t \in (\kappa, \alpha L + c) \) with strict inequality for some \( t \in (\kappa, \alpha L + c) \). We have used \( \hat{h}' = \frac{1}{\alpha L} \) to derive the last inequality. The assumption requires that \( f() \) must not decrease too sharply, i.e., that \( f'() \) does not become too negative. This condition is stronger than the condition in the case with verifiability in (3.41).

### 3.9.3 Comparison of prices between settings

#### 3.9.3.1 Prices in the Credence Goods Markets (\( V \)) and (\( noV \))

We show that \( t_{V}^* > t_{noV}^* \). We do this by comparing the two functions with the theory of monotone comparative statics (Milgrom and Shannon, 1994). Both \( \pi_{V}(t) \) and \( \pi_{noV}(t) \) have the same choice set, but different profit functions. Both profit function are assumed to have inner solutions, i.e., the restrictions defined in lemmas 3.1 and 3.2 are assumed to be satisfied. We introduce a variable \( z \) which is 0 if \( \pi(t)_{CG} = \pi(t)_{V} \) and 1 if \( \pi(t)_{CG} = \pi(t)_{noV} \). We introduce the new function \( u(t, z) \) given by

\[
u(t, z) = (1 - F(\hat{h}(t)) \cdot t - \kappa \int_{h(t)}^{1} h^{1-z} f(h)dh\]

98
and observe that \( u(t, 1) = \pi(t)V \) as \( (1 - F(\hat{h}(t))) \cdot (t - \kappa) = (1 - F(\hat{h}(t))) \cdot t - \kappa \int_{\hat{h}(t)}^{1} f(h)dh \). Furthermore, \( u(t, 0) = \pi(t)_{noV} \) as \( \pi(t)_{noV} = (1 - F(\hat{h}(t))) \cdot t - \kappa \int_{\hat{h}(t)}^{1} hf(h)dh \). Next, we check whether \( u(t, z) \) has (strictly) increasing differences in \((t, z)\) which would imply that \( t^*_V > t^*_{noV} \). The function \( u(t, z) \) has strictly increasing differences in \((t, z)\) if for all \( t^H, t^L \) in the choice domain such that \( t^H > t^L \), we have

\[
\begin{align*}
\pi_V(t^H) - \pi_V(t^L) &> \pi_{noV}(t^H) - \pi_{noV}(t^L) \\
\iff
u(t^H, 1) - u(t^L, 1) &> u(t^H, 0) - u(t^L, 0)
\end{align*}
\]

We obtain

\[
(1 - F(\hat{h}(t^H))) \cdot t^H - \kappa \int_{\hat{h}(t^H)}^{1} f(h)dh - (1 - F(\hat{h}(t^L))) \cdot t^L + \kappa \int_{\hat{h}(t^L)}^{1} f(h)dh
\]

\[
> (1 - F(\hat{h}(t^H))) \cdot t^H - \kappa \int_{\hat{h}(t^H)}^{1} hf(h)dh - (1 - F(\hat{h}(t^L))) \cdot t^L + \kappa \int_{\hat{h}(t^L)}^{1} hf(h)dh,
\]

which simplifies to

\[
\int_{\hat{h}(t^H)}^{1} hf(h)dh - \int_{\hat{h}(t^L)}^{1} hf(h)dh > \int_{\hat{h}(t^H)}^{1} f(h)dh - \int_{\hat{h}(t^L)}^{1} f(h)dh
\]

\[
\iff
\int_{\hat{h}(t^H)}^{1} f(h)dh > \int_{\hat{h}(t^L)}^{1} hf(h)dh.
\]

Both sides of the last inequality are positive since \( \hat{h}(t) \) is strictly increasing in \( t \) and \( f(\hat{h}(t)) > 0 \ \forall \hat{h}(t) \in [0, 1] \). Since \( \hat{h}(t) \in [0, 1] \) the inequality is always fulfilled in our model. We have shown that \( u(t, z) \) has strictly increasing differences in \((t, z)\) and consequently \( t^*_V > t^*_{noV} \).

### 3.9.4 Price regulation

#### 3.9.4.1 Without asymmetric information

From (3.29) welfare is given by

\[
\omega_{nA}(t) = \int_{0}^{\hat{h}(t)} f(h)(-ht)dh + \int_{\hat{h}(t)}^{1} f(h)(-\alpha ht - c)dh + \pi_{nA}(t).
\]

(3.50)
Chapter 3

The first term represents the utility of all patients who do not prevent. The second term represents the utility of the preventing patients. The last term denotes the profit of the physician as established by equation (3.10) for \( t \in (\frac{c}{1-\alpha}, L] \) and zero for prices out of this range. We divide \( \omega_{nA}(t) \) into four terms \( i = \{1, 2, 3, 4\} \) denoted by \( \omega_{nA_i/4} \) such that

\[
\omega_{nA_1/4} = \int_0^{\hat{h}(t)} f(h)(-ht)dh = -t \int_0^{\hat{h}(t)} f(h)dh,
\]

\[
\omega_{nA_2/4} = \int_{\hat{h}(t)}^1 f(h)(-\alpha ht - c)dh = \int_{\hat{h}(t)}^1 f(h)(-\alpha ht)dh + \int_{\hat{h}(t)}^1 f(h)(-c)dh
\]

\[
= t \int_{1}^{\hat{h}(t)} f(h)\alpha h dh + \int_{1}^{\hat{h}(t)} f(h)cdh,
\]

\[
\omega_{nA_3/4} = (t - \kappa) \int_0^{\hat{h}(t)} f(h)dh,
\]

\[
\omega_{nA_4/4} = (t - \kappa)\alpha \int_{\hat{h}(t)}^1 f(h)dh.
\]

We take the partial derivatives of \( \omega_{nA_i/4} \) for \( i = \{1, 2, 3, 4\} \) with respect to \( t \) and obtain

\[
\omega'_{nA_1/4} = -\int_0^{\hat{h}(t)} f(h)dh + t f(\hat{h}(t))\hat{h}(t)\hat{h}'(t),
\]

\[
\omega'_{nA_2/4} = \alpha \int_{1}^{\hat{h}(t)} f(h)dh + tf(\hat{h}(t))\alpha \hat{h}(t)\hat{h}'(t) + f(\hat{h}(t))\hat{c}\hat{h}'(t),
\]

\[
\omega'_{nA_3/4} = \int_0^{\hat{h}(t)} f(h)dh + (t - \kappa) f(\hat{h}(t))\hat{h}(t)\hat{h}'(t),
\]

\[
\omega'_{nA_4/4} = \alpha \int_{\hat{h}(t)}^1 f(h)dh + (t - \kappa)(-\alpha) f(\hat{h}(t))\hat{h}(t)\hat{h}'(t).
\]

We add \( \omega'_{nA_1/4} \) and \( \omega'_{nA_2/4} \) and simplify

\[
\omega'_{nA_1/4} + \omega'_{nA_2/4} = -\int_0^{\hat{h}(t)} f(h)dh + \alpha \int_{1}^{\hat{h}(t)} f(h)dh + f(\hat{h}(t))\hat{h}'(t) \left( -t\hat{h}(t) + \alpha \hat{h}(t) + c \right),
\]

\[
= -\int_0^{\hat{h}(t)} f(h)dh + \alpha \int_{1}^{\hat{h}(t)} f(h)dh + 0,
\]

\[
= -\int_0^{\hat{h}(t)} f(h)dh - \alpha \int_{\hat{h}(t)}^1 f(h)dh,
\]

where in the step from the first to the second line we used

\[ \alpha \hat{h}(t) - t\hat{h}(t) = (\alpha - 1)\hat{h}(t)t = (\alpha - 1)\frac{c}{(1-\alpha)}t = -c. \]
3.9. Appendix

Last, we calculate \( \frac{\partial \omega_{nA}(t)}{\partial t} \):

\[
\frac{\partial \omega_{nA}(t)}{\partial t} = \sum_{i=1}^{4} \omega'_{nA_i} = (1 + \alpha)(t - \kappa) f(\hat{h}(t))\hat{h}(t)\hat{h}'(t).
\] (3.51)

Since \( \hat{h}'(t) = -\frac{c}{(1-\alpha)t^2} < 0 \), we have \( \frac{\partial \omega_{nA}(t)}{\partial t} < 0 \) for \( t \in (\kappa, L) \). Since \( \omega_{nA}(t) \) is continuous on \([\kappa, L]\), the welfare maximizing price \( t_{\omega_{nA}} \) is given by \( t_{\omega_{nA}} = \kappa \).

3.9.4.2 The Credence Goods Market

In section 3.6 we stated that

\[
\frac{\partial \omega_V(t)}{\partial t} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L\hat{h}(t) - c + \kappa) < 0 \text{ for } t \in (\kappa, L),
\] (3.52)

and

\[
\frac{\partial \omega_{noV}(t)}{\partial t} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L\hat{h}(t) - c + \hat{h}(t)\kappa) < 0 \text{ for } t \in (t_\emptyset, L).
\] (3.53)

Here, we show, how these results are derived. Welfare in the credence goods setting, for prices \( t \in (\frac{c}{1-\alpha}, L) \) when all three patient groups exist, is given by

\[
\omega_{CG}(t) = \int_{0}^{\hat{h}} f(h)(-hL)dh + \int_{\hat{h}}^{\hat{h}(t)} f(h)(-\alpha hL - c)dh + \int_{\hat{h}(t)}^{1} f(h)(-t)dh + \pi_{CG}(t).
\] (3.54)

The profit function \( \pi_{CG}(t) \) depends on the institutional setting \( CG \in \{(V), (noV)\} \). We denote the terms of \( \omega_{CG}(t) \) as well as the two profit functions \( \omega_{CG_i} \) with \( i = \{1, 2, \ldots, 5\} \), respectively, such that

\[
\omega_{CG_1} = \int_{0}^{\hat{h}} f(h)(-hL)dh,
\]

\[
\omega_{CG_2} = \int_{\hat{h}}^{\hat{h}(t)} f(h)(-\alpha hL - c)dh,
\] (3.55)

\[
\omega_{CG_3} = \int_{\hat{h}(t)}^{1} f(h)(-t)dh = t \int_{1}^{\hat{h}(t)} f(h)dh,
\]

\[
\omega_{CG_4} = \pi_V(t) = \int_{\hat{h}(t)}^{1} f(h)(t - \kappa)dh = -\int_{1}^{\hat{h}(t)} f(h)(t - \kappa)dh, \quad \text{for } t \in (\kappa, \alpha L + c)
\]

\[
\omega_{CG_5} = \pi_{noV}(t) = t \int_{\hat{h}(t)}^{1} f(h)dh - \kappa \int_{\hat{h}(t)}^{1} hf(h)dh
\] (3.56)
Chapter 3

\[ -t \int_{1}^{\hat{h}(t)} f(h)dh + \kappa \int_{1}^{\hat{h}(t)} hf(h)dh \quad \text{for } t \in (t_{0}, \alpha L + c). \]  

(3.57)

Now, we take the partial derivatives of \( \omega_{CGi} \), \( i = \{1, 2, \ldots, 5\} \) with respect to \( t \) and obtain

\[ \omega'_{CG1} = 0, \]
\[ \omega'_{CG2} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L\hat{h}(t) - c), \]
\[ \omega'_{CG3} = \int_{1}^{\hat{h}(t)} f(h)dh + tf(\hat{h}(t))\hat{h}'(t), \]
\[ \omega'_{CG4} = \int_{1}^{\hat{h}(t)} f(h)dh - (t - \kappa)f(\hat{h}(t))\hat{h}'(t), \]
\[ \omega'_{CG5} = -\int_{1}^{\hat{h}(t)} f(h)dh - tf(\hat{h}(t)) - \kappa\hat{h}'(\hat{h}(t)) + \kappa\hat{h}(t)f(\hat{h}(t))\hat{h}' + \kappa F(\hat{h}(t))\hat{h}' \]
\[ = -\int_{1}^{\hat{h}(t)} f(h)dh - (t - \kappa\hat{h}(t)) f(\hat{h}(t))\hat{h}'(t). \]

With verifiability (V) Calculating \( \sum_{i=1}^{4} \omega'_{CGi} \) we obtain

\[ \sum_{i=1}^{4} \omega'_{CGi} = \frac{\partial \omega_{V}(t)}{\partial t} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L\hat{h}(t) - c + \kappa) \]
\[ = f(\hat{h}(t))\hat{h}'(-\alpha L\frac{t - c}{\alpha L} - c + \kappa) \]
\[ = f(\hat{h}(t))\hat{h}'(-t - \kappa). \]

Since \( f(\cdot) > 0 \) and \( \hat{h}' = \frac{1}{\alpha L} > 0 \), we have \( \frac{\partial \omega_{V}(t)}{\partial t} < 0 \) for \( t \in (\kappa, \alpha L - c) \). Thus, a price decrease in this range is welfare-improving.

With liability, without verifiability (noV) Calculating \( \sum_{i=1}^{3} \omega'_{CGi} + \omega'_{CG5} \) we obtain

\[ \sum_{i=1}^{3} \omega'_{CGi} + \omega'_{CG5} = \frac{\partial \omega_{noV}(t)}{\partial t} = f(\hat{h}(t))\hat{h}'(t)(-\alpha L\hat{h}(t) - c + \hat{h}(t)\kappa) \]
\[ = f(\hat{h}(t))\hat{h}'(t) \left(-\alpha L\frac{t - c}{\alpha L} - c + \frac{t - c}{\alpha L}\kappa\right) \]
\[ = f(\hat{h}(t))\hat{h}' \left(-t + \frac{t - c}{\alpha L}\kappa\right) \]  

(3.58)
The first order-condition for a critical point $t^*$ is given by
\[
\hat{f}(\hat{h}(t^*))\hat{h}'\left(-t^* + \frac{t^*-c}{\alpha L \kappa}\right) = 0.
\] (3.59)

Since $f(\cdot) > 0$ and $\hat{h}' = \frac{1}{\alpha L} > 0$, the condition boils down to
\[
\frac{t^*-c}{\alpha L \kappa} = t^* - \frac{c\kappa}{\kappa - \alpha L}.
\]

Since $f(\cdot) > 0$ and $\hat{h}' = \frac{1}{\alpha L} > 0$, the sign of (3.58) equals the sign of $(-t + \frac{t^*-c}{\alpha L \kappa})$.

If ($\alpha L \geq \kappa$), then for all prices $t > 0$, $\text{sgn}(\partial \omega_{\text{nol}}(t)/\partial t) < 0$. This implies that the welfare optimal price for $\alpha L \geq \kappa$ is given by $t_0$, the zero-profit price defined in (3.25). With ($\alpha L \geq \kappa$), there is no critical point for which $t^* > 0$.

If ($\alpha L < \kappa$), we obtain
\[
\text{sgn}(\partial \omega_{\text{nol}}(t)/\partial t) < 0 \text{ if } 0 < t < \frac{c\kappa}{\kappa - \alpha L} = t^* \text{, and}
\]
\[
\text{sgn}(\partial \omega_{\text{nol}}(t)/\partial t) > 0 \text{ if } 0 < t > \frac{c\kappa}{\kappa - \alpha L} = t^*.
\]

Hence, the critical point $t^*$ is a minimum. The welfare optimal price for $\alpha L < \kappa$ is given by $t_0$, the zero-profit price defined in (3.25).

### 3.9.5 Social insurance

#### 3.9.5.1 Social insurance without asymmetric information

In the benchmark case without asymmetric information, every patient in a bad state visits the physician. The expected utility for a patient who decides not to prevent is given by
\[
EU_{np}(\delta) = -h(1 - \delta)t - p.
\]

The expected utility of a patient who decides to prevent is given by
\[
EU_{p}(\delta) = -\alpha h(1 - \delta)t - c - p.
\]

The higher the patient’s type $h$, the larger are her prevention incentives. The population is divided into two groups, similar to the case without social insurance. Low types do not prevent, high types prevent. The indifferent type is denoted by the
Chapter 3

cutoff \( \tilde{h}_I(\delta) \) and given by

\[
\tilde{h}_I(\delta) = \frac{c}{(1 - \alpha)(1 - \delta)t}.
\]

The cutoff is independent of the insurance premium \( p \), because every patient has to pay it regardless of her prevention decision. The larger \( \delta \), i.e., the larger the rate of coverage, the higher is the cutoff, \( \frac{\partial h_I(\delta)}{\partial \delta} = \frac{c}{(1 - \alpha)(1 - \delta)^2t} > 0 \), because with cheaper treatment, prevention becomes less attractive. Once the rate of coverage has reached a threshold, no patient prevents anymore and the patient population solely consists of one group of not preventing patients. This threshold is obtained by setting \( \tilde{h}_I(\delta) = 1 \) which leads to \( t(1 - \delta) = \frac{c}{1 - \alpha} = t_0 \). Once the copayment is lower than the price \( t_0 \), no patient prevents. Since by assumption, \( t_0 < \kappa \), a situation without prevention does not occur without insurance, as the physician would stop supplying the market at such low prices. With social insurance, however, the amount earned by the physician is still above marginal costs.

Without asymmetric information, the only inefficiencies are distorted prevention decisions due to the monopoly price. The regulator can set the rate of coverage at a level that eliminates this inefficiency. The optimal rate of coverage \( (1 - \delta^*_nA) \) fulfills \( (1 - \delta^*_nA)t = \kappa \), such that the copayment equals marginal costs (derivation available on demand). Since the price in the market is \( t^*_nA > \kappa \), it follows that \( (1 - \delta^*_nA) \in (0, 1) \). The regulator can implement the first-best cutoff \( h^*_nA \), as in the case with price regulation discussed in section 3.6.2.1.

### 3.9.5.2 Social insurance in the Credence Goods Market

In settings with verifiability (and with or without liability), \( (V) \), physicians have incentives to overtreat patients. In settings with only liability, \( (L/noV) \), physicians have incentives to overcharge patients. The different institutional settings also have different implications for the optimal rate of coverage when social insurance is in place. In \( (V) \), contrasting the case of price policy, it is not possible to eliminate all inefficiencies as incentives for overtreatment remain at all copayment rates. In \( (L/noV) \), a similar logic as in the case of price regulation applies and the regulator should set the highest possible coinsurance rate in order to reduce the inefficiency stemming from untreated health problems.

In equilibrium, two situations are possible. First, a situation with three patient groups and second a situation with only two patient groups and no prevention. The first situation is similar to the situation from section 3.5 and illustrated in
This situation occurs in the coinsurance domain \((1 - \delta)t \in [t_0, t] \iff \delta \in [0, 1 - \frac{t_0}{t}]\). The second situation occurs when the rate of coverage is high and hence the copayment is low. This is the case in the coinsurance domain \((1 - \delta)t \in [0, t_0] \iff \delta \in (1 - \frac{t_0}{t}, 1]\). This situation is illustrated in figure 3.8. Such a situation does not occur without insurance, because we assume \(t_0 < t_0 < \kappa\) and at such low prices the physician does not supply his services. With social insurance, however, there is a wedge between between the price paid by the patient, the copayment \((1 - \delta)t\), and the price received by the physician, \(t\), such that this situation is possible.

Consider the situation illustrated in figure 3.7 which holds for relatively high copayments, \((1 - \delta)t \in [t_0, t] \iff \delta \in [0, 1 - \frac{t_0}{t}]\). The patient population splits into three groups similar to the case without social insurance. Again, due to supply-side moral hazard in the credence goods market at prices above costs, no patient visits and prevents at the same time. The expected utility of a patient who decides not to visit, nor to prevent, a strategy denoted \((nv, np)\), is given by

\[
EU_{nv,np}^I(\delta) = -hL - p.
\]

The expected utility of a patient who decides not to visit, but to prevent, a strategy denoted \((nv, p)\), is given by

\[
EU_{nv,p}^I(\delta) = -\alpha hL - c - p.
\]
The indifferent type is given by \( \hat{h}_I = \hat{h} = \frac{c}{(1-\alpha)L} \). Patients below the cutoff choose not to visit the physician, nor to prevent. The cutoff does not depend on the rate of coverage. Patients above the cutoff choose not to visit, but to prevent or to visit and not to prevent, a strategy denoted \((v, np)\). The expected utility of a patient who decides to visit and not to prevent is given by

\[
EU_{v, np}(\delta) = -(1-\delta)t - p.
\]

The indifferent type is given by

\[
\hat{h}_I(\delta) = \frac{(1-\delta)t - c}{\alpha L}. \tag{3.60}
\]

Types below \( \hat{h}_I(\delta) \) (and above \( \hat{h} \)) choose not to visit, but to prevent. Types above \( \hat{h}_I(\delta) \) choose to visit and not prevent. A rate of coverage of zero yields the same cutoff as in the market without social insurance. Coinsurance increases the patients’ benefits from physician visits. The higher the rate of coverage, the lower the cutoff \( \frac{\partial h_I(\delta)}{\partial \delta} = -\frac{t}{\alpha L} < 0 \). The introduction of social insurance thus leads to a reduction of prevention activities while it increases physician visits.

Now consider the second situation illustrated in figure 3.8 for relatively low copayments \((1-\delta)t \in [0, t_0) \iff \delta \in (1- \frac{t_0}{t}, 1]\).

**FIGURE 3.8**

The credence goods market with social insurance and low copayments.

Note: Credence goods market with social insurance in the case \((1-\delta)t \in [0, t_0) \iff \delta \in (1- \frac{t_0}{t}, 1]\).

In this situation, copayments are so low that no patient prevents. The patient population divides into two groups with the strategy not to visit, nor to prevent \((nv, np)\) below a marginal type and the strategy to visit and not to prevent \((v, np)\) above the marginal type. Patients choose their strategy depending on their expected
utility. Hence, for $(1 - \delta)t \in [0, t_0)$:

$$EU_{nv,np}^I(\delta) \leq EU_{v,np}^I(\delta) \iff -hL - p \leq -(1 - \delta)t - p \iff h \geq \frac{(1 - \delta)t}{L}.$$  

The marginal type is denoted $\hat{h}^I(\delta)$ and given by

$$\hat{h}^I(\delta) = \frac{(1 - \delta)t}{L}.$$  

(3.61)

In both situations, the claim of each visiting patient is $\delta t$. The premium is calculated by multiplying the claim $\delta t$ with the proportion of treated patients and given by $p(\delta) = P(\delta) = \delta t \int_{h(\delta)}^{\hat{h}(\delta)} f(h) dh$.

