Overconfident Consumers in the Marketplace
(First Draft in Preparation for the Journal of Economic Perspectives)

Michael D. Grubb

March 2015

When consumers sign contracts, expectations about future usage of the product or service matter. For instance, the value provided by car insurance depends on how likely a consumer believes she is to file a claim, the value provided by a gym membership depends on how often a consumer anticipates going to the gym, and the value provided by a cellular phone contract depends on how many gigabytes of data a consumer anticipates using. The standard modeling paradigm assumes that consumers have rational expectations. Like assuming risk neutrality, assuming rational expectations is often expedient, drastically simplifying models and eliminating the need to measure beliefs. Yet consumer beliefs often deviate substantially from rational expectations, and a large literature in psychology shows that they are often wrong in systematic ways. This has important consequences for contract design, firm profits, and consumer welfare.

The term overconfidence is used broadly in the psychology literature, referring to both overoptimism and overprecision. Overconfident consumers who exhibit overoptimism overestimate their own abilities or prospects, either in absolute terms or in relative terms (the above average effect). Overoptimism can cause consumers to missforecast their average future consumption or to overestimate their ability to successfully navigate contract terms. Moreover, even appropriately optimistic individuals who are good at estimating means can still be very bad at estimating variances. Overconfident consumers who exhibit overprecision place overly narrow confidence intervals around forecasts, thereby underestimating uncertainty. Overprecision can cause consumers to underestimate the variance of future consumption.

Consumer overconfidence, whether it is overoptimism or overprecision, causes consumers to missforecast the costs and benefits of offered contracts and to make poor choices as a result. For instance, overoptimism about self-control is a leading explanation for why individuals overpay for gym memberships they do not use (DellaVigna & Malmendier, 2006). Similarly, overprecision is a leading explanation for why individuals systematically choose the wrong calling plans, racking up large overage charges for exceeding usage allowances in the process (Grubb, 2009; Grubb & Osborne, 2015).

While overconfidence necessarily leads to individually suboptimal choices, this leaves open three important questions about how consumer overconfidence alters equilibrium market outcomes: (1) First, what will firms do to exploit consumer overconfidence? (2) Second, what are the equilibrium welfare consequences of consumer overconfidence for consumers, firms, and society? (3) Third, what are the implications of consumer overconfidence for public policy?

1 Boston College. E-mail: michael.grubb@bc.edu
The small but growing economics literature studying overconfidence suggests the following answers: First, firms introduce complicated pricing features to contracts in order to exploit consumer overconfidence and these pricing features are robust to competition. Second, the welfare consequences for firms and consumers depend importantly on whether overconfident consumers over or undervalue contracts, consumer heterogeneity, and market structure. Third, while overconfidence may harm consumers, consumer protection policy should be undertaken with caution. Even seemingly-innocuous nudges to improve consumer decision making may harm consumers when firms’ equilibrium responses are taken into account.

**Firms complicate contracts to exploit overconfidence:**

In the standard common-prior framework, firms design contracts with two goals in mind: to create surplus from trade and to extract that surplus as profit. In other words, firms aim to bake a large pie and keep a large slice for themselves. Relaxing the common-prior assumption to allow for consumer overconfidence adds an important new element to a firm’s contract design problem. Unlike standard consumers, overconfident consumers may overvalue or undervalue offered contracts relative to the true value they deliver.

Firms naturally prefer consumers to overvalue contracts as much as possible because, as I discuss later, contract overvaluation allows firms both to charge more for contracts and to sell more contracts. Firms selling to overconfident consumers design contracts with an additional goal in mind: either to maximize the amount by which consumers overvalue contracts or to minimize the amount by which consumers undervalue contracts, depending on the situation. In other words, firms aim to bake a large pie, keep a large slice for themselves, and make the piece served to customers appear larger than it is.

To understand implications for pricing, it is important to distinguish two ways in which overconfidence may cause consumers to misvalue offered contracts. First, overconfident consumers may misforecast their future usage of services. For instance, a driver who is overconfident about his own driving safety may undervalue car insurance because he underestimates the likelihood of filing a claim (Sandroni & Squintani, 2007, 2013). When consumers misforecast future usage, firms have an incentive to distort marginal prices or quality provision to exploit the mistake (e.g. DellaVigna and Malmendier (2004), Eliaz and Spiegler (2006, 2008), Grubb (2009), and Heidhues and Kőszegi (2010)). Second, overconfident consumers may be overoptimistic about their own abilities of self-control, prospective memory, or attention. Such consumers overestimate their abilities to navigate contract terms to take advantage of contract benefits or avoid contract costs. For instance, an overconfident consumer may overvalue an offer because she overestimates the likelihood she remembers to mail in a rebate. Firms have an incentive to complicate their contracts with precisely those terms that consumers overestimate their own abilities to navigate (e.g. DellaVigna and Malmendier (2004), Holman and Zaidi (2010), and Grubb (2015)).

I first consider implications for pricing when overconfident consumers misforecast their usage and then return later to discuss the case when overconfident consumers overestimate
their ability to navigate contract terms. When an overconfident consumer underestimates her future usage, firms have an incentive to inflate marginal prices above marginal cost. As the consumer underestimates the chance of paying marginal fees, she also underestimates the cost of any increases in marginal prices. Hence inflating marginal prices either contributes to contract overvaluation or mitigates contract undervaluation.

For example, suppose that a consumer will use a unit of service with probability 1 but underestimates his usage, believing he will use the unit only with probability 1/2. Moreover, suppose for a moment that these probabilities are fixed, independent of pricing. What happens if the firm raises the marginal price by one dollar but lowers the contract fixed fee by the same amount? This change is actuarially neutral; the total cost of the contract and total payments to the firm are unchanged because the two price changes cancel out. However, from the consumer’s perspective, the expected price of the contract is 50 cents lower. This decrease in perceived cost follows because the consumer knows he receives the fixed fee discount for sure but underestimates the likelihood of paying the offsetting increase in marginal price. Thus raising the marginal price by one dollar increases the amount by which the consumer overvalues the contract by 50 cents.