The utilitarian welfare function in the credence goods market with social insurance either represents the situation depicted in figure 3.7 with three patient groups for rate of coverages $(1 - \delta)t \in [t_0, t)$ or the situation depicted in figure 3.8 for rate of coverages $(1 - \delta)t \in [0, t_0)$. Social welfare for $(1 - \delta)t \in [0, t_0)$ is given by

$$\omega_C^I(\delta) = \int_0^{\hat{h}(\delta)} f(h)(-hL)dh + \int_{\hat{h}(\delta)}^{\hat{h}(\delta)} f(h)(-\alpha hL - c)dh + \int_{\hat{h}(\delta)}^1 f(h)(-t(1 - \delta))dh + \pi_{CG,I} - t\delta \int_{\hat{h}(\delta)}^1 f(h)tdh.$$  

(3.62)

The first two terms represent the patients’ welfare for each of the two groups, respectively. The last term is the premium paid by the population. Profits for both
cases are given by

\[ \pi_V(\delta) = \begin{cases} (t - \kappa) \int_{\delta}^{1} f(h) dh & \text{if } (1 - \delta)t \in [0, t_0) \\ (t - \kappa) \int_{\delta}^{1} f(h) dh & \text{if } (1 - \delta)t \in [t_0, t) \end{cases} \]  

(3.64)

and

\[ \pi_{L/noV}(\delta) = \begin{cases} t \int_{\delta}^{1} f(h) dh - \kappa \int_{\delta}^{1} hf(h) dh & \text{if } (1 - \delta)t \in [0, t_0) \\ t \int_{\delta}^{1} f(h) dh - \kappa \int_{\delta}^{1} hf(h) dh & \text{if } (1 - \delta)t \in [t_0, t). \end{cases} \]  

(3.65)

The regulator maximizes social welfare by setting the rate of coverage in each institutional setting. Importantly, the introduction of social insurance does not alter the physician’s incentives for overtreatment and overcharging, respectively. This is where the regulatory tool of social insurance differs from the tool of price regulation. In the following, we will analyse the optimal rate of coverage the two institutional settings \((V)\) and \((L/noV)\) separately.

**Institutional settings with verifiability (V).** With verifiability, an increase in the rate of coverage leads to more inefficient overtreatment through increased physician visits. The first-order condition for the maximization problem in the settings with verifiability \((V)\), is given by (derivation available on demand)

\[ f(\hat{h}(\delta^*)) \hat{h}'( - \alpha \hat{h}(\delta^*) L - c + \kappa) = 0. \]  

(3.66)

The welfare maximizing rate of coverage is given by \(\delta^* = 1 - \frac{\kappa}{t} \iff (1 - \delta^*)t = \kappa\). Patients pay the marginal costs. Patients who visit do not prevent and the physician still overtreats patients. Therefore inefficient treatment and prevention behaviour persists. Moreover, patients below the cutoff \(\hat{h}(\delta^*)\) do not visit and inefficiently suffer from untreated health problems. Still, social insurance increases social welfare by changing patients’ incentives compared to the situation without social insurance and a monopolistic physician by decreasing the inefficiency from untreated health problems at the cost of increasing both the inefficiencies of overtreatment and insufficient prevention.

**Institutional Settings with only Liability (L/noV).** With only liability, a similar logic as with price regulation applies. By increasing the rate of coverage, the regulator trades off the inefficiency of non-preventing patients with the inefficiency that patients suffer from untreated health problems. Overcharging is not inefficient
and hence in contrast to \((V)\), the inefficiency that patients visit the physician too often is not present. As with price regulation, the regulator should increase physician demand as much as possible. The optimal rate of coverage is therefore given by \(\delta^*_L/noV = 1\), i.e., by full insurance. This implies no prevention, as illustrated in figure 3.8. A welfare maximising regulator tolerates the inefficiency of no prevention for the sake of a decreased inefficiency of untreated health problems. We summarize the results of the whole section in the following result. A formal derivation is available on demand.

**Result 3.3.** Consider the introduction of compulsory social insurance with risk-independent premiums and fixed prices at monopolistic levels. The rate of coverage \(\delta \in [0, 1]\) is set by a regulator. Without asymmetric information, the first-best allocation is implemented by the rate of coverage that fulfills \((1 - \delta^*_A)t = \kappa\) such that \(\delta^*_V \in (0, 1)\). In the institutional settings \((V)\), the welfare-maximizing rate of coverage is also given by \((1 - \delta^*_V)t = \kappa\) such that \(\delta^*_V \in (0, 1)\), but does not solve the supply-side moral hazard problem. A positive share of patients prevents. In \((L/noV)\) the welfare-maximizing rate of coverage is given by \(\delta^*_L/noV = 1\), i.e. full insurance with no prevention.

A comparison of these results with price regulation yields remarkable insights. While price regulation is a better policy tool than social insurance in the institutional settings with verifiability, \((V)\), social insurance is a better policy tool in the setting with only liability, \((L/noV)\)\(^{44}\). The reason is that in \((V)\), the supply-side moral-hazard problem can be solved with price policy by setting prices equal to marginal costs. This is not the case in \((L/noV)\), because incentives for supply-side moral-hazard remain at any price above zero. With both price regulation and insurance, the regulator in \((L/noV)\) trades off the inefficiency of low prevention with the inefficiency of untreated health problems for the sake of the latter. With price regulation, the regulator’s urge to increase physician demand through lower prices decreases the physician’s profit. The regulator is constraint not to let the physician’s profit become negative, while this constraint is absent with insurance. This finding is presented in the following result (proof in appendix 3.9.6).

**Result 3.4.** In credence goods markets with the institutional settings with verifiability \((V)\), price regulation yields better welfare results than social insurance under a set of specific assumption. In credence goods markets with the institutional set-

\(^{44}\)While this is obvious for \((V)\) as the first-best is better than a not first-best solution, We proof this statement for \((L/noV)\) in appendix 3.9.6.
ting with only liability \((L/noV)\), social insurance yields a higher welfare than price regulation.

3.9.6 Price policy vs. social insurance in \((L/noV)\)

We show that welfare with social insurance at the optimal coinsurance rate is larger than welfare with price regulation at the optimal price in the credence goods market with only liability, \((L/noV)\). The welfare function is:

\[
\omega^I_{L/noV}(\delta) = \int_0^{\delta \cdot h_f(\delta)} f(h)(-hL)dh + \int_{\delta \cdot h_f(\delta)}^1 f(h)(-t(1 - \delta))dh \\
+ (t - \kappa) \int_{\delta \cdot h_f(\delta)}^1 f(h)dh - \int_{\delta \cdot h_f(\delta)}^1 f(h)\delta t dh.
\]

Welfare with social insurance at the welfare-maximising coinsurance rate \(\delta^* = 1\) is given by inserting \(\delta = 1\) into the welfare function. Using \(\hat{h}_f(\delta = 1) = \frac{(1-1)t}{L} = 0\) we obtain

\[
\omega^I_{L/noV}(\delta = 1) = \int_0^{\hat{h}_f(\delta)} f(h)(-hL)dh + \int_0^{1} f(h)(-t(1 - \delta))dh \\
+ (t - \kappa) \int_{\delta \cdot h_f(\delta)}^1 f(h)dh - t \int_{0}^{1} f(h)dh \\
= - \kappa \int_{0}^{1} f(h)dh.
\]

Welfare with price policy at the welfare-maximising price \(t_{\omega^*_L/noV} = t_\emptyset\) (at which the physician’s profit is zero) is given by inserting \(t = t_{\omega^*_L/noV} = t_\emptyset\) into (3.54):

\[
\omega_{L/noV}(t = t_\emptyset) = \int_0^{\hat{h}_f(t_\emptyset)} f(h)(-hL)dh + \int_{\hat{h}_f(t_\emptyset)}^{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)} f(h)(-\alpha hL - c)dh + \int_{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)}^1 f(h)(-t_\emptyset)dh \\
= - L \int_0^{\hat{h}_f(t_\emptyset)} h f(h)dh - \alpha L \int_{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)}^{\hat{h}_f(t_\emptyset)} h f(h)dh - c \int_{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)}^1 f(h)dh - t_\emptyset \int_{\hat{h}_f(t_\emptyset)}^1 f(h)dh. \\
(3.67)
\]

In order to show that \(\omega^I_{L/noV}(\delta = 1) > \omega_{L/noV}(t = t_\emptyset)\) we reformulate:

\[
- \kappa \int_0^{1} f(h)dh = - \kappa \int_0^{\hat{h}_f(t_\emptyset)} f(h)dh - \kappa \int_{\hat{h}_f(t_\emptyset)}^{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)} f(h)dh - \kappa \int_{\hat{h}_f(t_\emptyset) \cdot h_f(t_\emptyset)}^1 f(h)dh. \\
(3.68)
\]
Now we compare (3.67) and (3.68) integral-wise. We have

\[-\kappa \int_0^{\hat{h}} f(h)dh > -L \int_0^{\hat{h}} hf(h)dh,\]

because \(\kappa < L\) by assumption. Furthermore we have

\[-\kappa \int_{\hat{h}}^{\hat{h}(t_0)} f(h)dh > -\alpha L \int_{\hat{h}}^{\hat{h}(t_0)} hf(h)dh - c \int_{\hat{h}}^{\hat{h}(t_0)} f(h)dh,\]

because

\[-\alpha L \int_{\hat{h}}^{\hat{h}(t_0)} hf(h)dh - c \int_{\hat{h}}^{\hat{h}(t_0)} f(h)dh < -(\alpha L + c) \int_{\hat{h}}^{\hat{h}(t_0)} hf(h)dh\]

and \(\alpha L + c > \kappa\) by assumption. Last, we use the definition of \(t_0\) from (3.25):

\[t_0 = \frac{\kappa \int_{\hat{h}(t_0)}^{\hat{h}} hf(h)dh}{1 - F(\hat{h}(t_0))} = \frac{\kappa \int_{\hat{h}(t_0)}^{\hat{h}} hf(h)dh}{\int_{\hat{h}(t_0)}^{\hat{h}} f(h)dh},\]

and reformulate

\[-t_0 \int_{\hat{h}}^{1} hf(h)dh = -\kappa \int_{\hat{h}(t_0)}^{1} f(h)dh \cdot \frac{\int_{\hat{h}}^{\hat{h}(t_0)} hf(h)dh}{\int_{\hat{h}}^{\hat{h}} f(h)dh} = -\kappa \int_{\hat{h}(t_0)}^{1} hf(h)dh.\]

Therefore, we have shown that

\[\omega_{L/noV}(\delta = 1) > \omega_{L/noV}(t = t_0).\]
4 Advice markets. An experimental investigation.

We analyse markets for financial advice inspired by the model of Inderst and Ottaviani (2012a) in a market experiment. Product providers may pay commissions to advisors in order to steer advice in their favour against a consumer’s need. Our results show that – in line with prediction – conflicts of interest are created endogenously, and advice is biased in all conditions. Advice is considerably less biased than predicted in a benchmark condition with undisclosed commissions in which consumers are unaware of commission payments. Neither disclosure of commissions, nor imposing fines for wrong advice on the advisor lead to the predicted decrease of commission levels. Still, market welfare is largest with disclosure, because it increases consumers’ acceptance rate of recommendations. Our results point to several psychological traits documented in non-market advice settings - and show how their effects can differ on a market setting; for instance the “burden of disclosure” improves results in our setting, because prices are low enough. We find indications for social preferences in the behaviour of firms, advisors and consumers which should encourage further research.

4.1 Introduction

In financial or health care markets, consumers have to turn to experts for advice about which third-party product best fits their needs. For instance, consumers turn to investment advisors who possess superior information about which investment fund the consumer should purchase given his or her personal financial targets. Patients turn to doctors and pharmacists in order to receive advice in terms of the appropriate pharmaceutical product given their health situation. Not surprisingly, Chater et al. (2010) report that around 80% of all financial investment decisions are made in the presence of an advisor. In health care, this percentage is likely to be even higher as in many countries medications cannot be purchased without consultation of a health care provider.
The economic significance of such markets is great and has been growing. The share of the financial sector in the US has risen from around 3% of GDP in 1945 to 9% in the 2000s (Philippon, 2015), an observation that holds for most industrialized countries (Kerschbamer and Sutter, 2017). Similarly, health care expenditures in the US have risen from around 5% of GDP to more than 15% between 1960 and 2010 (Chernew and Newhouse, 2012). Both sectors combined today account for around one fourth of GDP.

A common phenomenon in such markets with advice is that the product providers make payments to advisors, often in the form of (hidden) commissions. These payments are of considerable size. Anagol et al. (2017) estimate that commissions in the life insurance market in India sum up to almost ten percent of the sum of premiums. In the US, payments from pharmaceutical companies to health care providers of USD 8.18B have been disclosed for 2016, paid by 1,481 companies to 631,000 physicians and 1,146 hospitals.1 In Germany, payments of EUR 562 Mio. have been disclosed.2 Commission payments may create conflicts of interest for advisors by rewarding advice which is not in the best interest of consumers.3 While commissions can have the positive effect to steer advice towards more efficient products, biased advice can harm consumers and lead to inefficient market outcomes (Inderst and Ottaviani, 2012a).

Although biased advice is a prevalent problem, many consumers do not seem to be aware of it. According to Chater et al. (2010), more than 80% consumers who recently bought financial products stated that they trusted the financial advise they received and that more than 80% of consumers thought that the advise they received was unbiased. The topic of consumer naivité is also analysed by Malmendier and Shanthikumar (2007) who find that especially small traders in financial markets blindly follow biased stock recommendations. Similarly it has been documented for health care that patients resist the idea that advice by their own doctor is possibly biased (Gibbons et al., 1998).

The persistence of biased advice can be explained by the credence goods characteristic of many markets with advice. Credence goods are defined by the fact that

---

1 The Affordable Care Act initiated by the Obama-administration in 2010 has mandated disclosure of such payments in the health care sector. This data is published by the governmental organization CMS (openpaymentsdata.cms.gov).
2 Data for 2016. Source: Press release of the Association of Research-Based Pharmaceutical Companies (vfa) and the FSA (Verein für Freiwillige Selbstkontrolle für die Arzneimittelindustrie), 19 June 2017. In contrast to the CMS, data by the vfa is only disclosed on the aggregate level unless recipients opt-in to full disclosure. Only a minority of providers chose to do so.
3 Evidence for biased advice in real world financial advice markets has recently been documented in several empirical studies in different contexts (Cummins and Doherty, 2006; Bergstresser et al., 2009; Bhattacharya et al., 2012; Anagol et al., 2017; Sane and Halan, 2017, forthcoming).
consumers cannot judge the quality of advice even after purchase (Darby and Karni, 1973; Emons, 1997). For instance, it may be hard or impossible to distinguish a bad investment outcome due to bad advice from a bad investment outcome due to bad luck. Likewise in health care patients often cannot judge whether they needed the medication received or whether they would have achieved the same health outcome with a different medication or no medication at all.

Given the potential problems associated with biased advice it comes at no surprise that the regulation of such markets is a highly debated issue. The most prominent political intervention is the disclosure of conflicts of interest which arise from third-party payments. Many commentators and politicians share the view that disclosure is a valid tool to improve market outcomes. Reinforced by the financial crisis 2007/2008, regulators in all parts of the world have launched new or revised existing legislatives involving stricter disclosure requirements. Member states of the European Union will have to put the MiFid II legislation into practice soon. It involves new disclosure requirements or the ban of commissions altogether. In Switzerland, the FidLeg legislation follows a similar direction and similar steps have been taken in the UK (Angelova and Regner, 2013). With respect to health care similar developments can be observed. For instance, the Affordable Care Act initiated by the Obama-Administration has mandated the disclosure of payments from product providers to health care providers. In Europe, the producers of pharmaceutical products have recently launched voluntary disclosure initiatives. As noted by Loewenstein et al. (2012), most scientific medical journals now require the disclosure of conflicts of interest of their authors.

The political focus on disclosure of conflicts of interest has not yet been justified by scientific results, however. There is scarce evidence that disclosure leads to the expected improved market outcomes. Recent theoretical works have studied markets with advice including third-parties and empathised that disclosure may work in some settings while not in others (Inderst and Ottaviani, 2012a,b).

4“Bad” investment outcomes are not necessarily bad in themselves, but rather bad compared to other options or benchmarks. This adds further complication to their detection.

5Initiated by the European Federation of Pharmaceutical Industries and Associations (EFPIA) and its EFPIA Disclosure Code. National member organisations are responsible for the implementation. In Germany, for instance, it is organized by the FSA, a club comprised of members of the Association of Research-Based Pharmaceutical Companies (vfa).

6Also in other markets than financial advice disclosure of conflicts of interest has become a frequently demanded policy tool. One example is policy regulation (Kartal and Tremewan, 2016). In this setting, the advisor is an scientific expert, the consumers are regulatory agencies who depend on the expert’s advice. The advice given may be influenced by firms or institutions who finance the research of the scientific expert. According to Kartal and Tremewan, regulating agencies, academic journals, NGO’s and government agencies have been increasingly demanding transparency.
number of experimental studies which focus on the interaction between advisor and consumer have reported several drawbacks of disclosure and thereby indicated that experimental research is helpful in revealing behaviour that is not predicted by theory.

Our research contribution is twofold. First, we conduct the first market experiment that investigates markets with advice including third party product providers based on the theoretical work by Inderst and Ottaviani (2012a) and Inderst (2015). We use an experimental approach to test the framework of Inderst and Ottaviani with experimental subjects, providing a test for the theory. Second, we compare the effect of disclosure to another policy which is designed to theoretically produce the same market welfare – fines for advisors for wrong advice – and to two markets with undisclosed commissions in which consumers are informed and uninformed, respectively, about the existence of commissions.

Our experiment is designed such that theory predicts endogenously arising conflicts of interest, as the two firms have different production costs and the more cost efficient firm has incentives to pay higher commissions than the less cost efficient firm. We set up a market with undisclosed commissions in which consumers are not informed about the possibility of commissions. This market theoretically leads to highly biased advice at the costs of consumers and to inefficient outcomes. In light of the preceding remarks regarding many consumers’ expectations of unbiased advice, this setting most likely resembles real world markets for advice which are untouched by regulation. We further study another market with undisclosed commissions, but this time consumers are informed about the possibility of commissions. This market theoretically results in strong welfare improvements over the market without consumer information. Finally, there are two markets with different policy regimes: Disclosure of commissions and fines for wrong advice. Both policy regimes are designed such that they theoretically improve market welfare by the same amount, but they work through different channels. Disclosure of commissions initiates a feedback process between the willingness to pay of consumers and firms’ reactions regarding prices and commissions (Inderst and Ottaviani, 2012b). The more cost-efficient firm is disproportionally discouraged to pay commissions, leading to less biased advice. Fines in turn directly affect the advisor by making him less responsive to differences in commissions, again disproportionally discouraging the more cost-efficient firm to pay commissions. While disclosure only works when consumers react appro-

\[7\text{This market is designed to implement the model with naive consumers by Inderst and Ottaviani (2012b); Inderst (2015).}\]
4.2 Related literature

This study builds on a series of recent theoretical work by Inderst and Ottaviani (2012c,a,b); Inderst (2015). The authors develop a model of markets with advice and show how conflicts of interest, created endogenously by firms competing over commissions, can bias advice and lead to market outcomes that are not in the best interest of the consumer and potentially inefficient. The authors use markets for financial advice as their prime example, but they are applicable to other markets such as health care and regulatory advice as well (see section 4.1). In the finance literature, conflicts of interest with a focus on portfolio management have been analysed for a longer time (see Stracca (2006) for a survey). Since the financial crisis, several works with particular focus to financial markets have contributed to

8The authors also published a series of applied papers discussing political applications of their work with a focus on financial advice (Inderst and Ottaviani, 2010; Inderst, 2011; Inderst and Ottaviani, 2013).
the understanding of conflicts of interest (Bolton et al., 2007; Stoughton et al., 2011; Gennaioli et al., 2015). A recent survey of both the theoretical and empirical literature on conflicts of interest in the financial service industry is provided by Burke et al. (2015). Connected to this literature is the empirical accounting literature on the consequences of regulatory disclosure in finance, recently surveyed by Leuz and Wysocki (2016). Leuz and Wysocki report that evidence on the empirical effects of disclosure is rare as very few studies are able to identify causal effects and call for more experimental research.

The model by Inderst and Ottaviani (2012a) bears similarities to the credence goods model by Dulleck and Kerschbamer (2006). However, there are some important differences between the models. Most importantly, in Dulleck and Kerschbamer (2006), there is only one expert who is both an advisor and a producer of the traded products, which we term expert providers (see chapter 1 of this thesis). Inderst and Ottaviani instead focus on experts who only sell advice, which we term expert advisors. Potential conflicts of interest in the framework of Dulleck and Kerschbamer (2006) are created by the expert himself through the prices he sets for two offered products. Dulleck and Kerschbamer (2006) find, however, that in equilibrium experts choose price vectors with equal mark-ups, such that no conflicts of interest occur. This result is rather unsatisfying given the vast empirical evidence for conflicts of interest and biased advice in credence goods markets. Inderst and Ottaviani (2012a) introduce firms which produce the traded goods into the setup and endogenously create conflicts of interest that persist in equilibrium. Furthermore, consumer utility is modelled differently in the two models. While in Dulleck and Kerschbamer (2006) consumer utility is unambiguously determined by the expert’s behaviour, it is a stochastic function of expert behaviour in Inderst and Ottaviani (2012a), similar to other credence goods models (Emons, 1997). The consequence of this design is that advice remains a credence good even in case of an unmatched recommendation, because consumers are not able to infer the quality of advice.\footnote{In Dulleck and Kerschbamer (2006) the problem of undertreatment is similar to the problem of an unmatched recommendation in Inderst and Ottaviani (2012a). In the setting of the former, however, consumers can unambiguously relate this problem to fraudulent expert behaviour, making advice an experience good.}

Our study also relates to a number of experimental studies that have analysed the effects of disclosure in settings with exogenous conflicts of interest, hence focussing on the interaction between advisors and consumers without product providers. Many of these studies document negative or neutral effects of disclosure on the quality of advice, consumer payoffs or market efficiency (Cain et al., 2005; Koch and Schmidt, 2010; Lacko and Pappalardo, 2010; Cain et al., 2011;
4.2. Related literature

Loewenstein et al., 2011, 2012; Ismayilov and Potters, 2013). Loewenstein et al. (2011) provide a survey of the findings. One example the authors give for a negative impact of disclosure on advisor behaviour is “moral licensing” which refers to the tendency that advisors feel morally entitled to give biased advice when conflicts of interest are disclosed. We do not find evidence for such an effect in our experiment. Another effect is “strategic exaggeration”, referring to the tendency of advisors to give even more biased advice in anticipation consumers’ discounting due to disclosure. In our setting we rather find the opposite effect, a strategic restrain: Even though conflicts of interest are larger with disclosure, advisors do not give more biased advice. Another example of a potentially negative effect of disclosure on the recipients of advice is that consumers feel pressured to follow biased advice in the face of disclosure, because a rejection can be interpreted as an accusation of corruption (“burden of disclosure”). Interestingly, the (“burden of disclosure”) may have a positive effect in our experimental setting, because acceptance rates of recommendation are rather too low than too high. Lacko and Pappalardo (2010) find negative effects from information overload following disclosure on the buyers of mortgages. Loewenstein et al. (2011) further report that consumers often fail to make the right strategic conclusions from disclosed conflicts of interest; this finding is supported by Ismayilov and Potters (2013) who report that advisors do not expect consumers to react drastically to disclosed conflicts of interest and hence find neutral effects of disclosure on advisor behaviour.

Only recently, a small literature has introduced endogenous conflicts of interest by giving the advisor the possibility to reject (exogenous) third-party payments. These studies have helped to draw a more differentiated picture of the effects of disclosure by analysing different forms of disclosure and introducing elements of reciprocity into the advisor-consumer relationship that may serve as policy alternatives to disclosure (Chater et al., 2010; Sah and Loewenstein, 2014; Kartal and Tremewan, 2016).10 Sah and Loewenstein (2014) show that disclosure can be beneficial when advisors can choose to avoid conflicts of interest. Similarly, the advisor can reject third-party payments in Kartal and Tremewan (2016). The authors find some positive effects of disclosure, but only in the short run. Church and Kuang (2009) also contribute to the evidence that disclosure may work in some situation while not in others. The authors report that disclosure serves consumers better

10Angelova and Regner (2013) experimentally analyse biased advice and potential remedies apart from disclosure. The authors investigate whether reciprocal behaviour between advisor and consumer may help to mitigate the problem of biased advice and conclude that mutual opportunities to reciprocate have positive effects on the truthfulness of advice. Beyer et al. (2013) show experimentally that advisors strongly react to financial incentives in a financial advice framing.
when they have the possibility to sanction advisors. Chater et al. (2010) report several experiments in which they investigate consumer behaviour in retail finance settings. Their main finding is that disclosure can work in some setting and not in others. For instance, they find in an internet experiment that disclosure of conflicts of interest only raises consumers’ awareness if the disclosure is accompanied by an explicitly displayed warning. The authors demand that policies involving disclosure of conflicts of interest need to be designed carefully in order to minimise potential negative effects of disclosure and make sure that disclosure works in the first place.

The cited experiments have shown that the experimental investigation of advice situations yields remarkable insights which have not yet been obtained by theory and empirical studies. All of them have in common, however, that they do not include a supply-side of the market and therefore potentially lack important dynamics which are needed to understand markets for advice. Empirical research on markets with advice is limited due to the complex nature of real-world advice markets and the uniqueness of many policy interventions (Leuz and Wysocki, 2016). Experimental research may partly fill this gap. Our contributions are first, to investigate whole markets for advice with endogenous conflicts of interest. Second, we compare different conditions including the ubiquitous policy of disclosure and fines for the advisor as a policy alternative.

4.3 Theoretical framework

4.3.1 The setup

The model  Our model combines elements of Inderst and Ottaviani (2012a) and Inderst (2015). Two firms, A and B, each offer a product, denoted A and B respectively, to a customer via an advisor at prices \( p_A \) and \( p_B \), respectively. Firms have production costs \( c_A \) and \( c_B \), respectively. Firm A is more cost efficient, i.e., \( c_A < c_B \). The consumer is in one of two possible states, \( \theta = A, B \). Each state represents the product the consumer needs in the respective state. The utility of the consumer depends on whether his purchase decision matches his state. When both are matched, the consumer obtains an utility of \( v_h \), \( v_h > 0 \), and \( v_l, 0 < v_l < v_h \) when the purchased product does not match the state. We denote the difference \( v_h - v_l \) by \( \Delta v \). The consumer is of type \( q \), representing the ex-ante probability that his state will be A – and vice versa \( 1 - q \) for state B. The probability \( q \) is drawn from a distribution function \( G(q) \) on \([0, 1]\) which we assume to be uniform such that
4.3. Theoretical framework

\[ g(q) = 1 \text{ and } G(q) = q \text{ for } q \in [0, 1]. \]

The consumer does not know his type \( q \), but the advisor observes \( q \). For selling products the advisor may receive commissions \( f_i \) for \( i \in \{A, B\} \) and \( f_i \geq 0 \) from the product providing firms \( A \) and \( B \).

We assume that the advisor bears a cost \( w \) from recommending an ex-post unsuitable product to the consumer. This disutility may reflect losses in future business or reputation due to bad consumer reviews\(^{12}\). We further assume \( \Delta v > 2w \) in order to obtain a benchmark compared to which disclosure can possibly increase welfare, which will be discussed further below\(^{13}\) We specify a welfare-neutral fine imposed by a regulating authority denoted \( l \), \( l \geq 0 \), which has to be paid by the advisor on top of \( w \). When \( l = 0 \) we say that no fines are implemented. Note that even when the advisor performs advice in the most honest way, i.e. recommends product \( A \) when \( q > \frac{1}{2} \) and product \( B \) when \( q < \frac{1}{2} \) (and any of the two products when \( q = \frac{1}{2} \)), there will be unmatched recommendations ex-post\(^{14}\).

The timing of the game is as follows.

Stage 1: Firms simultaneously set commissions \( f_A \) and \( f_B \), respectively.
Stage 2: Firms simultaneously set prices \( p_A \) and \( p_B \).
Stage 3: The advisor observes \( q \) and recommends a product.
Stage 4: The consumer makes a purchase decision.

At stage 4, we restrict the consumer’s choice to a decision between purchasing the recommended product or no product at all. We do this in order obtain greater flexibility in the experimental parametrization\(^{15}\).

**Unbiased, biased and first-best advice.** Following Inderst and Ottaviani (2012a) we first analyse advisor behaviour by assuming that consumers always follow the recommendation (which will be the case in equilibrium). The advisor faces a

---

\(^{11}\)By imposing a uniform distribution we follow Inderst (2015). This specification is done with regard to the experimental implementation. More generally, assuming a symmetric distribution, we follow Inderst and Ottaviani (2012a); Inderst (2015). It makes sure that unbiased advice implies the same expected consumer payoff for each of the two products.

\(^{12}\)Credence qualities of the good do not allow to attribute bad outcomes to biased advice on the individual level. Hence the bad reviews may reflect advice outcomes sampled over many consumers (e.g., by consumer protection organizations or by the press).

\(^{13}\)In order to show the potentially negative welfare effects of disclosure, Inderst and Ottaviani (2012a) focus on a different benchmark.

\(^{14}\)Due to this characteristic, the recommendation of the advisor is a credence good in the sense of Darby and Karni (1973), because the consumer is not able to evaluate the accurateness of a single recommendation. Wrong advice cannot be avoided completely, even if the advisor acts in the best interest of the consumer as the cutoff \( q = \frac{1}{2} \) still leads to wrong advice with a probability of 25 percent.

\(^{15}\)This modification has no effect on results as the consumer follows the recommendation in equilibrium. Specifically, we relax assumption (3) in Inderst and Ottaviani (2012a)). The consequences of this are discussed by Inderst and Ottaviani (2012a), page 792.
Chapter 4

trade-off between receiving commissions and the costs associated with recommending an unsuitable product. Given $q$, the advisor’s expected utility from recommending product $A$ is given by $f_A - (1 - q)(w + l)$. The expected utility from recommending product $B$ is $f_B - q(w + l)$. The advisor consequently chooses a cutoff $q^*$, above which he recommends product $A$ and below which he recommends product $B$. Assuming an interior solution (and fixed commissions for the moment), it is given by

$$q^*(f_A, f_B) = \frac{1}{2} - \frac{f_A - f_B}{2(w + l)}. \quad (4.1)$$

We say that advice is unbiased if $q^* = \frac{1}{2}$, because at this cutoff, the likelihood of recommending a product that matches the consumer’s needs is maximized. From (4.1), advice is only unbiased when $f_A = f_B$ and biased whenever $f_A \neq f_B$. When $f_A > f_B$, the advisor is steered towards recommending product $A$ and $q^* < \frac{1}{2}$. Vice versa for $f_A < f_B$.

Because commissions are welfare-neutral, we can already determine the first-best cutoff which balances costs and benefits of advice. Since $c_A < c_B$, product $A$ should be recommended more often than product $B$ from an efficiency point of view, trading off lower production costs with a higher probability that a recommendation leads to a mismatch. The marginal difference in production costs $c_B - c_A$, i.e., the efficiency gain from decreasing the cutoff, is constant; the marginal expected costs from a mismatches increase when the cutoff decreases. At the first-best cutoff, $q_{FB}$, marginal costs and benefits balance. As formulated by Inderst (2015), at the first-best cutoff, the total expected payoff of recommending product $A$, $qv_h + (1 - q)v_l - (1 - q)w - c_A$, equals that from recommending product $B$, $(1 - q)v_h + qv_l - qw - c_B$. In contrast to Inderst (2015), we assume that the costs of recommending an unsuitable product, $w$, are not welfare-neutral, accommodating for our interpretation of this cost as an irrecoverable loss. We denote the first-best cutoff by $q_{FB}$. It is given by

$$q_{FB} = \frac{1}{2} + \frac{c_A - c_B}{2(\Delta v + w)}, \quad (4.2)$$

which is below one half since $c_A < c_B$.

**Naive and wary consumers.** A risk-neutral consumer purchases the recommended product whenever his expected payoff, i.e., his willingness to pay, from purchasing the product is at least as large as the price. For a given value of $q$, the expected payoff of purchasing product $A$ is equal to $qv_h + (1 - q)v_l = q\Delta v + v_l$ and for product $B$ it is $qv_l + (1 - q)v_h = -q\Delta v + v_h$. The consumer’s expected payoff
4.3. Theoretical framework

depends on his expectations about the advisor’s cutoff, denoted \( \hat{q}^* \). A consumer who expects to receive a recommendation for product A whenever \( q > \hat{q}^* \) and a recommendation for product B whenever \( q < \hat{q}^* \) has an expected payoff of

\[
E(q\Delta v + v_l|q > \hat{q}^*) = \int_{\hat{q}^*}^{1} (q\Delta v + v_l) \frac{g(q)}{1 - G(\hat{q}^*)} dq
\]  

(4.3)

for product A and

\[
E(-q\Delta v + v_h|q < \hat{q}^*) = \int_{0}^{\hat{q}^*} (-q\Delta v + v_h) \frac{g(q)}{G(\hat{q}^*)} dq
\]  

(4.4)

for product B. Due to the symmetry of \( G \), the consumer’s willingness to pay is equal for both products when \( \hat{q}^* = \frac{1}{2} \). When \( \hat{q}^* < \frac{1}{2} \), product A is recommended sometimes when product B is more likely to match the consumer’s needs. This implies that the average probability that product B will lead to match when it is recommended is higher than when \( \hat{q}^* = \frac{1}{2} \). Hence, when the expected cutoff \( \hat{q}^* \) decreases, the consumer’s willingness to pay for product A decreases, while his willingness to pay for product B increases.

We follow Inderst and Ottaviani (2012c) as we distinguish between naive and wary consumers.\(^{16}\) Naive consumers do not take into account that advice is possibly biased. They expect unbiased advice, i.e., \( \hat{q}^* = \frac{1}{2} \). This assumption can be motivated by the empirical observation that a majority of people who buy financial products with advice expects advice to be unbiased.\(^ {17}\) When consumers instead are wary, they are aware of the incentives of the advisor. Although wary consumers cannot observe actual commissions when they are not disclosed, they form rational beliefs about commissions and these beliefs are correct in equilibrium (Inderst and Ottaviani, 2012a,c). Consumer awareness implies that the cutoff expected by the consumers, \( \hat{q}^* \), in equilibrium is equal to the cutoff chosen by the advisor, i.e., \( \hat{q}^* = q^* \).

Disclosure is assumed to be an eye-opener (Inderst and Ottaviani, 2012a), i.e., to make consumers wary.\(^ {18}\) In contrast to wary consumers with undisclosed commissions, consumers with disclosed commissions observe actual commissions can respond when commissions deviate from their beliefs. Hence, the best response

---

\(^{16}\)Inderst (2015) considers only naive consumers, Inderst and Ottaviani (2012a) consider only wary consumers. We combine these two approaches in our analysis. Inderst and Ottaviani (2012c) consider both wary and naive consumers, but in a different framework.

\(^{17}\)See section 4.1 and the discussions in Inderst and Ottaviani (2012c) and Inderst and Ottaviani (2012b).

\(^{18}\)If consumers stayed naive despite disclosure, we could analyse the case as if commissions were not disclosed.
functions of the firms with respect to commissions are different than in the case without disclosure.

**Prices and commissions.** At stages 1 and 2, firms independently and simultaneously set commissions and prices, respectively. At stage 2, firms use the consumer’s expectation about the advisor’s cutoff in order to set prices equal to his willingness to pay. That is, they use (4.3) and (4.4) in order to set

\[ p_A(\hat{q}^*) = E(q\Delta v + v_l|q > \hat{q}^*), \]  

and

\[ p_B(\hat{q}^*) = E(-q\Delta v + v_h|q < \hat{q}^*), \]  

respectively. With naive consumers \( \hat{q}^* = 1/2 \), and because of symmetry we have \( p_B(\hat{q}^* = 1/2) = p_A(\hat{q}^* = 1/2) \). The equilibrium price \( p^* \) with wary consumers is hence given by:

\[
p_B(\hat{q}^* = 1/2) = \int_0^{1/2} (-q\Delta v + v_h)\frac{g(q)}{G(1/2)}dq = \int_0^{1/2} (-q\Delta v + v_h)\frac{1}{G(1/2)}dq = \left[ (-\frac{1}{2}q^2\Delta v + qv_h) \cdot 2 \right]^{1/2}_0 = -\frac{1}{4}\Delta v + v_h + \frac{1}{4}v_l = p^* \text{ and} \]

\[
p_A(\hat{q}^* = 1/2) = \int_{1/2}^1 (q\Delta v + v_l)\frac{g(q)}{1 - G(1/2)}dq = \frac{3}{4}v_h + \frac{1}{4}v_l = p^*. \]

Wary consumers in contrast are willing to pay a lower price for product A, and a larger price for product B. Given an *expected* cutoff \( \hat{q}^* \), prices for wary consumers are given by

\[ p_A(\hat{q}^*) = v_l + \Delta v\frac{1}{2}(1 + \hat{q}^*) \]  

and

\[ p_B(\hat{q}^*) = v_h - \Delta v\frac{1}{2}\hat{q}^*, \]  

respectively. The derivation can be found in appendix 4.9.2.

Commissions at stage 1 are determined as mutually best responses. The expected profit of a firm is given by the price it charges, minus production costs, minus commissions paid to the advisor, multiplied with the probability that the
4.3. Theoretical framework

firm’s product is recommended (and sold). The respective profits are given by

\[
\begin{align*}
\pi_A &= (p_A(\hat{q}^*) - c_A - f_A)[1 - G(q^*)], \\
\pi_B &= (p_B(\hat{q}^*) - c_B - f_B)G(q^*),
\end{align*}
\]

(4.10)

where \( p_B(\hat{q}^* = 1/2) = p_A(\hat{q}^* = 1/2) = \frac{3}{4}v_h + \frac{1}{4}v_l \) when consumers are naive.

4.3.2 Equilibria in the experimentally tested market settings

In the experiment, we will consider four conditions as implementations of four theoretical settings: two settings with undisclosed commissions with naive and wary consumers, respectively; and two settings with disclosed commissions and undisclosed commissions with fines for the advisor, respectively. We consider a situation in which disclosure theoretically improves welfare compared to the setting with undisclosed commissions and naive consumers.

Undisclosed commissions, naive consumers, without fines (UN) / with fines (UNF). By calculating best responses from (4.10) with the prices in (4.7) and plugging the best responses into (4.1) we obtain the equilibrium cutoff:

\[
q^* = \frac{1}{2} + \frac{c_A - c_B}{6(w + l)}.
\]

(4.11)

We denote the equilibrium cutoff in the setting (UN) with no fines, \( l = 0 \), by \( q_{UN}^* \); and we denote the equilibrium cutoff in the setting (UNF) with fines by \( q_{UNF}^* \). Clearly, since \( c_B > c_A \), \( q^* < \frac{1}{2} \) in both settings. The setting \( UN \) serves as our experimental benchmark. It represents a typical situation in real-world market for advice with no disclosure requirements. We want to design our experiment such that market welfare can be improved by disclosure compared to this benchmark. Therefore, we require that the first-best cutoff \( q_{FB} \) in (4.2) is larger than \( q_{UN}^* \) in (4.11), i.e., \( q_{FB} > q_{UN}^* \), which is the case if \( \Delta v > 2w \). Hence, we have a case

\( 19 \)Throughout the chapter we refer to experimental treatments as experimental conditions, because in the credence goods literature the term treatment is often used to denote the expert’s actions (Dulleck and Kerschbamer, 2006).

\( 20 \)This setting does not receive much attention in Inderst and Ottaviani (2012a) as the authors focus on the possibly bad consequences of disclosure.

\( 21 \)As shown by Inderst and Ottaviani (2012a), disclosing commissions always increases the cutoff and leads to a cutoff above the first-best cutoff (Inderst and Ottaviani, 2012a). If we did not assume \( \Delta v > 2w \), the benchmark cutoff is already above the first-best cutoff which would mean that disclosure surely reduces welfare.
with overprovision of good A compared to the first-best allocation. The condition \( \Delta v > 2w \) can be interpreted such that the advisor can be influenced relatively easy by commission payments.\(^\text{22}\) Intuitively, due to its low costs, firm A has an incentive to increase commissions excessively. To illustrate this, equilibrium commissions of firms A and B, respectively, are given by (derivation in appendix 4.9.1)

\[
 f^*_A,j = \frac{3}{4} v_h + \frac{1}{4} v_l - w - l - \frac{2}{3} c_A - \frac{1}{3} c_B, \\
 f^*_B,j = \frac{3}{4} v_h + \frac{1}{4} v_l - w - l - \frac{2}{3} c_B - \frac{1}{3} c_A. 
\]

(C.12)

Clearly, since \( c_B > c_A \), \( f^*_A,j > f^*_B,j \) for both \( j \). The difference between both commissions determines the advisor’s cutoff. With naive consumers, an increased market share does not force firms to adjust prices, because consumers have a constant willingness to pay based on false expectations. Hence, firms have incentives to increase commissions. Due to its lower marginal costs, firm A can afford to pay a higher commission than firm B. Further, \( f^*_i,\text{UN} > f^*_i,\text{UNF} \) for \( i \in \{A,B\} \). This shows that with fines, firms proportionally pay lower commissions in equilibrium. The reason is that fines make the advisor less responsive to commissions and firms therefore have lower marginal gains from increasing commissions.

With and without fines, naive consumers realize negative payoffs in expectation (they are “fooled” (Inderst and Ottaviani, 2012c)), because they pay too high prices for product A, given that it is recommended so often. This effect is only partially countered by the low price for product B, because products B’s market share is lower.

We implement this setting by an experimental condition in which consumers are not informed about stage 1 of the stage game, i.e., about the possibility of commission payments. Therefore, consumers are uninformed about possible advisor incentives to give biased advice. Furthermore, commissions are not disclosed. The settings with undisclosed commissions are potentially closest to many real world

\(^\text{22}\)This seems to be realistic, as it implies that the consumer values a match at least twice as much as the advisor. In markets with financial advice consumers usually put up high stakes. While for a consumer an investment outcome can change all financially related determinants of wellbeing for the better or the worse, for an advisor who draws his income from recommendations to numerous clients, the impact of one wrong advice is smaller. A similar argument could be made for health care markets.
markets for advice as conflicts of interest\textsuperscript{23} are typically not disclosed in the absence of regulation.\textsuperscript{24}

**Undisclosed commissions, wary consumers, without fines (UW) / with fines (UWF).** Wary consumers correctly anticipate that advice is biased and adjust their willingness to pay by forming beliefs about advisor behaviour that are correct in equilibrium. In our setting with \( c_A < c_B \), firm A has incentives to pay a larger commission than firm B. Accordingly, consumers are willing to pay a lower price for product A, and a larger price for product B. Using the prices (4.8) and (4.9) in the calculation of the best responses, we obtain the equilibrium cutoff

\[ q^* = \frac{1}{2} + \frac{c_A - c_B}{6(w + l) + \Delta v}. \]  

The derivation can be found in appendix 4.9.2. We denote the equilibrium cutoff in the setting (UW) with no fines, \( l = 0 \), by \( q^*_{UW} \); and we denote the equilibrium cutoff in the setting (UWF) with fines by \( q^*_{UWF} \). We clearly see that the cutoffs with wary consumers are higher than with naive consumers, \( q^*_{UN} < q^*_{UW} \) and \( q^*_{UNF} < q^*_{UWF} \). Although consumer cannot observe actual commissions, they correctly anticipate the incentives of all players and that advice is biased in equilibrium. Contrary to naive consumers, wary consumers cannot be fooled systematically. Yet, commissions are still relatively high in equilibrium, as firms have incentives to increase commission payments.