The preceding example, in which usage is independent of pricing, leads to an implausible prediction: that the optimal marginal price is infinite. Ruling out such implausible predictions is in fact an important reason that the common prior assumption is standard in economics. However, a common prior is not necessary to rule out infinite marginal prices on consumer contracts because usage is endogenous. For instance, suppose that the contract in question is a car lease. If the marginal price per mile were too high, lessors would simply avoid driving their cars and the contracts could not be optimal. Thus, the size of the optimal pricing distortion is endogenously limited.

When consumers underestimate usage, contract marginal prices serve two roles. First, marginal prices affect consumers’ usage decisions and, hence, the surplus created by the sale of a contract. To maximize total surplus, firms would like to set marginal prices equal to marginal cost. Second, marginal prices serve as the stakes of a speculative bet about how much the consumer will use. When the consumer uses more than expected, the firm wins the bet, receiving the marginal payments in compensation. To maximize the contract overvaluation (or the fictional surplus) created by this speculative bet, firms would like to set marginal price as high as possible. The optimal marginal price is chosen to balance these two incentives, at the point above marginal cost where the additional gains from exploiting consumers’ mistaken usage forecasts are offset by the additional costs from distorting consumers’ true usage choices.

When overconfident consumers overestimate rather than underestimate future usage, the logic is similar, but the direction of predicted price distortions reverses: firms have an incentive to discount marginal prices below marginal cost. As consumers overestimate the chance of paying marginal fees, they also overestimate the value of any discount to marginal

---

2 Or to commitment prices that overcome self-control problems when they are present (DellaVigna and Malmendier, 2004).
prices. Hence discounting marginal prices either contributes to contract overvaluation or mitigates contract undervaluation.\(^3\)

Will overconfidence distort marginal prices up or down? The answer depends on whether overconfident consumers under or over-estimate future usage. That depends in turn on context and the nature of overconfidence. For instance, Svenson (1981) found that 83% of his American subjects placed themselves among the top 30% of their fellow lab subjects in driving safety. A natural conjecture is that such over-optimism about relative driving safety will lead drivers to underestimate the likelihood of filing a car-insurance claim. If so, insurance companies should distort the price of filing a claim upwards, which means raising deductibles.

The prediction that overconfidence increases deductibles is specific to insurance markets. In many cases, predictions about the consequences of overconfidence may be equally context specific. However, overoptimism about self-control and overprecision each lead to more general predictions. First, consider overoptimism about self-control. DellaVigna and Malmendier (2004) define investment goods as those that require costly effort at the point of consumption but yield future benefits. Similarly, leisure goods are those that yield an immediate payoff upon consumption but are costly later. For example, a gym workout is an investment good but credit card borrowing is a leisure good. Importantly, individuals with a self-control problem will under-consume investment goods (working out too little) but overconsume leisure goods (spending too much on credit cards). Moreover, those who are overoptimistic about their own levels of self-control will underestimate these problems. Thus, they overestimate their consumption of investment goods (overestimating their gym attendance) but underestimate their consumption of leisure goods (underestimating credit card borrowing). As a result, DellaVigna and Malmendier (2004) predict that marginal prices of investment goods will be discounted below marginal cost but that marginal prices of leisure goods will be inflated above marginal cost.

DellaVigna and Malmendier’s (2004) prediction is consistent with the fact that many gyms do not charge per-visit fees to members, despite per-visit marginal costs of $3 or more (DellaVigna & Malmendier, 2004). It is also consistent with evidence that high interest rates on credit card debt do not merely reflect the costs of default but are substantially above marginal cost: Ausubel (1991) finds that banks are able to resell credit card debt for an average premium of 20%.

Next, consider how firms may exploit overprecision. A car lessor who exhibits overprecision may correctly forecast her median mileage but underestimate the variance of her driving needs around the median. If mile q is below median total mileage, such a consumer overestimates her likelihood of driving it. If mile q is above median total mileage, such a consumer underestimates her likelihood of driving it. Therefore, if Q is median mileage, the lease contract should price the first Q miles below marginal cost and all later miles above marginal cost. If consumers can freely dispose of miles (for instance by lending the car to a friend for a weekend trip) then the mileage fee for the first Q miles should not be reduced below zero. While

---

\(^3\) The result that marginal prices are distorted upwards (downwards) when usage is underestimated (overestimated) may be found in DellaVigna and Malmendier (2004) and Grubb (2009).
the optimal contract will be fully nonlinear, Grubb (2009) shows that a three-part tariff is a good approximation. A three-part tariff charges a fixed fee for an included allowance of units followed by a constant marginal price for additional units. This coincides exactly with observed car leasing contracts, which typically offer an allowance of 36,000 miles with a three-year lease and charge 15 cents per mile for additional mileage. Overconfidence may explain the structure of car lease contracts as well as three-part tariffs in a variety of other settings. Table 1 gives some examples below:

<table>
<thead>
<tr>
<th>Product or Service</th>
<th>Consumers must forecast</th>
<th>Example Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car lease</td>
<td>Mileage</td>
<td>Fixed price for 36 months and 36,000 miles plus 15 cents per additional mile. (Toyota, June 23, 2014.)</td>
</tr>
<tr>
<td>Smart phone service</td>
<td>Data usage</td>
<td>$60/month for unlimited talk and text and 1GB of data plus $15 per additional 500MB of data. (Verizon, June 23, 2014)</td>
</tr>
<tr>
<td>Credit card.</td>
<td>Loan duration</td>
<td>Introductory offer with an initial balance transfer: Banks charge a balance transfer fee that is independent of the loan duration. Then there are zero financing charges for the first six months but a high interest rate thereafter. (e.g. Chase Slate card, June 23, 2014)</td>
</tr>
</tbody>
</table>

My discussion thus far has followed the literature’s focus on how overconfidence affects the prices specified by contracts. However, contracts often specify aspects of quality as well as price, and overconfidence should affect these terms as well. I conjecture that the preceding conclusions about marginal price distortions extend naturally to quality distortions. If a consumer overestimates the likelihood of using the q\textsuperscript{th} unit, then she will overvalue an increase in its quality just as she overvalues a discount to its marginal price. Thus, the firm should overinvest in quality of the q\textsuperscript{th} unit. Similarly, if a consumer underestimates the likelihood of using the q\textsuperscript{th} unit, then he will underestimate the cost of a quality reduction and the firm should underinvest in its quality. Just as overprecision leads firms to charge zero marginal price up to a usage allowance followed by high marginal charges thereafter, overprecision could also lead firms to offer high quality service up to a usage allowance followed by low quality thereafter.

These conjectures about optimal product quality may explain why T-Mobile offers cellular data plans that include an allowance of data at high speed but provide additional data beyond the allowance at slow speed. If overconfident consumers underestimate the variance of their total data usage, then inefficiently high speed should be provided up to an allowance after which inefficiently low speed should be provided. They may also explain coverage limits on car insurance. If overconfident consumers underestimate the variance of their accident losses then they underestimate the likelihood of large losses and the firm should respond by reducing coverage quality for large losses. Coverage limits accomplish this objective. Table 2 summarizes both examples.
Table 2: Product quality distortions.

<table>
<thead>
<tr>
<th>Product or Service</th>
<th>Consumers must forecast</th>
<th>Example Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart phone service</td>
<td>Data usage</td>
<td>$50/month for unlimited talk and text and 1GB of data at high speed. Additional data provided at slow speed. (T-mobile, June 23, 2014)</td>
</tr>
<tr>
<td>Car insurance</td>
<td>Losses</td>
<td>Premium, deductible, and coverage limit beyond which 0% of additional loss is covered. (Liberty Mutual, 2014)</td>
</tr>
</tbody>
</table>

Now return to the case where overconfident consumers overvalue contracts because they overestimate their abilities to navigate contract terms. Taking full advantage of a contract often requires follow through, remembering and then completing a costly task in the future, whether it be mailing in a rebate or canceling service once an introductory rate expires. Any contract term offering a future benefit after a costly task is completed presents two challenges. First, such terms create a *memory hurdle*. Follow through requires prospective memory, the ability to remember to take planned actions. Overconfidence about prospective memory leads consumers to overestimate the likelihood of remembering to complete tasks, and overvalue the contract to firms’ benefit (Holman & Zaidi, 2010). Second, such terms serve as a *self-control trap* that exploits overconfidence about self-control (DellaVigna & Malmendier, 2004). If the cost and benefit are correctly balanced, an overconfident consumer will anticipate having the self-control to complete the task in a timely fashion but in practice will either procrastinate and delay completing the task or fail to complete it altogether (O'Donoghue & Rabin, 1999). This is a second reason that overconfident consumers overvalue the contract to the firm’s benefit. Table 3 lists a few of the many contract terms that require follow-through, creating memory hurdles and self-control traps. In each case, while there may be other explanations, firms may include the contract terms simply to exploit overconfidence about prospective memory, overconfidence about self-control, or both.
Table 3: Barriers to follow-through: Memory hurdles and self-control traps

<table>
<thead>
<tr>
<th>Contract term</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mail-in Rebates</td>
<td>Overconfident consumers may overestimate the likelihood of remembering to mail-in a rebate or of having the self-control to avoid procrastinating until the deadline is missed.</td>
</tr>
<tr>
<td>Free trials or teaser rates with switching or cancelation costs</td>
<td>When a free trial or teaser rate expires, paying a switching or cancelation cost may avoid the newly higher fees. Overconfidence about self-control or prospective memory both lead one to overestimate the likelihood of switching or canceling and achieving the benefits rather than procrastinating or forgetting.</td>
</tr>
<tr>
<td>Auto-renewal</td>
<td>Auto-renewal makes switching or quitting relatively more costly and can lead to overestimation of switching or quitting.</td>
</tr>
<tr>
<td>Bonus cash back (quarterly activation required)</td>
<td>Some credit cards offer additional cash back conditional on customers actively opting in each quarter. Overconfident consumers may overestimate the likelihood of opting in.</td>
</tr>
</tbody>
</table>

Beyond remembering tasks and exercising self-control, navigating contract terms as intended often also requires consumers to pay attention. For instance, a checking account customer who does not pay attention to her account balance can easily and unintentionally pay her bank $35 to buy a cup of coffee at Starbucks with her debit card simply because she does not realize that her account balance is exhausted and that an overdraft fee applies. This in turn leads to bill shock when a notice of overdraft fees arrives in the mail. (Stango and Zinman (2014) find that “60% of overdrafters reported overdrafting because they ‘thought there was enough money in my account’”.) While card-processing terminals could be designed to ask a consumer, “Overdraft fee applies. Continue Yes/No?” they do not. The transaction-processing fee varies dramatically between zero and $35 but consumers must actively keep track of their balance to know which fee applies when taking out their wallets at Starbucks. Importantly, if banks designed overdraft fees simply to reflect the marginal cost of extending credit then they should want to disclose overdraft fees at the point of sale to increase the social efficiency of consumers’ choices. Instead, the lack of transparency creates an attention hurdle, which consumers who are overoptimistic about their attention levels overestimate the likelihood of clearing, leading to contract overvaluation rather than increased efficiency.

Overdraft fees are an example of surprise penalty fees, which firms charge for crossing a consumption threshold but do not disclose at the point of sale. Surprise loyalty discounts, which lower rather than raise marginal price after crossing a consumption threshold, are also attention hurdles. For instance, the last flight required to achieve elite status is effectively discounted by the value of elite rewards. Yet a frequent flyer must keep track of his mileage balance to know whether the implicit discount applies to his current trip (or whether a mileage run is required

---

(Sharkey, 2005)). Consumers who are overoptimistic about their attention levels will underestimate the likelihood of paying surprise penalty fees but overestimate the likelihood of collecting surprise loyalty discounts. In either case, consumers overvalue contracts with attention hurdles to firms’ benefit (Grubb, 2015). More examples are in Table 4.