We experimentally implement this setting by an experimental condition in which consumers are informed about stage 1 of the stage-game, i.e., about the possibility of commission payments, but where commissions are not disclosed to consumers. We will implement a condition without fines, i.e., \( l = 0 \) and one condition with fines, i.e., \( l > 0 \).

**Disclosed commissions, wary consumers, no fines (D).** With disclosure, consumers adjust their willingness to pay to the advisor cutoff they can expect given actual commissions and \( w + l \). Firms hence have to take into account that higher commissions directly reduce the consumers’ willingness to pay. Disclosure thus has

\textsuperscript{23}The difference in commissions determines the incentives of the advisor. That is why we use the term conflict of interest synonymously to the term difference in commissions of the two products.

\textsuperscript{24}An example is provided by Sane and Halan (2017, forthcoming). As argued by Inderst and Ottaviani (2012b) even if voluntary disclosure in the absence of policy intervention could increase firm profits, commitment problems often prevent such agreements from reaching complete disclosures.
a restraining effect on commissions. Without disclosure, such a feedback process is absent as a wary consumer has wrong beliefs about the cutoff when one firm deviates for instance by increasing its commission\(^{25}\). This leads to lower commissions with disclosure than without. Moreover, as shown by Inderst and Ottaviani (2012a), the advisor-cutoff always increases with disclosure, because the feedback process through prices has a more disciplining effect in the commission setting of the cost-efficient firm \(A\). We will present the results with disclosure here while a detailed derivation can be found in appendix 4.9.3.

Prices are now given by (4.8) and (4.9) with \(q^* = q^*\), because consumers can calculate the cutoff in (4.1) using the actual observed commissions \(f_A\) and \(f_B\). The respective expected profits of the firms are given by (we set \(l = 0\) and ignore it)

\[
\begin{align*}
\pi_A &= (p_A(q^*(f_A, f_B)) - c_A - f_A^D) [1 - G(q^*)] \\
\pi_B &= (p_B(q^*(f_A, f_B)) - c_B - f_B^D) [G(q^*)].
\end{align*}
\]

(4.14)

The equilibrium cutoff is then given by

\[
q^*_D = \frac{1}{2} + \frac{c_A - c_B}{6w + 2\Delta v}.
\]

(4.15)

The cutoff is unambiguously larger than \(q^*_UNF(l_D) = q^*_D\), where \(l_D\) denotes the resulting fine. We obtain

\[
6(w + l_D) = 6w + 2\Delta v \Rightarrow l_D = \frac{1}{3} \Delta v.
\]

(4.16)

A fine of this size adds considerable weight on the advisor’s incentives. It is increasing in \(v\), the value the consumer attaches to a match. That is, the further away advice initially departs from the equilibrium with disclosure, the higher the incentive-correcting fine is. The introduction of the fee unambiguously increases

\(^{25}\)A more detailed intuitive explanation hereof see Inderst and Ottaviani (2012b), p.499-500.
4.3. Theoretical framework

consumer welfare as the new cutoff is closer to the cutoff \( q^* = \frac{1}{2} \), the preferred cutoff for the consumer. The expected total of fines collected by the regulator, denoted \( F \), for any \( q^* \) is given by \( F = \int_q^{q^*} (1 - q) l_D dq + \int_0^q q l_D dq = q^2 - q^* + \frac{1}{2} \). We plug in \( q_{UNF}^* \) or \( q_{UWF}^* \) in order to obtain the fines collected by the regulator in our settings with naive and wary consumers, respectively. Welfare in the setting with fines and naive consumers is equal to welfare with disclosure.

Although both policy interventions are designed to be similar, they work through different channels. Theory predicts disclosure to start a feedback process: because consumers immediately adjust their willingness to pay if a firm raises commissions, firms have to adjust prices accordingly. Their marginal gains from raising commissions is therefore lower than without disclosure. This leads to lower commissions in equilibrium, and to a smaller difference in commissions, as the incentives to cut commissions are disproportionally strong for firm A. This in turn results in less biased advice. Fines work by directly making the advisor less responsive to differences in commissions. The results with fines, however, do not differ much between the settings with wary or naive consumers, because even when consumers are naive, firms are constrained in their commission setting. In table 4.1 we present an overview of the derived equilibrium cutoffs.

| TABLE 4.1 |
|---|---|---|
| **Equilibrium cutoff across settings.** | **Consumers** | **Denotation** | **Cutoff** |
| **Undisclosed commissions** | naive | (UN) | \( \frac{1}{2} + \frac{c_A - c_B}{6w} \) |
|  | wary | (UW) | \( \frac{1}{2} + \frac{c_A - c_B}{6w + 2\Delta v} \) |
| **Disclosed commissions** | wary | (D) | \( \frac{1}{2} + \frac{c_A - c_B}{6w + 2\Delta v} \) |
| **Fines and undisclosed commissions** | naive | (UNF) | \( \frac{1}{2} + \frac{c_A - c_B}{6(w+l)} \) |
|  | wary  | (UWF) | \( \frac{1}{2} + \frac{c_A - c_B}{6(w+l) + \Delta v} \) |
4.4 Experiment

4.4.1 Experimental design

We take the model from section 4.3 to the lab. We implement four conditions, two markets with undisclosed commissions with uninformed and informed consumers, respectively and two regulated markets with disclosure of commissions and fines for the advisor, respectively as shown in figure 4.2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Commissions</th>
<th>Information</th>
<th>Fines</th>
<th>Prediction</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNDIS-U</td>
<td>undisclosed</td>
<td>no</td>
<td>no</td>
<td>UN</td>
<td>72</td>
</tr>
<tr>
<td>UNDIS-I</td>
<td>undisclosed</td>
<td>yes</td>
<td>no</td>
<td>UW</td>
<td>96</td>
</tr>
<tr>
<td>DISCL</td>
<td>disclosed</td>
<td>yes</td>
<td>no</td>
<td>D</td>
<td>96</td>
</tr>
<tr>
<td>FINES</td>
<td>undisclosed</td>
<td>yes</td>
<td>yes</td>
<td>UWF</td>
<td>96</td>
</tr>
</tbody>
</table>

First, there is condition UNDIS-U, a market with undisclosed commissions in which consumers are not informed about the first stage of the game in which commission are determined. We use theoretical predictions of the setting with naive consumers (UN) to predict the outcomes for this condition. This seems reasonable as in the instructions consumer participants receive no clues about the possibility of commissions in the experiment. Second, there is condition UNDIS-I, a market with undisclosed commissions in which consumers are informed about stage 1 of the game. Here, we use the theoretical predictions of the setting with wary consumers (UW) to predict the outcomes for this condition. Admittedly, it is not granted that getting informed about possible commission payments makes consumers wary. We act on this assumption, however, as the possibility of commissions is salient from the experimental instructions and procedures. We further use control questions in order to make sure that subjects understand the consequences of commissions on advisor payoffs and firm incentives. Third, there is condition DISCL with disclosure of commissions in which consumers observe commissions. Fourth, there is condition FINES in which commissions are not disclosed, but in which fines for ex-post wrong advice are raised. For the setting with fines we use the theoretical predictions with wary consumers (UWF) for the prediction of outcomes, because consumer subjects are receive the same information as in UNDIS-I. The experiment was framed neutrally.

The experiment is designed such that UNDIS-U performs worst with respect to the quality of advice, consumer and market welfare. In UNDIS-I, DISCL and
FINES, the quality of advice, consumer and market welfare are predicted to improve compared to UNDIS-U. The predicted quality of advice in both conditions with regulation, DISCL to FINES, is higher than in UNDIS-I. The predicted quality of advice and total welfare for DISCL and FINES are almost equal, but different mechanisms are predicted to lead to these results. While with disclosure, we expect consumers to adjust their willingness to pay with regard to the observed commission payments. Following, firms have lower incentives to raise commissions. Furthermore, incentives for firm A are disproportionally affected, leading to a smaller conflict of interest. With fines, we expect that advisors react less strongly to conflicts of interest - a mechanism which likewise decreases incentives to pay commissions.

We let subjects play the one-shot-game for eight rounds. Subjects are assigned to matching groups with twelve players and one of the four roles – firm A, firm B, advisor or consumer – before the first round. Within each matching group, three independent games with four players each are played in each round. Matching groups and role assignments remain fixed throughout the experiment. Within matching groups, subjects are randomly matched before every round.

To elicit the advisors’ behaviours we employ a strategy method (Selten, 1967). The advisor is asked to reveal his full strategy. It contrasts the direct response method in which the advisor would only be asked to make a choice at one decision node. The strategy method has been used frequently by experimental economists in the last decades. In a survey, Brandts and Charness (2011) compare experimental results of both methods and do not find evidence that they lead to different results. The major advantage of the strategy method is that we obtain more data about the advisors’ behaviours than with the direct response method. Specifically, we can directly determine the advisor’s cutoff. In order to keep the task simple for advisors, we implement monotonic switching as we ask advisors at which cutoff they like to switch from recommending product B to recommending product A rather than asking them to make a decision at each possible probability \( q \) that the consumer needs product A. This method with an enforced single switching point prevents the possible problem that subjects choose more than one switching point and show inconsistent behaviour or randomly express indifference. The approach has been used by Andersen et al. (2006) and Tanaka et al. (2010) in tasks to elicit the risk aversion of subjects. Downsides of this method are first, that it imposes

\[26\text{Many authors refer to matching groups by the term (experimental) markets. We do not use this terminology in this paper in order to avoid confusion with our four settings which we refer to as markets.}\]

\[27\text{We use the term cutoff for both the theoretical and the experimental cutoff as both cutoffs represent the strategy choice of the advisor.}\]
strict monotonicity on revealed preferences and enforces transitivity. Second, it may bias the results, because subjects who would have made inconsistent choices with the standard method are now forced to make choices (Charness et al., 2013).\footnote{A practical reason for implementing monotonic switching is that we wanted to play the game as many rounds as possible, given time and cost restrictions. Had we left subjects make a decision for every possible scenario the advisors’ decisions would have taken considerably longer with negative effects on the experiment costs and on the subjects in the other roles who would have had longer waiting times.}

With respect to the probability $q$, we restrict the experimental world to eleven possible realizations of $q$ (“scenarios”) in ten percent-increments from zero to one, 0\%, 10\%, ..., 100\%, representing a discrete version of the signal from the continuous model presented in section 4.3. All scenarios are equally likely and this is communicated to all subjects. Advisors decide up to which scenario they recommend the product offered by firm B and starting from which scenario they recommend the product of firm A. Furthermore, advisors can indicate that they are indifferent between recommending one or the other product at the chosen scenario.\footnote{In case of such a choice, the advisor recommends each product with 50\% probability when this indifference-scenario gets realized, and the product of firm B up to the scenario below the indifference-scenario, and the product of firm A in any scenario above the indifference-scenario. We implement the possibility to choose an indifference-scenario in order to make sure that consumers could reasonably expect unbiased advice.} The advisor hence can choose among 23 possible strategies, including the two extreme decisions to recommend one or the other product in all scenarios. We present them in table 4.3. The complexity of the advisor choice in our experiment is high, but comparable to other experiments which make use of the strategy method. For instance, in a public goods game experiment by Fischbacher and Gächter (2010) participants need to make one choice for each of 21 possible scenarios.

We will from now on use the term quality of advice synonymously with the empirical expected share of correct recommendations which is calculated as the mean of the values of the column “Quality” in table 4.3. In the theoretical model, the link between the advisor’s cutoff and the quality of advice is unambiguous, because there is no heterogeneity in advisor-choices.\footnote{The quality of advice is not a linear, but a weakly concave function of the empirical cutoff. For illustration, two unbiased cutoff-choices at 0.5 have the same mean as two extreme cutoff-choices of zero and one. The expected recommendation quality is however higher with the two unbiased cutoff choices compared to the two extreme cutoff choices (77.3\% vs. 50.0\%).}

The game played is analogous to the game presented in section 4.3. At the first stage, the two firm players in every group of four players independently set commissions. At the second stage, firms observe all commissions in their group of four and independently choose product prices. At the third stage, the advisor in each group of four is informed about commissions and prices. Using the strategy
### TABLE 4.3
The 23 possible strategy choices of the advisor.

<table>
<thead>
<tr>
<th>#</th>
<th>Scenario</th>
<th>Share B</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%</td>
<td>0 0.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B/A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>B/A</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>B</td>
<td>B/A</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>13</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>18</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>19</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>21</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>22</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>23</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Note: The 23 possible recommendation strategies of the advisor; for instance, with strategy #5, product B is recommended when scenarios 0% or 10% are realized, and product A in all other scenarios. B/A indicates that both products are recommended with equal probability. The column “Share B” shows the expected recommendation probability of product B, the analogue to the cutoff in the continuous model. The column “Quality” shows the probability that advice will match the consumers’ needs.

Method with enforced monotonic switching, the advisor reveals his recommendation strategy. Before the fourth stage, one scenario is randomly realized by the computer and the product the consumer requires is randomly determined based on the probabilities from the realized scenario. At the fourth stage, the consumer is informed about the prices of the two products and the recommendation of the advisor which is calculated depending on the realized scenario. In the disclosure condition, the consumer additionally is informed about the commission payments from stage 1. The consumer then chooses to either buy the recommended product or no product at all. When the consumer purchases no product, payoffs of all players are zero. After all decisions have been made, payoffs are calculated and displayed.
4.4.2 Parametrization & predictions

We used the experimental currency ECU with exchange rate 10 ECU = 1 CHF. The parameters of the game are displayed in table 4.4. Firms could choose commissions and prices in 10-ECU increments between 0 and 220. The predictions of the discrete model are presented in table 4.5. They were obtained with numerical calculations (available on request).

<table>
<thead>
<tr>
<th>Role</th>
<th>Parameter</th>
<th>Value (in ECU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm A</td>
<td>Costs</td>
<td>(c_A) 20</td>
</tr>
<tr>
<td>Firm B</td>
<td></td>
<td>(c_B) 120</td>
</tr>
<tr>
<td>Advisor</td>
<td>Default Fine</td>
<td>(w) 40</td>
</tr>
<tr>
<td>Advisor</td>
<td>Fine (only in FINES)</td>
<td>(l) 60</td>
</tr>
<tr>
<td>Consumer</td>
<td>Payoff from match</td>
<td>(v_h) 220</td>
</tr>
<tr>
<td>Consumer</td>
<td>Payoff from no match</td>
<td>(v_l) 70</td>
</tr>
</tbody>
</table>

We expect the lowest advisor cutoff in UNDIS-U (1/22 - referring to decision no. 2 in table 4.3). When consumers are informed about commissions in UNDIS-I we expect, assuming that information makes consumers wary, more accurate advice realized in a higher cutoff (6/22 - referring to choice no. 7 in table 4.3). Advice is even less biased with disclosure (8/22 - referring to decision no. 9 in table 4.3) and with fines (9/22 - referring to decision no. 10 in table 4.3), assuming that consumers are wary in fines. Even if consumers are not wary in FINES, the predicted cutoff is as in DISCL.

The predictions for UNDIS-U are driven by the wrong expectations of consumers. Firm A can demand a higher price than if consumers’ expectations were correct and hence profits from raising commissions to 90 ECU, while firm B suffers from a low price for product B and is thus reluctant to increase commissions likewise. The difference in commissions amounts to 40, resulting in a large conflict of interest for the advisor which in turn results in a low cutoff. Since consumers are systematically fooled, their payoff is expected to be negative. Overall, the low accuracy of advice leads to a relatively low market welfare of 109.1 ECU. In UNDIS-I, firms need to adjust prices to consumers’ expectations, resulting in a price wedge between products. This puts some restraint on firm A’s incentives to pay commissions, resulting in a smaller conflict of interest for the advisor. Consumers and firm B are better off and market welfare increased to 119.2.
### TABLE 4.5
 Predictions (in ECU)

<table>
<thead>
<tr>
<th>Market</th>
<th>Undisclosed Commissions</th>
<th>Disclosed Commissions</th>
<th>Fines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer behaviour</td>
<td>naive</td>
<td>wary</td>
<td>wary</td>
</tr>
<tr>
<td>Condition</td>
<td>UNDIS-U</td>
<td>UNDIS-I</td>
<td>DISCL</td>
</tr>
<tr>
<td>Firm A</td>
<td>90</td>
<td>70</td>
<td>20</td>
</tr>
<tr>
<td>Firm B</td>
<td>50</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Firm A - Firm B</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Prices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product A</td>
<td>180</td>
<td>160</td>
<td>170</td>
</tr>
<tr>
<td>Product B</td>
<td>180</td>
<td>200</td>
<td>190</td>
</tr>
<tr>
<td>Product A - Product B</td>
<td>0</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>Advisor Behaviour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutoff (Share Product B)</td>
<td>$\frac{1}{22}$</td>
<td>$\frac{6}{22}$</td>
<td>$\frac{8}{22}$</td>
</tr>
<tr>
<td>Expected share correct Recommendations</td>
<td>0.545</td>
<td>0.718</td>
<td>0.755</td>
</tr>
<tr>
<td>Welfare</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>-28.2</td>
<td>6.8</td>
<td>5.9</td>
</tr>
<tr>
<td>Advisor</td>
<td>70.00</td>
<td>53.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Firm A</td>
<td>66.8</td>
<td>50.9</td>
<td>82.7</td>
</tr>
<tr>
<td>Firm B</td>
<td>0.5</td>
<td>8.2</td>
<td>21.8</td>
</tr>
<tr>
<td>Fines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Welfare</strong></td>
<td><strong>109.1</strong></td>
<td><strong>119.2</strong></td>
<td><strong>117.0</strong></td>
</tr>
</tbody>
</table>

Note: Predictions of the discrete model in ECU with parametrization as presented in table 4.4.
In both regulated markets, incentives for firms to pay commissions are strongly reduced, but for different reasons. With disclosure, consumers adjust their willingness to pay to actually observed commissions to which firms need to react by adjusting prices. This mechanism leads to low commission levels and low conflicts of interest in equilibrium. Disclosure works through a feedback process between consumers and firms, while the advisor behaves as in the case with undisclosed commissions. Fines in contrast aim at making the advisor less responsive to differences in commissions. Firms hence profit less from increasing commissions compared to the other conditions. Consumer expectation, however, have little influence with fines as is shown in the small differences between fines with naive and wary consumers, respectively. The advisor realizes a negative payoff in this setting, because commissions are low and fines are large. We summarize our predictions in the following hypotheses.

**Hypothesis 4.1. Undisclosed commissions with uninformed consumers.** In UNDIS-U, a large conflict of interest and correspondingly highly biased advice occur. The expected share of correct recommendations is lower than in all other conditions. There is no price difference between the two products. Consumer welfare is negative due to the combination of biased advice and high prices for product A. Market welfare is the lowest among all conditions.

**Hypothesis 4.2. Policy interventions.** In both DISCL and FINES we observe equally biased advice which is more accurate than in both conditions with undisclosed commissions. Commission levels are lower than in the markets with undisclosed commissions. For the same cutoff in DISCL and FINES, we observe a larger conflict of interest in FINES. Due to low commissions, both firms achieve a higher welfare than with undisclosed commissions, while the advisor is worse off. Total welfare does not differ significantly between the two policy conditions.

### 4.4.3 Experimental protocol

The experiment was programmed with z-tree (Fischbacher, 2007). All sessions were conducted in the DeScil (Decision Science Laboratory) at ETH Zurich, Switzerland. The experiment was organized and recruited with the software hroot (Bock et al., 2014). In total we held fourteen sessions in May and June 2017 with 12, 24 or 36 subjects per session and 360 subjects in total. Subjects were students of different departments from the universities in Zurich. See table 4.8 in the appendix. Four sessions were allocated to each

---

31 See table 4.8 in the appendix.
condition but *UNDIS-U*, for which two sessions were conducted. In total we had eight matching groups in *UNDIS-I*, *DISCL* and *FINES* and six matching groups in *UNDIS-U* with twelve subjects per matching group. All sessions with *UNDIS-I* and *FINES* had 24 participants, sessions with *DISCL* had 12, 24 or 36 participants; sessions with *UNDIS-U* had 36 participants.

Subjects were seated randomly using numbered cards. After seating was completed, the instructions were read out aloud in all conditions but *UNDIS-U*. In *UNDIS-U*, the instructions regarding the game played were not read out aloud as subjects in the consumer role received instructions in which stage 1 of the game was not mentioned. Before the experiment started, the subjects had to answer eight control questions. The experiment only started after all subjects had successfully answered the control questions.

First, subjects played eight rounds of the (neutrally framed) advice-game with random re-matching within matching groups after each round. After the eight rounds, subjects played a simple level-k game (Nagel, 1995) in their market group of twelve with the aim to elicit their sophistication with respect to strategic thinking. Subjects were asked to guess an integer value between 0 and 100. A winning price of CHF 10 was awarded to the subject with the closest guess to two thirds of the average guessed number. After that we conducted a one-shot dictator game with charity donations (Eckel and Grossman, 1996) in which each subject was asked to distribute an endowment of CHF 10 between their own pocket and a donation to the charity doctors without borders, an international medical humanitarian organisation. In the beginning of the experiment instructions, we announced that there would be a second part of the experiment, but neither the level-k game nor the dictator game were mentioned. This announcement was done to guarantee that subjects could anticipate when the session was over. After the dictator game subjects were asked to fill out a short socio-economic questionnaire. In one question we asked subjects

---

32. The first page of the instructions including information on the matching procedure was read out aloud.
33. Importantly, these subjects received incomplete, but no false information.
34. In *UNDIS-U*, the questions were adjusted for consumer subjects in order to not inform them about stage 1 of the game.
35. With random re-matching and no identifiability, reputation building is hardly possible.
36. On the last page of the introduction, subjects could find an online link (https://www.descil.ethz.ch/projects/1701-advicemarkets) and were told that may they take it home. On the web page run by the DeScil all donation certificates were made available for download.
about their willingness to take risks in the exact wording of Dohmen et al. (2011).\textsuperscript{37} Sessions lasted 80 minutes on average and subjects were paid CHF 41 on average.\textsuperscript{38}

### 4.5 Results

We first state our results concerning our hypotheses and go into further detail in the remainder of the section. Tables 4.6 and 4.7 present aggregate results for the four conditions.