<table>
<thead>
<tr>
<th>Product or Service</th>
<th>Consumers must attend to</th>
<th>Source of returns to attention.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checking account</td>
<td>Account balance</td>
<td>Debit card transaction fee rises from $0 to $35 when balance falls to zero.</td>
</tr>
<tr>
<td>Credit card</td>
<td>Account balance</td>
<td>Crossing credit limit triggers higher finance charges or over-limit fee.</td>
</tr>
<tr>
<td>Smart phone service</td>
<td>Data usage</td>
<td>Marginal price of data rises from $0 to $15 per 500MB after 1GB of usage.</td>
</tr>
<tr>
<td>Frequent flyer program</td>
<td>Mileage balance</td>
<td>Perks are awarded upon crossing various mileage thresholds.</td>
</tr>
</tbody>
</table>

More generally, if overconfident consumers over estimate their ability to clear hurdles and avoid traps in contract terms then this makes it profitable to add such hurdles and traps to contracts. If consumers overestimate their ability to avoid fees of some kind then firms should add them to contracts. If consumers overestimate their ability to collect discounts, such as mail in rebates, then firms should add them to contracts. While overconfidence is prevalent, consumers may also be underconfident about their abilities to complete some tasks. Including such tasks in contract terms would only lead consumers to undervalue contracts surplus to firms’ detriment. Thus, in equilibrium, we should only expect firms to complicate contracts with the sorts of tasks that consumers are overconfident they can complete.

**Complex pricing is robust to competition:**

Competition does not eliminate complex pricing designed to exploit overconfidence. Instead, its primary effect is to lower fixed fees.\(^5\) When competition forces firms to offer more value to consumers, lowering fixed fees is optimal because, unlike lowering marginal fees or adjusting other contract terms, this does not diminish either the true surplus generated by a contract or the amount by which consumers overvalue contracts. (Importantly, a firm benefits

---

\(^5\) I assume one-stop shopping or exclusive contracting, where consumers buy from only one firm. For non-exclusive contracts, competition ensures that marginal prices do not exceed marginal cost (Gottlieb 2008). Gabaix and Laibson (2006) show that competition need not give firms an incentive to educate or de-bias consumers.
from contract overvaluation whether it is a monopolist, an oligopolist, or one of many perfect competitors).

In fact, in a market for a homogeneous good, a firm that priced at cost would fail to capture any market share. An example adapted from Grubb (2009) illustrates this point nicely. Suppose that cellular service providers have marginal costs of 10 cents per minute and fixed costs of $40 per customer. Consumers value each minute of calling at 45 cents up to some satiation point, which they do not learn until after signing a contract. Consumer satiation points will either be 100, 400, or 700 minutes, each equally likely. However, overconfident consumers underestimate their uncertainty and believe they will be satiated at 400 minutes with probability 1. Cost based pricing yields true and perceived utility of 400(.45-.10)-40 = $100 to consumers. However, another firm could charge $60 for an allowance of 400 included minutes followed by 45 cent per minute charges for additional calling. In reality, this contract shifts $25 of surplus from consumers to firms, yielding expected profits of $25 and expected utility of $75. Absent overconfidence, no one would choose the contract. However, overconfident consumers perceive their expected utility to be higher, at 400(.45)-60 = $120, overvaluing the contract by $45, and will choose the contract over cost-based pricing.

Figure 1 shows the two pricing schemes graphically. An overconfident consumer expects to be in the shaded region of the figure where the contract line is in bold and the three-part tariff is below cost. The firm recognizes, however, that two thirds of the time the consumer will actually end up in the non-shaded regions where three-part tariff revenues are above cost. The contract serves as a bet about which region the consumer will end up in, and due to consumer overconfidence both firm and consumers believe they win on average.
Yes, but are consumers really overconfident?

Why would they be overconfident? Can’t a standard model explain these pricing schemes?\(^6\)

Many readers may be asking themselves these questions,\(^7\) so it is worth pausing to address them. The psychology literature provides clear evidence of overprecision. A typical study might pose the following question to a group of subjects: "What is the shortest distance between England and Australia?" Subjects would then be asked to give a set of confidence intervals centered on the median. A typical finding is that the true answer lies outside a subject's 90% confidence interval more than 50% of the time (Alpert & Raiffa, 1982; Lichtenstein, Fischhoff, & Phillips, 1982). Similar results have been found among managers answering industry or firm related questions (Russo & Schoemaker, 1992), finance professionals forecasting exchange rates (Önkal, Yates, Simga-Mugan, & Öztin, 2003), and software project managers predicting software development effort (Jørgensen, Teigen, & Moløkken, 2004). Overprecision is not an artifact of the questions selected by researchers (Klayman, Soll, González-Vallejo, & Barlas, 1999) or of noise in subjects’ interval reporting (Soll & Klayman, 2004).

The psychology literature also provides strong evidence for overoptimism, although the evidence is not as strong as previously thought. For instance, in one of the most cited studies of overoptimism, Svenson (1981) finds that 83% of American undergraduates and 51% of Swedish undergraduates place themselves in the top 30% of their fellow lab subjects in driving safety. Researchers have long interpreted this finding as evidence that both Americans and Swedes are overoptimistic about their driving safety. However, Benoît and Dubra (2011) show that, while Svenson’s (1981) results are consistent with overoptimism for both Americans and Swedes, rational expectations can only be rejected for Americans.\(^8\) Thus, Americans are indeed overoptimistic about driving safety but the international scope of the problem is unclear.

In contrast to overprecision, the psychology literature shows that overoptimism is highly context specific.\(^9\) Weinstein (1980) finds that overoptimism is highest when individuals have a high sense of control over outcomes and a positive outcome is very desirable. For example,

---

\(^6\) This question is adapted from the titles of Spiegler’s (2011b) article and chapter 13 of Spiegler’s (2011a) book, which include the question, “But Can’t We Get The Same Thing With A Standard Model?”.

\(^7\) Another common question is “Doesn’t overconfidence go away with learning?” With appropriate feedback, learning can mitigate overconfidence (Bolger & Önkal-Atay, 2004) but field studies of consumer choice show learning is no panacea. Choices may improve slowly (Grubb & Osborne, forthcoming), lessons may be forgotten (Agarwal, Driscoll, Gabaix, & Laibson, 2013), and individual performance avoiding traps in contracts peaks in middle age (Agarwal, Driscoll, Gabaix, & Laibson, 2009).

\(^8\) Benoît and Dubra (2011) show that up to 2k individuals may place themselves in the kth decile (e.g. up to 20% in the top 10%, up to 40% in the top 20%, and so on) without violating rational expectations. That I am aware of, no one has yet undertaken a survey to determine which results in the rest of the literature about the above average effect survive Benoît and Dubra’s (2011) critique.