For comparisons of outcomes between conditions we employ non-parametric test using mean values over all games over all periods in one matching group as one independent observation. Hence, our tests are based on six observations in UNREG-N and eight observations in each of the other conditions. For the comparison of conditions we use the standard two-tailed Mann-Whitney-U test (MWU). We report findings to be “significant” when the $p$-value is below 0.05 and “weakly significant” if the $p$-value is between 0.05 and 0.1. In case our theoretical predictions indicate a clear direction of an effect we may also employ one-sided tests. Furthermore we will complement our analysis with the Robust Rank-Order test (RRO) which was developed by Fligner and Policello (1981) as an alternative to the Mann-Whitney-U test.\textsuperscript{39} We will only refer the results of the RRO when it leads to different significance levels than the MWU, which is almost never the case. For the comparison of results with theoretical predictions we use the Wilcoxon Signed-Rank test (WSR).\textsuperscript{40} In some cases we additionally refer to the Sign test (S) in order to discuss results of the WSR.\textsuperscript{41}

\textsuperscript{37}Dohmen et al. (2011) show that this risk measure has a good behavioural validity.

\textsuperscript{38}The average hourly wage was comparatively high. The reason is that laboratory rules required us to make sure that no subject could possibly leave the experiment with a negative payout. Subjects could achieve negative payouts in the advice game, however. Particularly, the consumer payoff in the benchmark treatment is predicted to be negative. We paid a show-up fee of CHF 15 and CHF 15 for answering the control questions, equal to the highest possible loss in the experiment to avoid negative payoffs and to obey with the DeScil rules.

\textsuperscript{39}The RRO does not require that the compared distributions are identical and is thus less sensitive to changes in distributional assumptions (Feltovich, 2005). The test was recently used in market experiments by Minra et al. (2016a) and Huck et al. (2016a). Note that for small sample sizes ($n \leq 12$), critical values must be inferred from tables provided by Fligner and Policello (1981) and Feltovich (2005). The $p$-values of the RRO can only be reported within intervals.

\textsuperscript{40}The signed-rank test was, for instance, recently used by Hennig-Schmidt et al. (2011) and Fischbacher and Föllmi-Heusi (2013) to compare experimental results with predictions.

\textsuperscript{41}In contrast to the WSR, the Sign test does not use information provided by the ranking of absolute differences between the data and the prediction and is therefore less sensitive to outliers. See Rey and Neuhäuser (2011) for an overview on the applications of the WSR and the Sign Test.
## 4.5. Results

### TABLE 4.6

**Results I.**

<table>
<thead>
<tr>
<th></th>
<th>UNDIS-U</th>
<th>UNDIS-I</th>
<th>DISCL</th>
<th>FINES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Commissions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commission A (ECU)</td>
<td>55.42</td>
<td>54.84</td>
<td>60.10</td>
<td>58.39</td>
</tr>
<tr>
<td>(90**)</td>
<td>(70**)</td>
<td>(20**)</td>
<td>(30**/20**)</td>
<td></td>
</tr>
<tr>
<td>Commission B (ECU)</td>
<td>43.12(c)</td>
<td>43.33(e)</td>
<td>40.89</td>
<td>33.85(c,e)</td>
</tr>
<tr>
<td>(50)</td>
<td>(50*)</td>
<td>(10**)</td>
<td>(0**/0**)</td>
<td></td>
</tr>
<tr>
<td>Commission A - Commission B</td>
<td>12.29(B,c)</td>
<td>11.51(e)</td>
<td>19.21(B)</td>
<td>24.53(c,e)</td>
</tr>
<tr>
<td>(40**)</td>
<td>(20)</td>
<td>(10)</td>
<td>(30/20**)</td>
<td></td>
</tr>
<tr>
<td><strong>Prices</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price A (ECU)</td>
<td>141.0</td>
<td>142.9</td>
<td>138.7</td>
<td>144.3</td>
</tr>
<tr>
<td>(180**)</td>
<td>(160**)</td>
<td>(170**)</td>
<td>(180**/170**)</td>
<td></td>
</tr>
<tr>
<td>Price B (ECU)</td>
<td>183.9(a)</td>
<td>166.4(a,D)</td>
<td>194.5(D,F)</td>
<td>183.1(F)</td>
</tr>
<tr>
<td>(180)</td>
<td>(200*)</td>
<td>(190*)</td>
<td>(180/190*)</td>
<td></td>
</tr>
<tr>
<td>Price B - Price A</td>
<td>42.8(A)</td>
<td>23.4(A,D)</td>
<td>55.8(D,F)</td>
<td>38.8(F)</td>
</tr>
<tr>
<td>(0**)</td>
<td>(40**)</td>
<td>(20**)</td>
<td>(0**/20**)</td>
<td></td>
</tr>
<tr>
<td><strong>Advisor behaviour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutoff</td>
<td>0.353</td>
<td>0.370</td>
<td>0.350</td>
<td>0.383</td>
</tr>
<tr>
<td>(0.046**)</td>
<td>(0.273**)</td>
<td>(0.364)</td>
<td>(0.364/0.409)</td>
<td></td>
</tr>
<tr>
<td>Exp. correct recommendations</td>
<td>0.701(a)</td>
<td>0.669(a,E)</td>
<td>0.690</td>
<td>0.707(E)</td>
</tr>
<tr>
<td>(“Quality of advice”)</td>
<td>(0.545**)</td>
<td>(0.718**)</td>
<td>(0.755***)</td>
<td>(0.755**/0.764**)</td>
</tr>
<tr>
<td>Unbiased cutoff choices</td>
<td>0.292(C)</td>
<td>0.354</td>
<td>0.339</td>
<td>0.410(C)</td>
</tr>
<tr>
<td><strong>Trade &amp; matching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accepted recommendations</td>
<td>0.576</td>
<td>0.682</td>
<td>0.688</td>
<td>0.630</td>
</tr>
<tr>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
<td></td>
</tr>
<tr>
<td>Accepted A-recommendations</td>
<td>0.685</td>
<td>0.714</td>
<td>0.805</td>
<td>0.728</td>
</tr>
<tr>
<td>Accepted B-recommendations</td>
<td>0.400(A)</td>
<td>0.610(A,e)</td>
<td>0.478</td>
<td>0.448(e)</td>
</tr>
<tr>
<td>Matching frequency</td>
<td>0.361(b)</td>
<td>0.448</td>
<td>0.490(b)</td>
<td>0.438</td>
</tr>
<tr>
<td>(0.545**)</td>
<td>(0.718**)</td>
<td>(0.755***)</td>
<td>(0.755**/0.764**)</td>
<td></td>
</tr>
<tr>
<td>Matching frequency</td>
<td>0.627</td>
<td>0.656</td>
<td>0.712</td>
<td>0.694</td>
</tr>
<tr>
<td>(conditional on purchase)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matching groups</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Note: means over all matching groups and periods. Predictions in parentheses (FINES: predictions for (naive/wary) consumers). Differences between conditions: two-tailed MWU tests:

- UNDIS-U vs. UNDIS-I (a: \(p < 0.1\), A: \(p < 0.05\))
- UNDIS-I vs. DISCL (d: \(p < 0.1\), D: \(p < 0.05\))
- UNDIS-U vs. DISCL (b: \(p < 0.1\), B: \(p < 0.05\))
- UNDIS-I vs. FINES (c: \(p < 0.1\), E: \(p < 0.05\))
- UNDIS-U vs. FINES (c: \(p < 0.1\), C: \(p < 0.05\))
- DISCL vs. FINES (f: \(p < 0.1\), F: \(p < 0.05\))

Differences between conditions and predictions: Two-tailed WSR tests; ****: \(p < 0.01\), **: \(p < 0.05\), *: \(p < 0.1\).
All tests use means of each matching group over all individuals and periods as one independent observation.
Chapter 4

### TABLE 4.7
Results II.

<table>
<thead>
<tr>
<th>Welfare</th>
<th>UNDIS-U</th>
<th>UNDIS-I</th>
<th>DISCL</th>
<th>FINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm A (ECU)</td>
<td>25.90</td>
<td>29.53</td>
<td>29.22</td>
<td>25.36</td>
</tr>
<tr>
<td></td>
<td>(66.8**)</td>
<td>(50.9**)</td>
<td>(82.7**)</td>
<td>(82.7**/76.8**)</td>
</tr>
<tr>
<td>Firm B (ECU)</td>
<td>0.833a</td>
<td>-5.417a,D,E</td>
<td>4.062D</td>
<td>3.333E</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(8.2**)</td>
<td>(21.8**)</td>
<td>(21.8**/28.6**)</td>
</tr>
<tr>
<td>Advisor (ECU)</td>
<td>23.47</td>
<td>28.13E</td>
<td>29.53F</td>
<td>13.59E,F</td>
</tr>
<tr>
<td></td>
<td>(70.0**)</td>
<td>(53.3**)</td>
<td>(6.6**)</td>
<td>(-5.5**/-11.8**)</td>
</tr>
<tr>
<td>Consumer (ECU)</td>
<td>8.89</td>
<td>20.94</td>
<td>19.90</td>
<td>19.95</td>
</tr>
<tr>
<td></td>
<td>(-28.2**)</td>
<td>(6.8**)</td>
<td>(5.9**)</td>
<td>(3.2**/6.4**)</td>
</tr>
<tr>
<td>Fines (ECU)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.56</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(14.7/14.2)</td>
</tr>
<tr>
<td>∑</td>
<td>59.09</td>
<td>73.20</td>
<td>82.71</td>
<td>73.80</td>
</tr>
<tr>
<td></td>
<td>(109.1**)</td>
<td>(119.2**)</td>
<td>(117.0**)</td>
<td>(117.0**/114.2**)</td>
</tr>
</tbody>
</table>

### Controls

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-K Win-Value</td>
<td>28.9</td>
<td>29.7</td>
<td>30.9</td>
<td>29.2</td>
</tr>
<tr>
<td>Dictator Charity Donation (%)</td>
<td>28.4b</td>
<td>31.6</td>
<td>36.5b</td>
<td>29.0</td>
</tr>
<tr>
<td>Willingness to take Risks (0-10)</td>
<td>4.9</td>
<td>5.1</td>
<td>5.1</td>
<td>4.8</td>
</tr>
<tr>
<td>Gender=Female (%)</td>
<td>58.3</td>
<td>54.7</td>
<td>55.2</td>
<td>53.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matching Groups</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Note: means over all matching groups and periods. Predictions in parentheses (FINES: predictions for (naive/wary) consumers). Differences between conditions: two-tailed MWU tests:

- UNDIS-U vs. UNDIS-I (a: p < 0.1, A: p < 0.05)
- UNDIS-I vs. DISCL (d: p < 0.1, D: p < 0.05)
- UNDIS-U vs. DISCL (b: p < 0.1, B: p < 0.05)
- UNDIS-U vs. FINES (c: p < 0.1, C: p < 0.05)
- UNDIS-U vs. FINES (e: p < 0.1, E: p < 0.05)
- DISCL vs. FINES (f: p < 0.1, F: p < 0.05)

Differences between conditions and predictions: Two-tailed WSR tests; ****: p < 0.01, **: p < 0.05, *: p < 0.1.

All tests use means of each matching group over all individuals and periods as one independent observation.

### Result 4.1. Undisclosed commissions.
In UNDIS-U, firms create conflict of interests in favour of product A and advisors bias advice in the predicted direction. Both effects are, however, considerably and significantly less pronounced than predicted. Conflicts of interest are small, because firm A sets lower commissions than predicted. The advisor cutoff does not differ from the cutoff in other conditions, but the quality of advice is weakly significantly larger in UNDIS-U than in UNDIS-I. Contrary to predictions, the price for product A is significantly lower than the price for product B. Consumer welfare is positive and significantly higher than predicted due the combination of less biased advice and lower prices for product A. In line with predictions, market welfare in UNDIS-U is the lowest among all conditions, because consumers reject advice more often than in other conditions.
4.5. Results

Result 4.2. Policy interventions. In both DISCL and FINES firms’ commissions create conflicts of interest which are significantly larger than in the markets with undisclosed commissions. Advisors bias advice in both conditions in favour of product A. Despite the larger conflicts of interest, advisors do not bias advice stronger than in the conditions with undisclosed commissions, which is contrary to predictions for DISCL and in line with predictions for FINES. In both conditions, firms choose considerably and significantly higher commissions than predicted. Price differences are significantly larger than predicted due to low prices of product A. Market welfare in DISCL is the largest of all conditions, because consumers accept recommendations more often. In both policy conditions advisors have a significantly larger payoff than predicted, while firms are significantly worse off. Consumers have the same payoff in DISCL, FINES and UNDIS-I, which is weakly significantly larger than in UNDIS-U and above the predicted values in every condition.

4.5.1 Experimental conditions

4.5.1.1 UNDIS-U

In UNDIS-U, advice is considerably and significantly less biased than predicted as we observe a cutoff of 0.353 (prediction: 0.046). The difference in commission is only 12.29 ECU$^{42}$ and therefore around 30 percent of the predicted size. It is driven by firm A which pays significantly lower commissions than predicted (55.42 vs. 90), while firm B’s commissions are not significantly lower than predicted (43.12 vs. 50). Judging by theory, the high cutoff appears to be in line with the small conflict of interest. In contrast to theory, consumers only accept 57.6 percent of all recommendations, and more recommendations of product A (68.5%) than of product B (40.0%).

In contrast to the predictions, we observe price difference of more than 40, driven by a low price for product A (141.0 vs. 180) while the price for product B is close to predictions (183.9 vs. 180). Several explanations are possible for this finding. One is that consumers do not expect advice to be unbiased as described in section 4.3 and consequently demand a price wedge between the two products. Consumers’ risk-aversion could then explain why the price wedge does not expand around 180, but around a lower price of around 160. However, using the answers from the risk-questionnaire, we do not find evidence that participants behave differently depending on their degree of risk aversion in a panel regression on consumers’ decision to accept a recommendation for product A (table 4.9 in appendix 4.8).

$^{42}$From now on we suppress the currency notation.
Another explanation is that consumers’ low willingness to pay for product A expresses social preferences such as inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), because firm A has a larger profit potential than firm B due to its lower production costs. This conjecture about social preferences is supported by panel regressions on the consumers’ acceptance of product A recommendations (table 4.9). The regressions show that in UNDIS-U, consumers with higher charity donations in the dictator game are significantly less likely to accept a product A recommendation, given the price of product A and other controls. With respect to firm A’s behaviour, a panel regression on firm A’s commission choices reveals that a higher charity donation in the dictator game is associated with lower commission payments (table 4.11). Together with the results from table 4.9 this suggests that social preferences may play a role in the behaviour of both consumers and firms. Moreover, we do not observe a time trend for prices of product A, which could be explained by the fact that firms anticipate consumer responses already in the start of the experiment. This would be in line with findings from the literature on social preferences (Fehr and Schmidt, 2006). The regressions on the consumer acceptance in table 4.9 do not show a significant effect of dictator donations in the other conditions. A possible explanation is that in UNDIS-U, consumers overestimate firm’s profits due to the fact that they are not aware of commission payments.

Given that advice is less biased and product A is cheaper than predicted, consumer welfare is positive and therefore not negative as expected. Correspondingly, firm A’s welfare is significantly lower than predicted. Advisors on average earn 23.47 which is significantly less than the prediction of 70. The reasons are that commissions of firm A are lower than predicted, product A gets recommended less often than predicted and consumers reject recommendations in more than forty percent of the cases.

4.5.1.2 UNDIS-I

In UNDIS-I, advisors choose a cutoff around ten percentage points above the predicted value (0.370 vs. 0.273), in line with a low difference of commissions of 11.51. The quality of advice is, however, around five percentage points below predictions (0.669 vs. 0.718). One reason for this result is the high share of extreme 0%-cutoff choices which reaches 22.4 %, compared to the share of 0.118 in UNDIS-U.

We do not find similar effects for firm B. However, firm B has a much lower profit potential than firm A due to its high production costs. A possible analogy is given by Güth et al. (2003) who show that social preferences for equality become weaker (compared to efficiency preferences) when subjects’ own payoffs are at stake.
and shares of around ten percent in *DISCL* and *FINES*. The difference to the other conditions is not significant, however, as the large share in *UNDIS-I* is mainly due to two matching groups (see figure 4.6b below).

Commission levels are very similar to *UNDIS-U*, but the commissions of firm A are closer to the predicted level than in *UNDIS-I* (54.84 vs. 70). We observe a price difference between both products where product A is cheaper than product B, but both products have lower prices than predicted (product A: 142.9 vs. 160; product B: 166.4 vs. 200). A closer look at individual data reveals that the low average price of product B compared to the other conditions is driven by only three subjects who, on average, choose price-commission differences below production costs. Excluding these three players in *UNDIS-I* corrects the average prices of firm B to 181.7 (compared to 166.4 before). Consumer payoff is above predicted levels at 20.94 and the highest among all conditions, because of the combination of low commissions of firm A and low prices for both products.

Panel regressions on the consumers’ decision to accept a recommendation for product A show that social preferences, as measured by the charity donations from the dictator game, are not significantly correlated with the consumers’ decision (table 4.9). One possible explanation is when consumers in *UNDIS-I* correctly anticipate that firms pay commissions, they should expect lower firm profits and therefore less unequal payoffs in *UNDIS-I* than in *UNDIS-U*. Consumers’ awareness about possible commission payments may explain the price wedge, but it cannot explain why we observe lower prices than predicted for both products. Regressions on firms’ price setting behaviour show that dictator donations are significantly negatively correlated with the price of product A, and significantly positively correlated with the price for product B (tables 4.13 and tables 4.14), suggesting that social preferences may help in explaining the price wedge between the two products.\(^{45}\)

\(^{44}\)This choice surely leads to a non-positive and possibly to a negative payoff. Out of the 720 experimental games played, we observe that firms B choose such a price-commission difference of less than production costs in 57 (7.92\%) cases. Only six out of the 90 firm B players choose price-commission combinations below production costs in more than two periods. One player makes such a choice in seven periods and three players consistently choose price-commission combinations below production costs in all eight periods. These three players are all in condition *UNDIS-I* and the player with seven choices is in *UNDIS-U*.

\(^{45}\)Moreover, regressions show that firm B’s commissions are negatively correlated with giving in the dictator game (table 4.12). A possible explanation for this result is that players with large donations belief that advisors also have social preferences and will not punish firm B for lower commissions, while players with lower donations do not belief so. While we cannot test this hypothesis, we may find an analogy in the finding from the literature on cooperation in public good games which has shown that subjects’ donations and beliefs about the donations of others are positively correlated (Fischbacher and Gächter, 2010).
4.5.1.3 DISCL and FINES

In both DISCL and FINES, the advisor cutoff is close to predictions and therefore does not differ significantly between the two conditions. The average cutoff in DISCL is 0.350 (prediction: 0.364) and 0.383 in FINES (0.364/0.409). With respect to the quality of advice, both DISCL and FINES perform significantly worse than predicted due to the heterogeneity in cutoff choices (see figure 4.4) (0.690 vs. 0.755 in DISCL and 0.707 vs. 0.755/0.764 in FINES).

DISCL. In DISCL, the difference in commissions amounts to 19.21, which is larger than the prediction of 10.\(^{46}\) This result is interesting as the larger conflict of interest – as compared to the two conditions with undisclosed commissions (DISCL vs. UNDIS-U: \(p = 0.0452\) (MWU), DISCL vs. UNDIS-I: \(p = 0.1889\) (MWU)) – does not lead to more biased advice, while theory predicts that it should. This effect also contradicts the psychological trait of moral licensing (Loewenstein et al., 2011) on the advisor part which should lead to more biased advice with disclosure.\(^{47}\) Behavioural traits like guilt aversion Charness and Dufwenberg (2006)\(^{48}\) may help to explain the result, as disclosure reveals the incentives for giving biased advice. Social preferences consistent with giving in dictator games like inequity aversion or altruism are less likely to play a role here when judged by the regression results from table 4.15 as the regressions show that contributions in the dictator game are positively correlated with the advisor’s cutoff only in UNDIS-I.

Commission levels are higher than expected for both firms. Hence, disclosure does not have the predicted effect of lowering commission levels. This indicates that the feedback process from the consumers’ willingness to pay over prices to commissions described by Inderst and Ottaviani (2012b) does not work to full extent in our experiment.

While the price for product B is only slightly larger than predicted (194.5 vs. 190), the price for product A is significantly below the predicted level (138.7 vs. 170), opening up a price wedge of 55.8 (prediction: 20). The question of why we observe such low prices for product A is not easily answered. One possible explanation is

\(^{46}\)WSR: \(p = 0.164; S: p = 0.070\). The WSR is not significant although only one matching group is below the prediction, because the difference in commissions in this matching group is \(-15.83\) which greatly differs from the prediction and thus leads to a high rank in the WSR (see figure 4.1). The Sign-test consequently indicates a (weak) significant difference.

\(^{47}\)Possibly the increased conflict of interest represents some form of moral licensing on part of firm A as the commissions of firm A in DISCL with 60.10 are larger than in the other condition, though not significantly so.

\(^{48}\)Lying aversion (Ellingsen and Johannesson, 2004; Gneezy, 2005; Gneezy et al., 2013) is another possibility, but giving biased advice is not necessarily lying.
4.5. Results

FIGURE 4.1
Difference between commissions across conditions and matching groups.

Note: matching groups (black); mean (red); prediction (green).

that firm A subjects expect consumers to expect a lower cutoff and therefore have a lower willingness to pay than consumers actually do.

In terms of market welfare, disclosure achieves the best result of all conditions. The reason is that consumers accept more recommendations than in the other conditions, specifically 80% of product A recommendations and 68.8% of all recommendations. In contrast, the similarly high acceptance rate of 68.2% in UNDIS-I is mainly due to the low prices of product B and a correspondingly higher acceptance rate of product B recommendations. A higher acceptance rate of product B recommendations has a lower welfare effect, because product B is recommended less often than product A.