\(^9\) Early work on overoptimism about trivia knowledge documented the hard-easy effect, individuals were overly optimistic about answering hard questions but overly pessimistic about answering easy questions (Lichtenstein, Fischhoff, and Phillips (1982). Later work argues that the hard-easy effect may be a statistical artifact of how questions are classified as hard or easy (Klayman, Soll, González-Vallejo and Barlas, 1999) or may arise because individuals observe question difficulty imperfectly, so rationally do not fully update confidence to reflect difficulty (Moore and Healy, 2008). Thus difficulty may not be as important an element of context as previously believed.
Weinstein (1980) finds that subjects are overoptimistic about avoiding lung cancer but not about avoiding cancer in general. As a result, studies of overoptimism in economically relevant contexts are especially important. One such study is Ericson (2011), which shows that individuals are overoptimistic about their prospective memory in a context highly relevant to the contract memory hurdles listed in Table 3. Ericson’s (2011) subjects anticipated claiming a payment in six months time with an average probability of 76% (inferred from their choices), but only 53% actually claimed the payment.

There is a variety of possible explanations for overconfidence. Van den Steen (2004, 2011) explains overconfidence by combining unbiased but heterogeneous prior beliefs with selection effects. Overprecision may result from anchoring and adjustment (Tversky & Kahneman, 1974). Even if beliefs are unbiased, a consumer who mistakenly evaluates a contract’s expected cost by computing the price of his expected usage, $P(E[\text{usage}])$, rather than his expected bill, $E[P(\text{usage})]$ will act as if his usage forecasts are overprecise. This mistake is an example of the flaw of averages, which Savage (2009) argues is common. Overoptimism has been explained as resulting from egocentrism (Kruger, 1999), incompetence (Kruger & Dunning, 1999), confirmatory bias (Koriat, 1980; Rabin & Schrag, 1999), or self-serving biases that optimally improve mental health (Brunnermeier & Parker, 2005; Taylor & Brown, 1988), personal motivation (Benabou & Tirole, 2002), or performance (Compte & Postlewaite, 2004). Fortunately, identifying the correct explanation(s) for overconfidence is not necessary to measure overconfidence or model its effects.

Moving beyond the psychology lab, what evidence is there that consumer overconfidence is important in the marketplace? Consumer beliefs can be inferred from contract choices and compared to later usage to identify bias. For instance, using a structural model, Grubb and Osborne (2015) estimate that cellular phone consumers are overprecise, underestimating the noise in their forecasts of their own future demand for calls by 62%. Contract choice and consumption data are also useful without imposing so much structure. For instance, Ausubel (1999) finds that, when responding to credit card offers, “consumers are at least three times as responsive to changes in the introductory interest rate as compared to dollar-equivalent changes in the post-introductory interest rate.” This is inconsistent with the standard model but consistent with overoptimism about the likelihood of repaying or refinancing debt in time to avoid paying post-introductory rates, perhaps due to overoptimism about self-control (Ausubel & Shui, 2005; Heidhues & Kőszegi, 2010). In a similar vein, DellaVigna and Malmendier (2006) show that, at the New England health club they study, users who choose a monthly membership could have saved an average of more than 40% by foregoing a membership and paying per-visit. This is inconsistent with the standard model but consistent with overoptimism about gym attendance due to overoptimism about self-control.

Often the largest challenge is not rejecting the standard model but determining which the right alternative is. For example, DellaVigna and Malmendier’s (2006) finding of overpayment for gym memberships rejects the standard model but can be explained equally well by a self-control problem with or without overconfidence. DellaVigna and Malmendier (2006) must marshal additional evidence that gym members overestimate the likelihood of canceling their
memberships to conclude that gym users are overconfident. Even then, the data do not distinguish between complementary forms of overconfidence, overoptimism about self-control and overoptimism about prospective memory.

It remains a challenge for the literature to directly test the relevance of overconfidence across a broader array of markets. However, until this challenge is overcome, the existing evidence of overconfidence suggests that we should be open minded about the role of consumer overconfidence whenever consumers sign contracts. This is particularly true in markets exhibiting the sort of pricing features described in Tables 1-4, which can all be used to exploit overconfidence. While rational expectations based price discrimination models can explain almost any pricing pattern, such as rebates (Narasimhan, 1984), menus of three part tariffs (Grubb, 2009), or (allowing for inattention) surprise penalty fees (Grubb, 2015), this does not mean they are the best or only important explanations. (Spiegler (2011a, 2011b) discusses model choice absent distinguishing empirical evidence.)

What are the equilibrium welfare consequences of consumer overconfidence?

I now turn to address the second question of this essay, and explain how the welfare consequences for firms and consumers depend importantly on consumer heterogeneity, market structure, and on whether overconfident consumers over or undervalue contracts.

Equilibrium contracts distort allocations and lower social welfare on the intensive margin:

The combination of consumer overconfidence and the resulting equilibrium pricing by firms typically distort consumer quantity choices. For instance, consumer overprecision in forecasting future demand leads to three-part tariffs, which sharply depart from marginal cost pricing (Grubb 2009). This distorts consumers’ consumption decisions relative to efficient quantity choices made under marginal cost pricing. In market settings in which outcomes would be efficient absent overconfidence, this naturally implies that overconfidence lowers social welfare.

A caveat to the preceding prediction is that overconfidence could be welfare improving if the distortions used to exploit overconfidence are countervailing to other distortions already present in the marketplace. Existing results already show that other forms of boundedly rational behavior by consumers may raise social welfare in equilibrium. For instance, Handel (2013) shows consumer inertia can raise social surplus by preventing insurance markets unraveling in a Cutler and Zeckhauser (1998) style “death spiral” due to adverse selection. Moreover, Grubb and Osborne (2015) show providing bill-shock alerts to compensate for consumer inattention may...

---

10 Particularly when beliefs cannot be inferred from contract choices, directly surveying consumers about their beliefs may be helpful, as suggested by Manski (2004) and implemented by DellaVigna and Malmendier (2006).

11 In fact, Grubb (2009) rules out his rational expectations based explanation for three-part tariff menus in the context of cellular phone service based on observed calling patterns.

12 I measure welfare of an overconfident consumer as expected utility with respect to the true distribution of outcomes. See Bernheim (2009) and Beshears, Choi, Laibson, and Madrian (2008) for alternative views on the best approach.
lower social surplus. In principal, consumer overconfidence could also be socially beneficial in some settings.

**Overconfident consumers may over or under value equilibrium contracts.**

Will overconfident consumers overvalue or undervalue equilibrium contract offers? The answer affects whether overconfidence raises or lowers industry profits, harms or helps consumers, and whether overconfidence expands or contracts the market. If overconfident consumers are overoptimistic about their own levels of self-control, prospective memory, or attention, and hence their ability to take advantage of a contract’s potential value, then they overvalue contracts (DellaVigna & Malmendier, 2004; Grubb, 2015; Holman & Zaidi, 2010). For example, an individual overconfident about his own self-control overestimates his future gym attendance and hence his value of a gym membership. In contrast, if overconfident consumers misforecast their future usage because they misforecast their future valuations for the service then they might under value contracts. Undervaluing car insurance due to overoptimism about driving ability is a clear example. Less obvious is the fact that overprecision in demand forecasts can also lead to contract undervaluation (Grubb, 2009).

**Contract overvaluation benefits firms but hurts consumers and society in competitive markets**

How do overconfident consumers who overvalue contracts fare compared to a counterfactual world in which they have rational expectations? If the market pass-through rate is less than or equal to one, as it is in a perfectly competitive market, then contract overvaluation leads price and quantity of contracts to rise, benefiting firms. Existing customers are made worse off by the amount of the price increase. New customers lured into the market by their overvaluation of contracts are also worse off because their true value for a contract is less than the price. Luring new customers into the market may improve social welfare (at consumers’ expense) if it offsets inefficiently low sales otherwise resulting from firms’ market power. In a perfectly competitive market that lacks any other source of inefficiency, however, overconfidence is unambiguously bad for social welfare.

Figure 1 depicts these consequences of overconfidence graphically for a perfectly competitive market, via the use of two tricks that I now explain.\(^{13}\) Contracts can typically be described by a fixed payment \(P\) and a vector of additional terms \(p\). For instance, an insurance contract charges a fixed premium \((P)\) but also specifies coverage limits, deductibles, and co-insurance rates \((p)\). Similarly, a wireless calling plan charges a monthly fee \((P)\) but also specifies a data allowance, overage rate, roaming fees, and other terms \((p)\). My first trick is standard: Given rational expectations (RE), I focus on the fixed payment \((P_{RE})\) as the “price” of the contract. Additional terms I treat like any other dimensions of product quality; I fix them at their equilibrium values \((p_{RE})\) and suppress them from the graphical analysis. One can think of the contract

\(^{13}\) Previous graphical treatments of welfare given distinct curves describing demand and consumer valuations include Bernheim and Rangel (2009), Madrian (2014), and Spinnewijn (2014).
simply as a product that delivers true expected utility $U$ for price $P_{RE}$. Given rational expectations, this yields familiar demand and supply curves for contracts in Figure 1.

Analyzing demand and supply curves on the same figure for the case of overconfidence requires a second trick. As discussed earlier, consumer overconfidence will lead firms to complicate a contract’s additional terms to include memory hurdles, self-control traps, or attention hurdles, or to create three-part tariffs or quality distortions. To make these more complicated contracts comparable to those offered under rational expectations, I partition the contract’s fixed payment into two parts. These are the “price” of the contract ($P_{OC}$) and a second fixed fee $F$ to be included with the additional terms $p_{OC}$ that are suppressed from the analysis. The second fixed fee is chosen so that, gross of the price $P_{OC}$, the additional contract terms offered to overconfident consumers ($F, p_{OC}$) yield the same true expected utility $U$ as the additional contract terms ($p_{RE}$) offered under rational expectations. The result is that, regardless of whether one considers the case of rational expectations or overconfidence, one can treat offered contracts simply as products that deliver true expected utility $U$. Hence Figure 1 plots contract demand and supply curves under both rational expectations and consumer overconfidence.

---

Figure 1: Consequences of contract overvaluation due to overconfidence in a competitive market.

---

14 For instance, given rational expectations, a wireless plan might charge $50 per month for unlimited talk time. In this case the “contract” is “unlimited talk” and its “price” is $50. Given overconfidence, a wireless plan might charge $40 per month for 500 included minutes and charge $0.50 per minute thereafter. This might lead consumers to make fewer calls, thereby losing $5 in value from foregone calls, and to pay $10 in additional calling charges. In this case the “contract” is “500 include minutes and $0.50 thereafter with a $15 refund” and its “price” is $55. The $15 adjustment to the definition of the “contract” under overconfidence ensures that, in either case, the “contract” offers comparable value. The difference in price between the overconfident ($55) and rational expectations contracts ($50) therefore reflects the true difference in their offered values.
Demand: Given rational expectations, a consumer’s value for a contract is the difference between its utility and her outside option. The rational expectations demand curve is downward sloping, rather than a horizontal line, because consumers have heterogeneous outside options. Given overconfidence, the demand curve is shifted upwards by $\Delta D$, the amount that consumers overvalue the offered contract.

Supply: Overconfidence induces firms to charge distortionary marginal prices with the aim of increasing contract overvaluation. These lead to a deadweight loss for each customer served resulting from distortions on the intensive margin. The supply curve given overconfidence is shifted up by this amount, $\Delta C$, which reflects firms’ increased cost of delivering the same true utility $U$ to consumers. (If contract overvaluation results from overoptimism about ability to navigate contract terms, then this shift must be smaller than the shift in demand, as otherwise the distortions would not be optimal for firms.)

Equilibrium: Overconfidence causes the equilibrium price to rise from $P_{RE}$ to $P_{OC}$, and equilibrium quantity to increase from $Q_{RE}$ to $Q_{OC}$. While the overconfident demand curve determines price and quantity with overconfidence, the rational expectations demand curve is still relevant for welfare calculations because it describes consumers’ true valuations. Firm profits increase because price increases more than cost. Dead weight loss to society includes waste on the intensive margin, captured by the upward shift in the supply curve, and waste on the extensive margin, due to inefficiently high contract sales.