FINES. Our results from FINES show that the fine has the intended effect of making the advisor less responsive to differences in commissions. The difference in commissions of 24.53 is larger than in the other conditions (UNDIS-U vs. FINES: $p = 0.0528$ (MWU)), UNDIS-I vs. FINES: $p = 0.0929$ (MWU)), but the average cutoff of 0.383 is not significantly different. The effect of fines to improve advice quality is also reflected in the expected share of correct recommendations in FINES 70.7 percent and therefore the highest among all conditions and significantly higher than in UNDIS-I ($p = 0.0208$). FINES also shows the largest share of unbiased cutoff choices (41.0%), which is significantly larger than in UNDIS-U (MWU: $p = 0.0487$; RRO: $p < 0.1$). Further, panel regressions using data from all conditions show that
advisors choose a higher cutoff with fines for a given difference in commissions and other covariates ((M1) and (M2) intable 4.16).

Similar to DISCL, commissions of both firms are considerably above the predicted levels. Therefore the two conditions with policy intervention contrast the two conditions with undisclosed commission, where commissions fall below the predictions. Yet, in FINES commissions of firm B are weakly significantly lower than in both conditions with undisclosed commissions (UNDIS-U vs. FINES: 0.0707 (MWU); UNDIS-I vs. FINES: 0.0587 (MWU)). A possible explanation for the high commissions of firm A in FINES are other-regarding social preferences of firm A, as the high level of commissions helps advisors to achieve a welfare of 13.59 which is significantly above the predicted negative values (-5.5/-11.8). This conjecture is supported by the regression in table 4.11 which shows that donation in the dictator game are positively correlated with commissions of firm A. Interestingly, table 4.12 indicates the opposite correlation for firm B. Also here, a possible explanation is inequality aversion, but due to the low payoff potential of firm B as firm B eventually only achieves a welfare of 3.333 just above zero. Similar to the other conditions, the price for product A is below predicted levels while the price of product B is close to the predictions.

4.5.2 Advisor and consumer behaviour in detail

In this section we present further details of the results and focus on comparing results across conditions.

4.5.2.1 Advisor behaviour

Cutoff and quality of advice As predicted, advice in all conditions is steered in the direction of product A. With a mean cutoff below 50%, product A is expected to be recommended more often than product B. Indeed, product A was recommended in 65.3% (470/720) of all stage-games, ranging from 61.8% in UNDIS-U to 69.3% in UNDIS-I. Contrary to predictions, the cutoffs in the two conditions with undisclosed commissions do not differ. They further do not differ from the cutoff in the regulated conditions which are close to their predicted values. Cutoffs range only between 0.350 in DISCL and 0.383 in FINES.

Advisors react to differences in commissions in the predicted direction. Figure 4.3a shows that the cutoff decreases as the difference in commissions increases. This is also indicated by the regressions in tables 4.15 and 4.16 which show negative coefficients for the difference in commissions.
4.5. Results

**FIGURE 4.2**
Advisor reaction to commission differences.

(a) Average advisor cutoff choice depending on difference in commissions.

(b) Expected share of correct recommendations ("Quality of Advice") depending on difference in commissions.

Purely self-regarding advisors should choose a cutoff of one and zero respectively, when the absolute difference in commissions is larger than 40 ECU. As figure 4.3a shows, the experimental means at the extreme commission differences are still well below one and above zero, respectively in all conditions. These results are in line with a vast number of findings from the experimental literature. For instance, Dulleck et al. (2011) find in a market experiment on credence goods that a significant share of suppliers behaves honestly despite contrary monetary incentives.

As table 4.6 shows, the mean share of expected correct recommendations vary between 66.9% in UNDIS-I and 70.7% in FINES. Although the differences between conditions do not seem to be large, the share is significantly higher in both FINES and UNDIS-U than in UNDIS-I (UNDIS-U vs. UNDIS-I: \( p = 0.0707 \), UNDIS-U vs. DISCL: 0.5286, UNDIS-I vs. DISCL: \( p = 0.0208 \)). These results can be explained by the different distributions of cutoff choices between conditions. It seems that in FINES and to a lesser extent in UNDIS-U advisors choose less extreme cutoffs than in UNDIS-I. This is illustrated by the histograms of the cutoff-choices in figure 4.4.

**0%-cutoff choices** Figure 4.4 shows that in UNDIS-I, a larger share of advisors, 22.4%, chooses the 0%-cutoff. Despite the share in UNDIS-I is looming large compared to the other conditions, the differences are not significant, because they are mainly driven by two matching groups with exceptionally large values (see fig-
FIGURE 4.4
Distribution of advisor cutoff choices across conditions.

Note: Y-axes show fractions. Vertical lines show mean (solid) and theoretical prediction (dashed).

Figure 4.6b). Figure 4.6a shows the frequency of 0%-cutoff choices (decision 1 in figure 4.3) for non-negative differences between the commissions of firm A and firm B. Advisors choose the extreme cutoff in less than ten percent of all games when the difference in conditions are low. With increasing differences in commissions, the frequencies of 0%-cutoff choices increase in all conditions.

FIGURE 4.5
0%-cutoff choices.

(a) Share of 0%-cutoff choices depending on difference in commissions. (b) Share of 0%-cutoff choices across conditions and matching groups. Note: matching groups (black); mean (red)

Unbiased cutoff choices. Figure 4.4 shows a concentration of cutoff-choices around cutoffs of 50%. We denote the choices of the three middle cutoffs (advisor
choices 11 to 13 in table 4.3) as unbiased cutoff choices, because these recommendation strategies lead to the highest possible probability of a match of the recommended product and the required product. Figure 4.7 shows the frequency of unbiased cutoff choices depending on the absolute difference of commissions. In line with predictions, we observe higher frequencies of unbiased cutoff choices for small commission differences than for large differences. It seems that frequencies drop from commission differences of 0 to differences of 20 and remain steady from thereon. A possible explanation is that there are subjects, who choose unbiased behaviour, regardless of the incentives. This observation is in line with other experiments on credence goods (Dulleck et al., 2011). On the other hand many, but not all subjects choose the unbiased cutoff when there is no conflict of interest. A possible explanation for this behaviour are preferences for efficiency (Charness and Rabin, 2002), because deviating from the honest cutoff towards a slightly lower cutoff increases expected market welfare due to more recommendations of the cost-efficient product A.

4.5.2.2 Consumer behaviour

Recommendation rejections. Theory predicts that in equilibrium, consumers accept all recommendations. In the experiment, consumers accept recommendations only in 57.6% (UNDIS-U) to 68.8% (DISCL) of the cases. Rates do not differ significantly between conditions (UNDIS-U vs. DISCL: \( p = 0.1519 \)). Nevertheless, the absolute differences are relevant for explaining welfare differences between markets (see below). One explanation for the higher acceptance rates with
disclosure is the somewhat unintuitive observation from the literature that disclosure can actually increase trust on the consumer side. Another explanation is provided the “burden of disclosure”, a psychological trait that describes consumers’ increased pressure to comply with advice in the face of disclosure (Loewenstein et al., 2011; Sah et al., 2013).

Trade frequencies are higher for product A than for product B: While consumers accept product A recommendations in 68.5% (UNDIS-U) to 80.5% (DISCL) of the cases, consumers only accept recommendations of good B in less than fifty percent of the cases in all conditions but in UNDIS-I, where the rate is around 60%. There are strong indications that prices can explain the difference. For one part, the lower rejection rate of product B recommendations in UNDIS-I comes along with lower prices of product B. The price explanation is further supported by the finding that prices for product A are considerably below predicted levels, while prices for product B are close to predictions (with the exception of UNDIS-I). Furthermore, the regressions on consumers’ purchase decisions (tables 4.9 and 4.10) show that higher prices significantly reduce consumers’ likelihood to purchase a product.

Matching. Consumers’ experience a match with their need in 43.89% (316/720) of all games in total, mismatches in 20.97% (151/720) and neither in case of rejections which occur in 35.14% (253/720) of the cases. Matching rates range from 36.1% in UNDIS-U and 49.0% in DISCL (UNDIS-U vs. DISCL: p = 0.0502 (MWU)). In all conditions the matching frequencies are significantly below predictions and also below the levels we would expect given the observed cutoffs and the predicted trade frequency of 100%. This is because consumers decide not to purchase the recommended product in 35.14% of all games. The experience of a mismatch does not seem to influence consumer behaviour over periods. Consumers’ frequency of accepting a product recommendation in period $t$ after having experienced a mismatch in period $t - 1$ is 63.04%, it is almost equal at 63.82% after not having experienced a mismatch in period $t - 1$. This possibly indicates that consumers do not change their beliefs about advice depending on experience.

4.5.3 Market welfare & agents’ payoffs

Market welfare (table 4.7), calculated as the sum of all payoffs, fell significantly short of the predicted levels in all conditions, mainly because consumers reject a

---

49 Pearson et al. (2006) show this effect for patients in health care.
50 The effect has been shown to be stronger when disclosure is provided by the advisor and not by third parties and might therefore be stronger in real-world face-to-face advice, because consumers may attribute disclosure presented by an advisor as disclosure provided by the advisor.
non-negligible share of recommendations. Market welfare is lowest in \textit{UNDIS-U} with 59.09 and highest in \textit{DISCL} with 82.71. For the comparison of conditions with \textit{UNDIS-U} one-tailed tests may be justified due to our theoretical prior that consumer welfare is lowest in \textit{UNDIS-U} (\textit{UNDIS-U} vs. \textit{DISCL}: \( p = 0.0602 \)). We hence find weak evidence that total market welfare is larger in \textit{DISCL} than in \textit{UNDIS-U}. Figure 4.8 (left side) shows results on matching group level. The heterogeneity in matching group averages explains why market welfare does not vary significantly between \textit{UNDIS-U} and \textit{UNDIS-I} as well as \textit{FINES}.

Consumers in all conditions are significantly better off than theory predicts. In theory, consumers cannot earn high payoffs, because firms set prices as close as possible to consumers’ willingness to pay – and in case of naive consumers, the willingness to pay for the more often recommended product A is too high in the first place. In the experiment, prices for product A in all conditions are low enough to give room for consumer profits, however. Even in \textit{UNDIS-U} in which consumers were not informed about possible commission payments, their average payoff reaches 8.89 and is therefore more than 30 higher than predicted. Consumer welfare is around 20 in all other conditions, but it is just not significantly higher than in \textit{UNDIS-U} due to heterogeneity in matching groups’ averages (see figure 4.9, right side) (\textit{UNDIS-U} vs. \textit{UNDIS-I}: \( p = 0.1062 \), \textit{UNDIS-U} vs. \textit{DISCL}: \( p = 0.1209 \), \textit{UNDIS-U} vs. \textit{FINES}: \( p = 0.1372 \), \textit{UNDIS-U}).
Payoffs of both firms are lower than predicted by theory in all conditions. The mean welfare of firm A players only ranges between 25.36 ECU in DISCL and 29.53 in ECU and therefore falls significantly short of the predictions, ranging from 50.9 in UNDIS-I and 82.7 in DISCL and FINES. In UNDIS-U and UNDIS-I this is because product A gets recommended less often than predicted and consumers do not accept recommendations in all cases. These effects are stronger than the profit-enhancing effect that commissions are lower than predicted. In DISCL and FINES the effect is likewise driven by consumers’ recommendations rejections. It is however not driven by cutoffs, but by commissions which are above the predicted levels.

Advisor welfare is smaller than predicted when commissions are undisclosed and larger than predicted when commissions are disclosed. In the conditions with undisclosed commissions, this is because firm A pays lower commissions and consumers reject more recommendations than predicted. In the regulated conditions both firms choose higher commissions, which has a stronger effect than recommendation rejections. Nevertheless, due to the fines for mismatches, advisor welfare is significantly or weakly significantly lower in FINES than in the other conditions using one-sided MWU-tests (FINES vs. UNDIS-I: \( p = 0.0043 \); vs. UNDIS-U: \( p = 0.0773 \); vs. DISCL: \( p = 0.0027 \)). The amount of fines collected by the regulator in FINES is almost as expected (11.56 vs. 14.7/14.2). This is although, contrary to the predictions, consumers accepted recommendations in only 63% of all interactions. At the same time, however, trade of product A took place in 72.8 percent of all cases while trade of product B in less than 45 percent of the cases, counteracting this effect.

4.5.4 Developments over time

Our results only show few indications for changing behaviour over time. This finding is expected since we let subjects play one-shot games with random and anonymous re-matching in every period which should limit reputation effects. Learning effects may play a role in theory, but we do not find much evidence for it. We present some figures showing the development of key variables over time in appendix 4.7.

4.5.5 Strategic sophistication

We find some effects of strategic sophistication as measured by the level-k game on firm’s decisions. We elicited strategic sophistication, because our four-stage advice game is fairly complicated, especially for firms and, to a lesser degree, for advisors. The regressions in table 4.11 show that a lower strategic sophistication
(a higher level-k number) is significantly associated with lower commissions of firm A in \textit{UNDIS-U} and with higher commissions of firm A in \textit{DISCL}. Both effects are consistent with our results, as commissions in \textit{UNDIS-U} are below the predicted level and above the predicted level in \textit{DISCL}. Hence, a lack of strategic sophistication compared to the rational agents in the theoretical model could partly help to explain observed commission patterns. We find no effect for firm B, however – potentially, because firm B players’ strategic decisions are simpler than those of firm A, as the high production costs limit the number of potential decision which are possible without going into losses.

Furthermore, the regression on prices of firm B suggests that a lower strategic sophistication (a higher Level-K number) leads to a lower price for product B in \textit{DISCL}. This finding is in line with the observation that the price for product B is (weakly significantly) above the predicted level. It possibly can be explained by the difference in commissions of 19.21 from which sophisticated agents can expect a lower cutoff than actually observed and hence a higher willingness of consumers to pay for product B. With respect to advisor behaviour we find that level-k numbers are positively associated with the cutoff in both conditions with undisclosed commissions (table 4.15).

These results show that strategic sophistication as measured by a simple level-k test has some explanatory power in our advice game. This is despite recent research has shown that subjects’ level-k choices are not necessarily a predictor for strategic sophistication (Georganas et al., 2015).

4.6 Discussion & Conclusion

Empirical evidence on the effect of policies such as disclosure on market dynamics in markets for advice has so far been limited due to the complexity of real-world markets for advice (Leuz and Wysocki, 2016). We have conducted an experiment to study markets with advice following the model of Inderst and Ottaviani (2012a) including firms, advisors, consumers and endogenously created conflicts of interest. We have investigated the effects of consumer information about commissions and different policy interventions, particularly disclosure of commission payments.

The dynamics studied in our model are highly relevant for current policy discussions not only the area of financial advice and health care, but also in other

\footnote{As a different measure of strategic sophistication we use the indicator whether subjects study in fields with mathematical contents (engineering, maths, physics and economics)\textsuperscript{52}, but we do not find effects worth mentioning for this variable.}
markets with similar structures such as consulting or policy advice. Disclosure of commissions is commonly believed to be a remedy for shortcomings in such markets. Research on the other hand has shown that the effects of disclosure are far less than well understood. This study has shown that the market dynamics with and without disclosure in an experimental setting can differ significantly from theoretical predictions.

We observe – in line with the theoretical analysis in Inderst and Ottaviani (2012a) – endogenously created conflicts of interest and biased advice towards the more cost-efficient product in all conditions, with and without disclosure. This basic result breaks with the publicly perceived wisdom that disclosure always works out in the best interest of consumers. We find that consumer welfare in unregulated markets is already higher than predicted and disclosure does not increase consumer welfare significantly. Market welfare is highest with disclosure, because consumers accept more recommendations of product A than in other treatments, whereas we do not observe the theoretically predicted decrease of commission levels. However, the differences in market welfare between the conditions are not significant, with the exception that disclosure of commissions leads to a weakly significant increase in market welfare compared to an unregulated market with consumers who are not informed that commissions may be paid. In particular, fines for the advisor do not yield significant improvements compared to the same benchmark.

We find indications for social preferences in the behaviour of firms, advisors and consumers. While advisors, for instance, react to financial incentives in the predicted direction, they do so less strongly than predicted by theory. A considerable share of subjects in all treatments chooses to give unbiased advice. Moreover, for very large differences in commissions advisors do not make the corresponding theoretically predicted extreme cutoff choices. Possible behavioural traits at play here are guilt aversion (Charness and Dufwenberg, 2006) and lying aversion (Ellingsen and Johannesson, 2004; Gneezy, 2005; Gneezy et al., 2013). Erat and Gneezy (2012) show that a considerable fraction of their subjects are reluctant to lie, independent of social preferences for outcomes. Giving biased advice is not necessarily the same as lying in other context, however, and therefore such conclusions have to be taken with care. Notwithstanding, distributional preferences may play a role in our study as all actions have consequences concerning the distribution of welfare between subjects (see the discussion in Dulleck et al. (2011)). For instance, the advisor does

\[53\text{Gibson et al. (2013) show that preferences for lying are heterogeneous both within and between subjects. This is in line with our finding that only a few subjects consistently choose unbiased advice.}\]
not only influence the consumers expected payoff by his own choices, but also the expected payoff of the firms. Also, the decisions of firms influence the payoffs of both the advisor and the consumer.

Our regression analyses reveal that decisions are correlated with the donations from a dictator game with charity donations which adds to the discussion on the predictive power of dictator games (Georganas et al., 2015). Several of our findings concerning these correlations hint into the direction of inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or preferences for efficiency (Charness and Rabin, 2002; Bolton and Ockenfels, 2006). Inequity aversion may help to explain why firms set lower or higher commissions, depending on whether consumers or advisors are disadvantaged (consumers in UNDIS-U, advisors in FINES). Preferences for efficiency may explain why some advisors choose unbiased advice despite the absence of conflicts of interest. Our conjectures about social preferences are in line with Kerschbamer et al. (2017) who investigate credence goods experimentally and find that the behaviour of a large majority of subjects is consistent with preferences for efficiency and inequity aversion. Social preferences could, for instance, moreover explain why we observe lower prices than predicted for the cost-efficient product A, but prices close to predicted levels for product B. Most notably, we observe a price difference between the products when consumers are not informed about commissions, while theory predicted that that this would not be the case. This price difference in turn can explain why – across all conditions – consumers accept more recommendations for product A than for product B. The low prices for product A link to consumer welfare which is significantly higher than predicted, particularly in the unregulated market with uninformed consumers where consumer welfare is further supported by advice being less biased than predicted. Future research could possibly help to understand the role of social preferences in markets for advice in greater detail.

While theory predicts sharp declines of commission levels in our two policy treatments, experimental results only show modest reductions of commission levels and only for product B. Further research is needed in order to investigate the drivers of firm behaviour in markets with advice. While the relationship between advisors and consumers has been the focus of many studies (see section 4.2), there appears to be a research gap for the role of third-party suppliers in markets for advice.

---

54 We may be inclined to be cautious with the interpretation of these results with respect to external validity as firms in real-world markets are often big and do not act like as individuals. However, both financial and health care advice, are often carried out in face-to-face settings (Chater et al. (2010) report that 80% of individuals who recently bought financial products did so in a face-to-face setting).
In our experiment, the increased welfare with disclosure stems from higher recommendation acceptance rates of consumers, particularly for product A. A possible explanation for this is the “burden of disclosure” (Loewenstein et al., 2011). It describes the psychological phenomenon that consumers are less willing to reject advice when conflicts of interest are disclosed, because a rejection could be interpreted as an implicit accusation of unethical behaviour from consumer to advisor. In our case, the “burden of disclosure” actually works in favour of market outcomes. Another possible explanation is that disclosure increases trust on the receiver-side, as has been shown in health care (Pearson et al., 2006). These observations also point to a potential analogy with Fung et al. (2007) who report in the context of public information disclosure that the positive effects of disclosure are more likely to be a consequence of changed consumer behaviour rather than changed advisor behaviour (see the discussion in Loewenstein et al. (2011)).

In one condition we impose fines on the advisor for wrong advice as a theoretically equivalent policy to disclosure. In line with predictions, advisors react less strongly to conflicts of interest, but the conflict of interest is greater than in the other conditions in the first place. This result – in line with the theory of Inderst and Ottaviani (2012a) – shows how equilibrium effects can offset primary effects when introducing regulatory measures into a dynamic market with several actors. Hence, the condition with fines does not improve market outcomes significantly.

In our experiment, advice is a credence good, i.e., consumers need to rely on the advice given by an advisor and cannot evaluate whether bad advice or (bad) luck in the case of a bad outcome (or: good advice or (good) luck in case of a good outcome) is responsible for the outcome. In their experimental study of credence goods markets, Dulleck et al. (2011) find that liability can help to improve market outcomes. Our FINES treatment can be interpreted as an implementation of some degree of liability but not full liability as in Dulleck et al. (2011). Our setting may well reflect some real life markets better than the assumption of full liability, as in reality full liability is hard to implement when it is difficult to juridically distinguish between right and wrong advice. Contrary to Dulleck et al. (2011), our treatment with fines does not improve market outcomes significantly, but a direct comparison of the effects is not possible due to crucial differences in the models. Nevertheless, our results indicate that further research on different forms of liability in credence goods markets might be insightful.
FIGURE 4.10
Distribution of the difference of commission A and commission B choices across treatments.

Note: vertical lines show mean (solid) and theoretical prediction (dashed).

FIGURE 4.11
Distribution of commission A choices across treatments.

Note: vertical lines show mean (solid) and theoretical prediction (dashed).

4.7 Appendix I: additional figures

Figure 4.13 shows the distribution of price choices of firm A for all four conditions.
Chapter 4

**FIGURE 4.12**
Distribution of commission B choices across treatments.

Note: vertical lines show mean (solid) and theoretical prediction (dashed).

**FIGURE 4.13**
Distribution of prices A choices across treatments.

Note: vertical lines show mean (solid) and theoretical prediction (dashed).

**FIGURE 4.14**
Distribution of prices B choices across treatments.

Note: vertical lines show mean (solid) and theoretical prediction (dashed).
TABLE 4.8
Field of study of participants.