There are two losses to consumer surplus. First is the area shaded light gray, corresponding to the real price increase for existing customers. Second is the area shaded dark gray, corresponding to the amount new customers pay above their true valuations. The magnitude of the two effects depends on the pass-through rate, which measures the fraction of any per-unit cost increase that would be passed-through to consumers as a price increase. Given a constant market pass-through rate, $\rho$, the true price increase for a contract offering utility $U$ is simply a weighted average of the shifts in demand and supply:

$$P_{OC} - P_{RE} = (1 - \rho)\Delta D + \rho \Delta C.$$  

Due to contract overvaluation, however, consumers perceive a price decrease of $\rho(\Delta D - \Delta C)$, which increases sales according to the elasticity of demand.\(^{15}\) A high pass-through rate protects infra-marginal consumers from their overconfidence by limiting price increases but, for the same reason, leads more consumers on the extensive margin to buy at prices above their true valuations.

When the pass-through rate is zero (supply is perfectly inelastic), firms raise the price of a contract offering utility $U$ by the amount consumers overvalue the contract, $\Delta D$. Consumers perceive no effective change in the price. Firm profits per customer, however, rise by the price increase less the deadweight loss from distortions on the intensive margin, $(\Delta D - \Delta C)$. When

\(^{15}\) In similar analyses using the market pass-through rate, Agarwal, Chomsisengphet, Mahoney, and Stoebel (Forthcoming-a, Forthcoming-b) describe the effect of regulating hidden fees on up-front prices, and Farrell (2008) describes the welfare loss when consumers underestimate aftermarket costs.
the pass-through rate is positive, firms pass fraction $\rho$ of these potential profits back to consumers. Consumers perceive this to be a price cut of $\rho(\Delta D - \Delta C)$ when in truth it is merely reduces the price increase to $(1 - \rho)\Delta D + \rho\Delta C$. When the pass-through rate is one (supply is perfectly elastic), firm profits are always zero and consumers become residual claimants of social surplus. The price increase, $\Delta C$, is equal to the deadweight loss on the intensive margin.

If firms have market power, Figure 1 does not apply. However, the expressions which relate true and perceived price changes to the market-pass through rate $\rho$ are the same. Thus, given a market pass-through rate less than one, contract overvaluation leads to a true price increase but additional sales due to a perceived price drop, which is good for firms but bad for consumers, whether firms have market power or not. A difference from the competitive case is that the additional contract sales may be socially valuable if they offset otherwise inefficiently low sales due to market power. Finally, and unlike in a competitive market, the market pass-through rate can exceed one with market power. In this case, some effects of overconfidence can be reversed: With a sufficiently high pass-through rate, contract overvaluation benefits infra-marginal consumers by lowering prices.

**Contract undervaluation harms firms and society but may benefit some consumers in competitive markets**

When overconfident consumers misforecast their future valuations for services, they may undervalue contracts. In this case, overconfidence shifts demand downwards relative to rational expectations and depresses the market price. This is bad for firms but good for infra-marginal consumers who enjoy the lower price. On the extensive margin however, some consumers with true contract valuations above the price will stop buying to their own detriment. Thus, the overall effect of overconfidence on consumer surplus may still be negative. Moreover, absent any source of inefficiency other than overconfidence, overconfidence remains unambiguously bad for social welfare in a perfectly competitive market.

Figure 2 is similar to Figure 1 but depicts the case in which overconfidence depresses the market price for a contract delivering true expected utility $U$. Overconfidence causes the equilibrium price to fall from $P_{RE}$ to $P_{OC}$, and equilibrium quantity to fall from $Q_{RE}$ to $Q_{OC}$. Firm profits fall with demand. Dead weight loss to society includes the area between the supply curves (waste on the intensive margin) and the lost surplus due to inefficiently low contract sales (waste on the extensive margin). There are two changes to consumer surplus. First, infra-marginal customers benefit from the true price decrease, benefiting by the area shaded light gray. Second, consumers on the extensive margin lose the area shaded dark gray because their undervaluation causes them to forego purchasing contracts with true expected value above their price. The net effect for consumer surplus may be positive or negative.
Does competition (partially) protect consumers from overconfidence?

Suppose that overconfidence leads consumers to overvalue contracts. Then in competitive markets, as discussed above, overconfident consumers are worse off than they would be in a counterfactual world with rational expectations. Thus, competition does not completely protect consumers from overconfidence. However, one may still ask whether competition partially protects consumers from overconfidence. First, does increased competition benefit overconfident consumers? Second, consider policies that directly reduce contract overvaluation, for instance by limiting the use of contract terms described in Tables 1-4, or by reducing overconfidence. Does competition reduce the consumer harm from overconfidence and thereby limit the potential benefit from such policies?

A common assumption in the literature is that competition increases the market pass-through rate and that there is full market coverage, meaning that all consumers buy and are infra-marginal. Under this assumption, the answer to both questions is “yes”. Infra-marginal consumers always benefit from lower prices, so benefit from competition. Moreover, if competition raises the pass-through rate it also limits the amount by which contract overvaluation raises prices, and hence limits the cost of overconfidence to infra-marginal consumers. Thus competition reduces consumers’ benefit from policies that directly reduce contract overvaluation.

Perhaps surprisingly, however, when the full-market coverage assumption is relaxed, the answer to both questions is “not necessarily”. It is possible both for competition to harm the average overconfident consumer and for competition to increase the returns to consumer protection policies that directly reduce contract overvaluation. The reason is that competition

---

16 For example, DellaVigna and Malmendier (2004) assume marginal cost is constant and all consumers have the same outside option. Thus all consumers buy and pass-through is 0 for monopoly but 1 for perfect competition.
and low prices have very different welfare consequences for marginal consumers than for infra-marginal consumers. In particular, while lower prices always benefit infra-marginal consumers who pay less, lower prices can actually harm marginal consumers who overvalue contracts. As a result, competition can have counter-intuitive effects on consumer welfare.

Accounting for contract overvaluation, consumers just indifferent to buying are actually strictly worse off buying than not. These consumers would be protected by high prices that keep them out of the market. Reduced prices due to competition can tempt them to buy when they should not. It is therefore possible that policies which successfully increase competition, lower prices, and expand sales also lower consumer surplus.