<table>
<thead>
<tr>
<th>Department</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>14</td>
<td>3.89</td>
<td>3.89</td>
</tr>
<tr>
<td>Economics</td>
<td>10</td>
<td>2.78</td>
<td>6.67</td>
</tr>
<tr>
<td>Education</td>
<td>1</td>
<td>0.28</td>
<td>6.94</td>
</tr>
<tr>
<td>Engineering</td>
<td>82</td>
<td>22.78</td>
<td>29.72</td>
</tr>
<tr>
<td>Maths or Physics</td>
<td>31</td>
<td>8.61</td>
<td>38.33</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>107</td>
<td>29.72</td>
<td>68.06</td>
</tr>
<tr>
<td>Psychology</td>
<td>10</td>
<td>2.78</td>
<td>70.83</td>
</tr>
<tr>
<td>Other Social Sciences</td>
<td>33</td>
<td>9.17</td>
<td>80.00</td>
</tr>
<tr>
<td>Other</td>
<td>59</td>
<td>16.39</td>
<td>96.39</td>
</tr>
<tr>
<td>No Student</td>
<td>13</td>
<td>3.61</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>360</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 4.15
Distribution of dictator donations to charity over all conditions.
FIGURE 4.16
Commissions and Prices over time.

FIGURE 4.17
Cutoff over time.
4.8 Appendix II: regression tables
## Table 4.9

Consumer decision A.
Analysis of consumers’ acceptance (0/1) of a recommendation for product A across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price A</strong></td>
<td>-0.005***</td>
<td>-0.006***</td>
<td>-0.005***</td>
<td>-0.006***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td><strong>Dictator Donation</strong></td>
<td>-0.005**</td>
<td>-0.000</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.92)</td>
<td>(0.61)</td>
<td>(0.61)</td>
</tr>
<tr>
<td><strong>Level-K number</strong></td>
<td>-0.001</td>
<td>0.002</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td>(0.44)</td>
<td>(0.95)</td>
<td>(0.23)</td>
</tr>
<tr>
<td><strong>Risk Taking</strong></td>
<td>0.032</td>
<td>-0.035*</td>
<td>0.036</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.10)</td>
<td>(0.26)</td>
<td>(0.13)</td>
</tr>
<tr>
<td><strong>Woman</strong></td>
<td>-0.012</td>
<td>0.179*</td>
<td>0.126</td>
<td>-0.095</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.07)</td>
<td>(0.33)</td>
<td>(0.32)</td>
</tr>
<tr>
<td><strong>Engin./Math/Econ Student</strong></td>
<td>-0.072</td>
<td>0.180*</td>
<td>-0.009</td>
<td>-0.110</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.10)</td>
<td>(0.93)</td>
<td>(0.14)</td>
</tr>
<tr>
<td><strong>Comm A -Comm B</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>Sum of Commissions</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.002*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

Panel probit regressions showing average marginal effects; p-values in parentheses, standard errors clustered at subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
TABLE 4.10
Consumer decision B.
Analysis of consumers’ acceptance (0/1) of a recommendation for product B across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price B</td>
<td>-0.009***</td>
<td>-0.003**</td>
<td>-0.007**</td>
<td>-0.012***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Dictator Donation</td>
<td>-0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.51)</td>
<td>(0.54)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.013***</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.95)</td>
<td>(0.60)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>0.068***</td>
<td>0.003</td>
<td>0.031</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.87)</td>
<td>(0.32)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Woman</td>
<td>-0.092</td>
<td>0.335***</td>
<td>0.308*</td>
<td>-0.228</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.00)</td>
<td>(0.06)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Engin./Math/Econ Student</td>
<td>-0.197</td>
<td>0.069</td>
<td>-0.065</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.67)</td>
<td>(0.64)</td>
<td>(0.88)</td>
</tr>
<tr>
<td>Comm A - Comm B</td>
<td></td>
<td></td>
<td></td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.88)</td>
</tr>
<tr>
<td>Sum of Commissions</td>
<td></td>
<td></td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.43)</td>
<td></td>
</tr>
</tbody>
</table>

Panel probit regressions showing average marginal effects; p-values in parentheses, standard errors clustered at subject level. * p < 0.10, ** p < 0.05, *** p < 0.01
TABLE 4.11
Commission choice firm A across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictator Donation</td>
<td>-0.11</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.29*</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.79)</td>
<td>(0.83)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>-0.32***</td>
<td>0.10</td>
<td>0.48***</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.59)</td>
<td>(0.00)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>-2.29*</td>
<td>3.22**</td>
<td>0.00</td>
<td>-1.49</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.03)</td>
<td>(1.00)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Woman</td>
<td>6.87</td>
<td>9.71</td>
<td>-10.96*</td>
<td>7.16</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.19)</td>
<td>(0.07)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Engin./Math/Econ Student</td>
<td>20.91***</td>
<td>10.97*</td>
<td>-2.41</td>
<td>-3.89</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.06)</td>
<td>(0.68)</td>
<td>(0.71)</td>
</tr>
<tr>
<td>Period</td>
<td>-0.94</td>
<td>-0.18</td>
<td>0.66</td>
<td>-2.05***</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.80)</td>
<td>(0.50)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Constant</td>
<td>72.25***</td>
<td>29.79**</td>
<td>43.60***</td>
<td>58.84***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

N 144 192 192 184

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.
* p < 0.10, ** p < 0.05, *** p < 0.01

TABLE 4.12
Commission choice firm B across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictator Donation</td>
<td>0.05</td>
<td>-0.11</td>
<td>-0.02</td>
<td>-0.28***</td>
</tr>
<tr>
<td></td>
<td>(0.58)</td>
<td>(0.21)</td>
<td>(0.80)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>0.18</td>
<td>0.17</td>
<td>0.05</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.39)</td>
<td>(0.75)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>-0.97</td>
<td>1.08</td>
<td>-2.63</td>
<td>3.58***</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.57)</td>
<td>(0.17)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Woman</td>
<td>1.74</td>
<td>7.83</td>
<td>-0.63</td>
<td>15.31**</td>
</tr>
<tr>
<td></td>
<td>(0.79)</td>
<td>(0.20)</td>
<td>(0.93)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Engin./Math/Econ Student</td>
<td>-13.08***</td>
<td>0.75</td>
<td>7.49</td>
<td>7.19</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.91)</td>
<td>(0.43)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Period</td>
<td>-1.90</td>
<td>-0.17</td>
<td>-0.40</td>
<td>-0.83</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.84)</td>
<td>(0.68)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Constant</td>
<td>48.11***</td>
<td>29.04*</td>
<td>54.15***</td>
<td>24.07*</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.07)</td>
<td>(0.00)</td>
<td>(0.06)</td>
</tr>
</tbody>
</table>

N 144 192 192 192

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.
* p < 0.10, ** p < 0.05, *** p < 0.01
### TABLE 4.13
Price choice firm A across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comm A -Comm B</td>
<td>0.46***</td>
<td>0.26*</td>
<td>0.33***</td>
<td>0.28***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.06)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Sum of Commissions</td>
<td>0.41***</td>
<td>0.44***</td>
<td>0.22**</td>
<td>0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.03)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Dictator Donation</td>
<td>0.19**</td>
<td>-0.33**</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.15)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>-0.10</td>
<td>-0.10</td>
<td>0.27</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.74)</td>
<td>(0.34)</td>
<td>(0.60)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>3.71***</td>
<td>-0.49</td>
<td>1.89</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.87)</td>
<td>(0.48)</td>
<td>(0.34)</td>
</tr>
<tr>
<td>Woman</td>
<td>-14.47**</td>
<td>18.60</td>
<td>20.38**</td>
<td>10.75</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.17)</td>
<td>(0.04)</td>
<td>(0.64)</td>
</tr>
<tr>
<td>Engineering/Math Student</td>
<td>4.11</td>
<td>-15.81</td>
<td>26.12**</td>
<td>9.82</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.17)</td>
<td>(0.02)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Period</td>
<td>0.56</td>
<td>1.55</td>
<td>-0.60</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.20)</td>
<td>(0.61)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Constant</td>
<td>81.85***</td>
<td>98.41***</td>
<td>61.16***</td>
<td>87.65**</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
### TABLE 4.14
Price choice firm B across conditions.

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comm A -Comm B</td>
<td>-0.22**</td>
<td>-0.17**</td>
<td>-0.18***</td>
<td>-0.27***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Sum of Commissions</td>
<td>0.31**</td>
<td>0.15**</td>
<td>0.16***</td>
<td>0.22**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Dictator Donation</td>
<td>-0.14</td>
<td>0.42***</td>
<td>-0.07</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.00)</td>
<td>(0.11)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>-0.12</td>
<td>-1.73***</td>
<td>-0.22**</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>-1.04</td>
<td>-3.19</td>
<td>2.82**</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.36)</td>
<td>(0.03)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Woman</td>
<td>0.23</td>
<td>18.04</td>
<td>9.78**</td>
<td>-7.62</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(0.31)</td>
<td>(0.03)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Engineering/Math Student</td>
<td>-10.83</td>
<td>21.34</td>
<td>-11.04**</td>
<td>-7.63*</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.17)</td>
<td>(0.02)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Period</td>
<td>-0.75</td>
<td>0.82</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.40)</td>
<td>(0.81)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>Constant</td>
<td>176.17***</td>
<td>226.81***</td>
<td>174.84***</td>
<td>175.91***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

N: 144 192 192 192

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.
* p < 0.10, ** p < 0.05, *** p < 0.01
### TABLE 4.15
Advisor cutoff choice across conditions

<table>
<thead>
<tr>
<th></th>
<th>(UNDIS-U)</th>
<th>(UNDIS-I)</th>
<th>(DISCL)</th>
<th>(FINES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comm.A -Comm.B</td>
<td>-0.0016*</td>
<td>-0.0049***</td>
<td>-0.0024***</td>
<td>-0.0027***</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Sum of Commissions</td>
<td>-0.0000</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.77)</td>
<td>(0.39)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Price A-Price B</td>
<td>0.0003</td>
<td>-0.0014***</td>
<td>-0.0001</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td>(0.00)</td>
<td>(0.73)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Sum of Prices</td>
<td>-0.0007</td>
<td>0.0001</td>
<td>-0.0003</td>
<td>-0.0002</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.64)</td>
<td>(0.32)</td>
<td>(0.65)</td>
</tr>
<tr>
<td>Dictator Donation</td>
<td>0.0006</td>
<td>0.0014***</td>
<td>-0.0004</td>
<td>-0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.02)</td>
<td>(0.42)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Level-K number</td>
<td>0.0022***</td>
<td>0.0015***</td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.94)</td>
<td>(0.67)</td>
</tr>
<tr>
<td>Risk Taking</td>
<td>0.0126</td>
<td>0.0073</td>
<td>0.0071</td>
<td>-0.0076</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.34)</td>
<td>(0.41)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Woman</td>
<td>0.0671*</td>
<td>-0.0627</td>
<td>0.0201</td>
<td>-0.0261</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.14)</td>
<td>(0.71)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Eng./Math/Econ Student</td>
<td>-0.0264</td>
<td>-0.0535</td>
<td>-0.0894*</td>
<td>-0.0318</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td>(0.15)</td>
<td>(0.08)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>Period</td>
<td>-0.0104*</td>
<td>-0.0079</td>
<td>-0.0125***</td>
<td>-0.0021</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.23)</td>
<td>(0.01)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.4608***</td>
<td>0.3659***</td>
<td>0.5227***</td>
<td>0.6038***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

N 144 192 192 192

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
### TABLE 4.16
Advisor cutoff choice; using data from all conditions.

<table>
<thead>
<tr>
<th></th>
<th>(M1)</th>
<th>(M2)</th>
<th>(M3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNDIS-U</td>
<td>0.001 (0.97)</td>
<td>0.011 (0.80)</td>
<td>0.027 (0.81)</td>
</tr>
<tr>
<td>DISCL</td>
<td>0.031 (0.36)</td>
<td>0.069 (0.11)</td>
<td>0.096 (0.27)</td>
</tr>
<tr>
<td>FINES</td>
<td>0.068** (0.02)</td>
<td>0.106*** (0.00)</td>
<td>0.148 (0.13)</td>
</tr>
<tr>
<td>Comm.A -Comm.B</td>
<td>-0.003*** (0.00)</td>
<td>-0.005*** (0.00)</td>
<td>-0.005*** (0.00)</td>
</tr>
<tr>
<td>(Comm.A-Comm.B) x DISCL</td>
<td>0.002** (0.01)</td>
<td>0.002** (0.01)</td>
<td>0.002** (0.03)</td>
</tr>
<tr>
<td>(Comm.A-Comm.B) x FINES</td>
<td>0.002** (0.02)</td>
<td>0.002** (0.02)</td>
<td>0.002** (0.03)</td>
</tr>
<tr>
<td>(Comm.A-Comm.B) x UNDIS-U</td>
<td>0.002*** (0.01)</td>
<td>0.002*** (0.01)</td>
<td>0.002*** (0.01)</td>
</tr>
<tr>
<td>Sum of Commissions</td>
<td>0.000 (0.40)</td>
<td>0.000 (0.49)</td>
<td></td>
</tr>
<tr>
<td>Price A-Price B</td>
<td>-0.001*** (0.00)</td>
<td>-0.001*** (0.00)</td>
<td>-0.001*** (0.00)</td>
</tr>
<tr>
<td>Sum of Prices</td>
<td>-0.000** (0.05)</td>
<td>-0.000** (0.03)</td>
<td></td>
</tr>
<tr>
<td>Dictator Donation</td>
<td>0.001** (0.01)</td>
<td>0.001** (0.02)</td>
<td></td>
</tr>
<tr>
<td>Dictator Donation x DISCL</td>
<td>-0.002** (0.01)</td>
<td>-0.002** (0.02)</td>
<td></td>
</tr>
<tr>
<td>Dictator Donation x FINES</td>
<td>-0.002*** (0.00)</td>
<td>-0.002*** (0.00)</td>
<td></td>
</tr>
<tr>
<td>Dictator Donation x UNDIS-U</td>
<td>-0.001 (0.33)</td>
<td>-0.001 (0.14)</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>-0.038 (0.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman x DISCL</td>
<td>0.051 (0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman x FINES</td>
<td>0.013 (0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman x UNDIS-U</td>
<td>0.073 (0.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng./Math/Econ Student</td>
<td>-0.063 (0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng./Math/Econ St. x DISCL</td>
<td>-0.027 (0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng./Math/Econ St. x FINES</td>
<td>0.029 (0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng./Math/Econ St. x UNDIS-U</td>
<td>-0.008 (0.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Taking</td>
<td>0.008 (0.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Taking x DISCL</td>
<td>-0.002 (0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Taking x FINES</td>
<td>-0.016 (0.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Taking x UNDIS-U</td>
<td>-0.001 (0.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period</td>
<td>-0.007 (0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period x DISCL</td>
<td>-0.005 (0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period x FINES</td>
<td>0.004 (0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period x UNDIS-U</td>
<td>-0.004 (0.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.426*** (0.00)</td>
<td>0.478*** (0.00)</td>
<td>0.517*** (0.00)</td>
</tr>
</tbody>
</table>

Panel OLS regressions; p-values in parentheses, standard errors clustered at subject level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
4.9 Appendix III: theoretical framework

4.9.1 Undisclosed commissions with naive consumers

We calculate the best responses from (4.10). With naive consumers, prices for both products are given by 
\[ v_h + \frac{1}{4}v_l. \]
Inserting these prices into (4.10), additionally using the uniform distribution and inserting 4.1 for the cutoff \( q^* \) using we obtain:
\[
\pi_A = \left( \frac{3}{4}v_h + \frac{1}{4}v_l - c_A - f_A \right) \left( \frac{1}{2} + \frac{f_A - f_B}{2(w + l)} \right),
\]
\[
\pi_B = \left( \frac{3}{4}v_h + \frac{1}{4}v_l - c_B - f_B \right) \left( \frac{1}{2} - \frac{f_A - f_B}{2(w + l)} \right).
\]

For firm A, we derive the first derivative and formulate the first-order-condition. Using the product rule we obtain:
\[
\frac{\partial \pi_A}{\partial f_A} = - \left( \frac{1}{2} + \frac{f_A - f_B}{2(w + l)} \right) + \left( \frac{3}{4}v_h + \frac{1}{4}v_l - c_A - f_A \right) \frac{1}{2(w + l)} = 0.
\]
After a few reformulations we obtain the best response of firm A given the commission of firm B:
\[
f_A = \frac{\frac{3}{4}v_h + \frac{1}{4}v_l - c_A - (w + l) + f_B}{2}.
\]
(4.17)

Likewise we derive the first derivative for firm B and formulate the first-order-condition. Using the product rule we obtain:
\[
\frac{\partial \pi_B}{\partial f_B} = - \left( \frac{1}{2} - \frac{f_A - f_B}{2(w + l)} \right) + \left( \frac{3}{4}v_h + \frac{1}{4}v_l - c_B - f_B \right) \frac{1}{2(w + l)} = 0.
\]
After a few reformulations we obtain the best response of firm B given the commission of firm A:
\[
f_B = \frac{\frac{3}{4}v_h + \frac{1}{4}v_l - c_B - (w + l) + f_A}{2}.
\]
(4.18)

Plugging (4.18) into (4.17) we obtain the equilibrium best response of firm A:
\[
f_A = \frac{\frac{3}{4}v_h + \frac{1}{4}v_l - c_A - (w + l) + \frac{3}{4}v_h + \frac{1}{4}v_l - c_B - (w + l) + f_A}{2}
\]
\[\Leftrightarrow\]
\[
\frac{3}{4}f_A = \frac{9}{16}v_h + \frac{3}{16}v_l - \frac{1}{2}c_A - \frac{3}{4}(w + l) - \frac{1}{4}c_B
\]
\[\Leftrightarrow\]
Chapter 4

\[ f_A = \frac{3}{4} v_h + \frac{1}{4} v_l - (w + l) - \frac{2}{3} c_A - \frac{1}{3} c_B. \]

The best response of firm B can be calculated vice versa by plugging (4.17) into (4.18).

4.9.2 Undisclosed commissions with wary consumers

In this section we assume, as in Inderst and Ottaviani (2012a), that consumers instead are wary. Although they cannot observe commissions directly, they form beliefs about them that are correct in equilibrium. We follow the approach of Inderst and Ottaviani (2012a). Prices for products A and B, respectively, equal the willingness to pay and take into account the expected cutoff of the advisor \( \hat{q}^* \). For a given expected cutoff prices are is given by

\[
p_A(\hat{q}^*) = \int_{\hat{q}^*}^{1} (qv_h + (1 - q)v_l) \frac{1}{1 - G(\hat{q}^*)} dq
= v_l + \Delta v \frac{1}{2} (1 + \hat{q}^*)
\]

and

\[
p_B(\hat{q}^*) = \int_{0}^{\hat{q}^*} ((1 - q)v_h + qv_l) \frac{1}{G(\hat{q}^*)} dq
= v_h - \Delta v \frac{1}{2} \hat{q}^*.
\]

The difference between the two prices is given by \( p_A^* - p_B^* = (\hat{q}^* - \frac{1}{2}) \Delta v \). The consumer takes into account that advice is potentially biased. If \( q^* < \frac{1}{2} \), the consumer gets recommend product A more often than with unbiased advice. Following the consumer is willing to pay more for product B than for product A, given a recommendation for either product. The difference \( p_A^* - p_B^* \) is negative if \( q^* < \frac{1}{2} \).

We use the advisor’s cut-off from (4.1) given by

\[
q^* = \frac{1}{2} - \frac{f_A - f_B}{2(w + l)}.
\]

(4.19)

For a given \( \hat{q}^* \), firms best responses are given by

\[
f_A = \frac{2}{3} p_A(\hat{q}^*) + \frac{1}{3} p_B(\hat{q}^*) - w - l - \frac{2}{3} c_A - \frac{1}{3} c_B
\]

and

\[
f_B = \frac{2}{3} p_B(\hat{q}^*) + \frac{1}{3} p_A(\hat{q}^*) - w - l - \frac{2}{3} c_B - \frac{1}{3} c_A,
\]
respectively. The difference between the two best responses yields $f_A - f_B = \frac{1}{3}(p_A(\hat{q}^*) - p_B(\hat{q}^*) - c_A + c_B)$. By plugging the best responses into (4.19) we obtain

$$\hat{q}^* = \frac{1}{2} - \frac{p_A(\hat{q}^*) - p_B(\hat{q}^*) - c_A + c_B}{6(w + l)}.$$ 

By plugging in the expected prices we obtain

$$\hat{q}^* = \frac{1}{2} - \frac{(\hat{q}^* - \frac{1}{2})\Delta v - c_A + c_B}{6(w + l)}. \quad (4.20)$$

In equilibrium $\hat{q}^* = q^*$. Plugging $q^*$ for $\hat{q}^*$ in (4.20) and solving for $\hat{q}$ yields

$$q^* = \frac{1}{2} + \frac{c_A - c_B}{6(w + l) + \Delta v}. \quad (4.21)$$

4.9.3 Disclosure

With disclosure, the profits of firms $A$, and $B$ are given by

$$\pi_A = (p_A(q^*(f_A, f_B)) - c_A - f_A^D) \left[1 - G(q^*)\right] \quad (4.22)$$

and

$$\pi_B = (p_B(q^*(f_A, f_B)) - c_B - f_B^D) \left[G(q^*)\right], \quad (4.23)$$

respectively. To calculate the respective best response we take the derivative of profits with respect to commissions. Now, with disclosure, we have to take into account how the consumer’s willingness to pay, and accordingly the price, for the respective products changes when commissions change. The price for product $A$ is given by

$$p_A(q^*(f_A)) = \int_{q^*}^{1} (qv_h + (1 - q)v_l) \frac{g(q)}{1 - G(q^*)} dq = \frac{1}{1 - q^*} \int_{q^*}^{1} (qv_h + (1 - q)v_l) dq = \frac{1}{1 - q^*} \int_{q^*}^{1} (q\Delta v + v_l) dq = \frac{1}{1 - q^*} \left[\frac{1}{2}q^2\Delta v + qv_l\right]_{q^*}^1 = \frac{1}{1 - q^*} \left(\frac{1}{2}\Delta v(1 - q^*)^2 + (1 - q^*)v_l\right) = \frac{1}{1 - q^*} \left(\frac{1}{2}\Delta v(1 - q^*)(1 + q^*) + (1 - q^*)v_l\right) = v_l + \frac{1}{2}\Delta v(1 + q^*) \quad (4.24)$$
where we have used the uniform distribution in the second step. Similarly for product $B$

$$p_B(q^*(f_B)) = \int_0^q ((1-q)v_h + qv_l) dq = \int_0^q (1-q)v_h dq = \frac{1}{q^*} \int_0^{q^*} ((1-q)v_h + qv_l) dq = \frac{1}{q^*} \int_0^{q^*} \left[ \frac{1}{2} q^2 \Delta v + qv_l \right] dq = 1 - \frac{1}{2} \Delta v q^*. \quad (4.25)$$

For firm $A$ we have, using (4.22), (4.24), (4.1) and the uniform distribution:

$$\frac{\partial \pi_A}{\partial f_A} = \left( \frac{\partial p_A^*}{\partial q^*} \frac{\partial q^*}{\partial f_A} - 1 \right) (1-q^*) + (p_A^* - c_A - f_A) \frac{\partial q^*}{\partial f_A} \frac{1}{q^*} = 0$$

We insert

$$\frac{\partial p_A^*}{\partial q^*} = \frac{1}{2} \Delta v,$$

$$\frac{\partial q^*}{\partial f_A} = -\frac{1}{2(w+l)}$$

and

$$q^* = \frac{1}{2} - \frac{f_A - f_B}{2(w+l)}. \quad (4.26)$$

After insertion and some reformulation we obtain the best response of firm $A$:

$$f_{A,D} = q^* \Delta v - (1 - q^*) 2(w+l) - v_l - c_A. \quad (4.27)$$

For firm $B$, we have, using (4.23), (4.25), (4.1) and the uniform distribution:

$$\frac{\partial \pi_B}{\partial f_B} = \left( \frac{\partial p_B^*}{\partial q^*} \frac{\partial q^*}{\partial f_B} - 1 \right) q^* + (p_B^* - c_B - f_B) \frac{\partial q^*}{\partial f_B} \frac{1}{q^*} = 0$$

We insert

$$\frac{\partial p_B^*}{\partial q^*} = -\frac{1}{2} \Delta v,$$

$$\frac{\partial q^*}{\partial f_B} = \frac{1}{2(w+l)}$$

and

$$q^* = \frac{1}{2} - \frac{f_A - f_B}{2(w+l)}. \quad (4.28)$$

After insertion and some reformulation we obtain the best response of firm $B$:

$$f_{B,D} = (1 - q^*) \Delta v - 2(w+l)q^* + v_l - c_B.$$
By inserting (4.27) and (4.28) into (4.1) we obtain the equilibrium cut-off:

\[ q^*_D = \frac{1}{2} + \frac{c_A - c_B}{6(w + l) + 2v}. \]  (4.29)

4.9.3.1 A mixed population with naive and wary consumers

It can be shown that when a small fraction of consumers is wary instead of naive we can still expect the same equilibrium as all consumers were naive. Details available on request.

4.10 Appendix IV: instructions

In the following we present the instructions for condition \textit{UNDIS-U}. 

173
INSTRUCTIONS FOR THE EXPERIMENT

Thank you for your participation in this experiment! Please do not talk to any other participants until the experiment is over.

We are going to read you the instructions first. After that, you are going to have time to go through the instructions yourself. Before the experiments starts, we ask you to answer correctly two times four control questions. If you have a question, please raise your hand. The experimenter will see you to answer your question. Then we start with the first part of the experiments, which consists of eight rounds. The following part two consist of two decisions independent of part one and a questionnaire. The instruction for part two will be shown on your screen.

Depending on your decisions during the experiment, you can earn money. In the experiment, the currency ECU (experimental currency unit) is used. At the end of the experiment, you will be paid according to the following exchange rate:

1 ECU = 0.1 CHF

After the end of the second part of the experiment, we are going to call you in order and give you your payment according to the exchange rate in cash. Additionally, you get a show-up fee of 15 francs as well as 15 francs for completing the questionnaire. With this endowment, it is possible to compensate for possible losses during the experiment.

Group assignment and roles
You are going to be assigned randomly to a group consisting of 12 participants. The assignment of a group stays the same during the entire experiment. Within the group, the following roles are assigned:

- Three players A1
- Three players A2
- Three players B
- Three players C

At the beginning of the experiment, the computer determines randomly which role you take. This information will be shown on your screen. Your role stays the same during the entire experiment.
PROCEDURE OF PART 1 OF THE EXPERIMENT

Part 1 of this experiment consists of eight rounds, each of which consists of identical sequence of four decisions. During these eight rounds, you only interact with other members in your group. In one round, you interact within a subgroup, consisting of four players (A1, A2, B and C). These subgroups are randomly and secretly assigned before each round.

BRIEF OVERVIEW

At the end of each round, player C has to take the decision whether he buys a product recommended by player C or no product. There are two products offered by player A1 and A2 (product P1 by player A1 and product P2 by player A2). Player C needs one of these products in each round, but does not know which one.

The possibility of player C needing a certain product, is depending on the scenario. Several scenarios are possible and player C does not know which scenario is realized in each round.

Player B decides for all possible scenarios, which product he recommends to player C. By this decision, player B can influence the possibility that his recommendation is consistent with the product needed by player C.

The computer determines the scenario for each round. All scenarios have the same possibility. For each scenario, the computer determines which product C is needing.

Player A1 and player A2 decide on the price of their product. When player C buys one of their products, this is their revenue. In case of a sale, A1 and A2 have to pay production costs. Player B receives for each recommendation a possible payment, which player A1 and A2 can determine at the beginning of each round. In case the recommendation by player B is not consistent with the product needed by player C, player C will get a deduction of his profit in this round. Player C gets a higher payoff when he buys the product he needs.

--- Please turn over ---
**DER ABLAUF EINER RUNDE**

**Step 1: player A1 and A2** decide at the same time and independently on the possible payment for player B. This becomes due if player C (in step 4) buys the respective product. These payment can be chosen freely between 0 and 220 ECU and in steps of 10 ECU.

**Step 2: player A1 and A2** find out which payment both players have decided on in step 1 and decide simultaneously and independently the prices for their product 1 and 2. These are the prices player C has to pay in case he buys a product. The prices can be chosen between 0 and 220 ECU in steps of 10 ECU.
**Step 3:** Player B gets to know all the decisions taken in step 1 and 2 and he now decides on his recommendation for each scenario for player C.

**Step 4:** A scenario is realized. The computer determines according to the scenario which product player C is needing. **Player C gets to know** which product is recommended by player B as well as the price for the product. **Player C does not know** which product he needs, nor which scenario is realized or which payment player A1 and A2 have decided on. Player C has the decision to buy the recommended by player B product or no product at all.

At the end all player receive the information about the general outcomes and their gain in this round. The relevant parameter in ECU are:

- Costs of production A1: 20 ECU
- Costs of production A2: 120 ECU
- Player B receives a deduction of 40 ECU when player C does not buy the needed product.
- Player C gets 220 ECU when he buys the product needed and 70 ECU when he buys the product not needed and 0 ECU when he does not buy any product.
DETAILED EXPLANATIONS

Scenarios and needed product

There are 11 scenarios and in each round before step 4, one scenario is realized. In each round, each scenario has the same possibility. The decisions taken by player A1 and A2 (step 1 and 2) hold for all scenarios and are the same for all scenarios. Player B takes a recommendation for each scenario for player B (discussed later).

The 11 scenarios have the names S-0, S-10, S-20, S-30, S-40 etc. till S-90 and S-100. The number in the name of each scenario stands for the possibility that player C needs product P1 in case this scenario is realized.

Example:

Steps 1-3 are over…

1. The computer determines randomly and secretly which scenario is realized.


2. The computer randomly draws which product player C needs. He is using the possibility of the realized scenario.

Then Step 4 follows.

Only after step 4 you find out, which product player C really needed.
Recommendation Decision Player B (Step 3)

Player B decides through inputs in two fields, up to which scenario he recommends product P2 and from which scenario on he recommends product P1.

Example: Player B wants to recommend product P1 if scenario S-1 to S-20 is realized and he wants to recommend product P2 for S-30 and higher scenarios.

Illustration and examples.

Consequences of the decision taken by player B

We have seen that player B determines in which scenarios he recommends product P1 and Product P2. By this he influences the possibility by which he recommends product P1 or product P2.

In each round, a scenario is realized and then determined which product player C needs. Because of this, B’s decision is also influencing the possibility that the product recommended is consistent with the product needed. This possibility is the biggest with 77.3% when player B recommends S-[50] and S-[50], meaning he recommends both product with the same possibility, or chooses one of the two the adjoining options. The accuracy of the recommendation decreases symmetrical to 50% when player B exclusively recommends one or the other product.

It follows a table with possible decision of player B and the consequences in sum.
### Decision Player B

<table>
<thead>
<tr>
<th>Scenario</th>
<th>I recommend product P2 to</th>
<th>I recommend product P1 from scenario</th>
<th>Recommendation</th>
<th>Consequences of the Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-0</td>
<td></td>
<td></td>
<td>P1</td>
<td>S-0</td>
</tr>
<tr>
<td>S-10</td>
<td></td>
<td></td>
<td>S-10</td>
<td>S-0</td>
</tr>
<tr>
<td>S-20</td>
<td></td>
<td></td>
<td>P2</td>
<td>P2</td>
</tr>
<tr>
<td>S-60</td>
<td></td>
<td></td>
<td>P2</td>
<td>P2</td>
</tr>
<tr>
<td>S-100</td>
<td></td>
<td></td>
<td>P2</td>
<td>P2</td>
</tr>
</tbody>
</table>

### Realized Scenario

Each scenario has the same probability (1/11) of occurring. The table below shows the recommendation for each scenario and the accuracy of the recommendation based on the realized scenario.

### Accuracy of the Recommendation

<table>
<thead>
<tr>
<th>Realized Scenario</th>
<th>Recommendation</th>
<th>Accuracy of the Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-0</td>
<td>P1</td>
<td>50.0% 50.0%</td>
</tr>
<tr>
<td>S-10</td>
<td>P1,P1</td>
<td>54.5% 45.5%</td>
</tr>
<tr>
<td>S-20</td>
<td>P1,P2,P1</td>
<td>59.1% 40.9%</td>
</tr>
<tr>
<td>S-30</td>
<td>P1,P2,P2</td>
<td>62.7% 37.3%</td>
</tr>
<tr>
<td>S-40</td>
<td>P1,P2,P2,P1</td>
<td>66.4% 33.6%</td>
</tr>
<tr>
<td>S-50</td>
<td>P1,P2,P2,P2</td>
<td>69.1% 30.9%</td>
</tr>
<tr>
<td>S-60</td>
<td>P1,P2,P2,P2</td>
<td>71.8% 28.2%</td>
</tr>
<tr>
<td>S-70</td>
<td>P1,P2,P2,P2</td>
<td>73.6% 26.4%</td>
</tr>
<tr>
<td>S-80</td>
<td>P1,P2,P2,P2</td>
<td>75.5% 24.5%</td>
</tr>
<tr>
<td>S-90</td>
<td>P1,P2,P2,P2</td>
<td>76.4% 23.6%</td>
</tr>
<tr>
<td>S-100</td>
<td>P1,P2,P2,P2</td>
<td>77.3% 22.7%</td>
</tr>
</tbody>
</table>

Example: Player B recommends product 2 for scenarios till scenario S-10 and product 1 for scenario beginning at S-20.

You find these decisions in the fifth row in the first two columns. Further right you see which product is recommended by player B in each of the 11 scenarios (columns S-0 ... to S-100). In the last two columns on the right you see what consequences the decision by player B from the first two columns has on the probability that player C really needs the recommended product. The possibility in this example for player C needing the product is 66.4% and 33.6% for not needing the product.
Summary payoffs

The potential payoff were calculated after each round. After part 2 of the experiment, two out of the eights rounds are chosen randomly for payoff. The sum of these profits is paid, all other rounds are not considered.

Table: Overview on the concrete payoffs

<table>
<thead>
<tr>
<th>Player</th>
<th>When?</th>
<th>Profit</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cost of production:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 ECU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ 120 ECU</td>
</tr>
<tr>
<td>A1</td>
<td>When player C buys product P1</td>
<td>Price of product (determined</td>
<td>Payment to player B (determined in Step 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in Step 2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>When player C does not buy</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>product P1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>When player C buys product P2</td>
<td>Price of product (determined</td>
<td>Payment to player B (determined in Step 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in Step 2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>When player C does not buy</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>product P2</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>When player C buys the product</td>
<td>Payment of provider of the</td>
<td>• When player C does not need the product:</td>
</tr>
<tr>
<td></td>
<td>recommended</td>
<td>bought by player C (determined</td>
<td>40 ECU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in Step 1)</td>
<td>• When player C needs the product: 0 ECU</td>
</tr>
<tr>
<td></td>
<td>When player C buys no product</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>When player C buys the product</td>
<td>• When he needs the product:</td>
<td>The price of the product bought</td>
</tr>
<tr>
<td></td>
<td>recommended</td>
<td>220 ECU</td>
<td>(determined in Step 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• When he does not need the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>product: 70 ECU</td>
<td></td>
</tr>
<tr>
<td></td>
<td>When player C buys no product</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

You have now time to study the instructions in private and for answering the eight control questions on your screen.

https://www.descil.ethz.ch/projects/1701-advicemarkets
Concluding remarks

In this thesis I presented a literature categorization and three research studies on credence goods markets using theoretical, laboratory and field experimental methods. With applications to health care and markets for (financial) advice I have applied my work to two of the most relevant examples of credence goods markets. The variety of methods mirrors the approach taken by the literature on credence goods. While theory often helps to become aware of possible problems which may occur in real world markets, laboratory and field experiments provide additional insights into human behaviour under controlled conditions or in real world settings, respectively.

The field experiment in chapter 2 represents the first field experiment in the literature on credence goods which studies the impact of patient, physician- and market characteristics on overtreatment in health care in a developed country. While it addresses the general literature on credence goods, it furthermore contributes to the literature on physician-induced demand and adds to the understanding of a wide-range of topics such as practice capacities, patient information in the age of the internet and socio-economic aspects of the physician-patient relationship.

Chapter 3 of this thesis contributes to the applied theoretical literature by analysing credence goods health care markets with prevention under different institutional settings. This research adds a novel point of view to questions of inefficient prevention behaviour in health markets under supply-side moral hazard. Specifically it shows that market regulations which lead to inefficiently low levels of prevention activities might be granted, because they help to avoid other inefficiencies such as untreated health problems due to low physician demand. Furthermore, the study raises awareness that credence goods markets with different institutional setting may require different levels of policies or different policies altogether.

Chapter 4 of this thesis contributes to the experimental literature on credence goods markets by presenting the first experimental test of the model by Inderst and Ottaviani (2012a). In this study I focus on expert advisors who recommend third-party products and potentially receive commissions from the suppliers of these
Concluding remarks

products. Thereby I combine two research field which are usually analysed in isolation as the focus on expert advisors contrasts the the focus on expert providers who sell both diagnosis and services in the two other chapters. Expert advisors who receive commissions are an important market characteristic of many real world markets, e.g. in markets for financial advice and health care. The presented experiment addresses the politically highly relevant issue of endogenous conflicts of interest and the effects of disclosure of these conflicts. Two main findings are that unregulated markets perform better than expected with respect to consumer welfare and that disclosure works differently than predicted by theory. The experiments generally reveals that other-regarding preferences may play an important role in the behaviour of all actors in the market, and suggests several avenues for future research.

It seems likely that credence goods markets will remain economically important in the future. Advancements in technology create new markets in which expertise knowledge is required and hence have the potential increase the informational asymmetry between experts and consumers. However, advancements of technology may as well work in the opposite way for existing markets with expertise. For instance, automated advice (“robot advice”) in financial markets may reduce the information gap between advisors and consumers; in the same way automatised diagnostics in health care may increase patient knowledge about the required treatments. Irrespective of which trend will predominate in the long-run, credence goods markets will remain highly relevant in many sectors of the economy. The characteristics of credence goods markets lead to market failures and merit political regulation. In order for political interventions to have the desired effects in credence goods markets, policy makers need to understand the problems and to impact of different policies. It was the aim of this thesis to contribute to this understanding.
5 Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


## List of Tables

1.1 A categorization of credence goods. ........................................ 3  
1.2 Examples of credence goods in the literature. ............................. 6  
1.3 Problems associated to credence characteristics ........................... 9  
1.4 Categorization of characterisation/definitions of credence goods. ...... 12  
2.1 Experimental conditions. .................................................... 20  
2.2 Experimental conditions and number of observations. .................. 29  
2.3 Balance of covariates. ..................................................... 32  
2.4 Description of variables on dentists, visits, and the market. .......... 34  
2.5 Descriptive statistics of variables on dentist, visit, and market character- 
istics. ............................................................................. 37  
2.6 Overtreatment recommendations per conditions. Number of observa- 
tions in parentheses. ....................................................... 38  
2.7 Regressions on a binary measure of overtreatment. ...................... 40  
2.8 Regression results on the likelihood (models 1 and 2) and amount 
(models 3 and 4) of overtreatment. ......................................... 41  
2.9 $p$-values of pairwise comparisons of means. ............................... 57  
2.10 $p$-values of pairwise comparisons of means. .............................. 57  
3.1 Patient utility depending on health state and treatment. ................. 67  
3.2 Physician payoffs & behaviour .............................................. 75  
3.3 Patient payoffs ................................................................. 75  
4.1 Equilibrium cutoff across settings. .......................................... 129  
4.2 Experimental conditions. $N = 360$. .......................................... 130  
4.3 The 23 possible strategy choices of the advisor. ........................... 133  
4.4 Parametrization ............................................................... 134  
4.5 Predictions (in ECU) ............................................................ 135  
4.6 Results I. ..................................................................... 139  
4.7 Results II. .................................................................. 140  
4.8 Field of study of participants. .................................................. 159
LIST OF TABLES

4.9 Regressions on consumer decision A. ........................................ 162
4.10 Regressions on consumer decision B. ................................. 163
4.11 Regressions on commission choice of firm A. ....................... 164
4.12 Regressions on commission choice of firm B. ....................... 164
4.13 Regressions on price choice of firm A. ............................... 165
4.14 Regressions on commission choice of firm B. ....................... 166
4.15 Regressions on the advisor cutoff choice. ......................... 167
4.16 Regressions on the advisor cutoff choice with data from all conditions. 168
List of Figures

2.1 Map of the canton of Zurich, Switzerland. Observations per district and population densities. ................................................. 31
2.2 Distribution of key variables. .................................................. 36
2.4 Cost estimates per number of fillings for overtreatment recommenda-
tions. ..................................................................................... 38
2.5 Distribution of overtreatment extent. ......................................... 42
2.6 Waiting time for the next possible appointment (days) by overtreat-
ment. ......................................................................................... 45
2.7 Distribution of diagnosis fees and points charged per diagnosis (n=180). 48
2.8 Points charged by SES (left) and patient information (right) (n=180). 48
2.9 Composition of points charged absolute (left) and relative (right) scales. 50
2.11 Distribution of point values, diagnosis costs and overtreatment. .... 54
2.12 Cumulative distribution of diagnosis fees and points charged by in-
formation status (n=180). .......................................................... 54
2.13 Appearance of the test patient and accessory in the SES variation. . 55
2.14 Screenshots of the mentioned web forum in the informed condition. . 55
2.15 Model fit. ................................................................................ 58
3.1 Benchmark without Market. ..................................................... 69
3.2 Market without asymmetric information. .................................... 71
3.3 Supply and demand. ................................................................. 76
3.4 The Credence Goods Market. ................................................... 78
3.5 Inefficiencies with fixed prices. ................................................ 80
3.6 Inefficiencies with equilibrium prices in (V) and (L/noV). ............... 85
3.7 The credence goods market with social insurance and high copayments. 105
3.8 The credence goods market with social insurance and low copayments. 106
4.1 Difference between commissions across conditions and matching groups. 145
4.2 Advisor reaction to commission differences. ................................. 147
4.4 Distribution of advisor cutoff choices across conditions. ................. 148
LIST OF FIGURES

4.5 0%-cutoff choices. .................................................. 148
4.7 Share of unbiased advisor cutoff choices depending on difference in commissions. .................................................. 149
4.8 Market welfare across conditions and matching groups. .......... 151
4.9 Consumer payoff across conditions and matching groups. ....... 151
4.10 Distribution of the difference of commission A and commission B choices across treatments. ............................... 157
4.11 Distribution of commission A choices across treatments. ......... 157
4.12 Distribution of commission B choices across treatments. ......... 158
4.13 Distribution of prices A choices across treatments. ............... 158
4.14 Distribution of prices B choices across treatments. ............... 158
4.15 Distribution of dictator donations to charity over all conditions. .. 159
4.16 Commissions and Prices over time. ............................... 160
4.17 Cutoff over time. .................................................. 160
Curriculum Vitae

Personal Details

Name: Felix C.H. Gottschalk
Place of birth: Darmstadt, Germany
Date of birth: October 18, 1986
Nationality: German

Current Position

2013–2017: Ph.D. Student, ETH Zurich, Switzerland
Department of Management, Technology and Economics (D-MTEC)
Chair of Risk and Insurance Economics
Supervisor: Prof. Dr. Wanda Mimra

Education

2010–2012: Master of Arts, Economics
University of Zurich, Switzerland

08/2008–01/2009: Study Abroad
Bilkent University Ankara, Turkey

2006–2009: Bachelor of Science, Economics
University of Mannheim, Germany

1997–2006: Abitur
Lichtenbergschule, Gymnasium der Stadt Darmstadt

Practical Experience

2011: SwissSign AG
2011: Ifo Institute for Economic Research
2010: Fresenius SE
2009: Deutsche Bank AG

Zurich, 2017