Moreover, while competition that raises the pass-through rate reduces the true price increase that results from contract overvaluation it simultaneously increases any perceived price decrease due to contract overvaluation. Thus while competition mitigates the cost of overconfidence to infra-marginal consumers, it increases the costs to marginal consumers, more of whom are lured into the market to pay more than their true valuations. Therefore, while we might say that competition partially protects infra-marginal consumers from overconfidence, we might also say that market power partially protects marginal consumers from overconfidence (by pricing them out of the market). If demand is elastic then the latter effect may be important, and competition may increase returns to policies that reduce contract overvaluation.

**Pooling and cross-subsidization:**

Whether or not competition limits the harm to consumers from overconfidence, thus far in my discussion it seems that the harm is limited when the pass-through rate is high and market demand is very inelastic. This is only true, however, when all consumers are equally overconfident.

Consider an example inspired by Bubb and Kaufman (2013) and Grubb (2015). There is a competitive market for checking accounts with a perfectly elastic supply of accounts. A bank’s marginal cost of processing an overdraft transaction is zero but the typical fee is $35. All consumers opening checking accounts believe they will pay attention to their balance and avoid overdraft fees. Half of them have rational expectations and do avoid fees. Half, however, are overconfident about their attention and in fact incur $100 in overdraft fees due to inattention.

Given the assumption that supply is perfectly elastic, banks cannot profit from consumer overconfidence. If all consumers were overconfident, annual fees for checking accounts would be $100 below cost and banks would break even with overdraft fee revenue. The market pass-through rate of one ensures that 100% of the overdraft fees, which overconfident consumers fail to anticipate paying, are passed back to consumers through lower account fees. The only harm to overconfident consumers would be on the extensive margin. Some who valued checking accounts below cost would be lured to open an account by underestimating its true cost.

---

17 This prediction is reversed if competition decreases rather than increases the pass-through rate, as is certainly possible (Bulow & Pfleiderer, 1983; Weyl & Fabinger, 2013).
How is the welfare of overconfident consumers affected when half of consumers are rational? One might hope that the presence of rational consumers might provide a positive externality that helps protect overconfident consumers. We know, for instance, that consumers with low search costs can benefit others by lowering equilibrium prices for all. Unfortunately, however, the presence of rational consumers does not protect the overconfident from harm. On the contrary, rational consumers can exert what Armstrong (2014) dubs a negative rip-off externality on the overconfident, raising the prices they pay.

In this example, when equal numbers of consumers are rational and overconfident, annual checking account fees are priced only $50 below cost. Banks still pass 100% of overdraft fee revenue back to consumers in lower fees. Now, however, the account fee reduction is shared equally between overconfident consumers who pay the overdraft fees and rational consumers who do not. As a result, infra-marginal overconfident consumers now pay $50 above cost for their accounts. This overpayment does not accrue to firms who still make zero profit, but to rational consumers who receive the $50 as a cross-subsidy. The only overconfident consumers to benefit from the presence of rational consumers are those on the extensive margin who are dissuaded opening an account when the account fee is $50 below cost rather than $100 below cost.

In short, even when market conditions (a high pass-through rate) prevent firms from exploiting infra-marginal overconfident consumers it does not mean that they are safe. Rational consumers who choose the same contracts and receive cross-subsidies may exploit them instead. In this case, consumer protection policy may be important not to shift surplus from firms to consumers, but to redistribute surplus among consumers to ensure that the costs of the banking system are spread equitably rather than being borne by a few. In practice this issue is important: in 2013 US overdraft fees totaled $32 billion (Andriotis, 2014) but Stango and Zinman (2009, 2014) find that these fees are paid by less than half of account holders. In fact, just 16% of account holders pay over 70% of overdraft fees and this group tends to be lower income than the general population, meaning that the cross-subsidies are regressive (Parrish & Frank, 2011).

Consumer protection policy should be adopted with caution:

Overconfidence may harm consumers in many ways. Indeed firms continually invent new ways to exploit consumer overconfidence because the returns to such exploitative innovation equal or exceed returns to traditional innovation that increases product value (Heidhues, Kőszegi, & Murooka, 2012). I hope that regulatory intervention can help ameliorate the problem. Indeed specific suggestions for regulating consumer credit, by “Prohibiting large penalties for deferring small amounts of repayment” (Heidhues & Kőszegi, 2010), or regulating retail banking, by requiring point-of-sale overdraft warnings (Armstrong & Vickers, 2012; Grubb, 2015), are promising. Moreover, Agarwal et al. (Forthcoming-a) estimate that the 2009 Credit Card Accountability Responsibility and Disclosure (CARD) Act fee reductions have saved U.S.
consumers $12.6 billion per year. In general, however, regulatory attempts to help overconfident consumers require both creativity and caution for at least three reasons:

First, correcting individual decision making errors will in general have smaller benefits when equilibrium considerations are taken into account. In fact, as discussed earlier, overconfidence that causes contract undervaluation can benefit infra-marginal consumers by suppressing prices and thus de-biasing could actually harm consumers once equilibrium price changes are accounted for.

Second, de-biasing some consumers can make things worse for those who remain overconfident. For example, in the preceding example about checking account pricing, suppose that half of overconfident consumers were debiased and became aware of their tendency to overdraft. Bubb and Kaufmann’s (2013) model predicts that they would close their existing accounts and open high-fixed fee accounts with low overdraft fees (perhaps at a credit union). Doing so could save them $50 by avoiding paying cross-subsidies to rational consumers through overdraft fees. Unfortunately, the remaining overconfident consumers would now be outnumbered by rational consumers 2 to 1. Thus 1/3 rather than 1/2 of their overdraft fees would be rebated to them and they would pay 2/3 rather than 1/2 as a subsidy to rational consumers. Their banking costs would rise from $50 above cost to $67 above cost.

Third, de-biasing consumers is likely to be extremely difficult. More practical are pricing restrictions or Thaler and Sunstein (2008) style “nudges” that limit firms’ ability to exploit overconfidence. However, pricing restrictions would often need a detailed understanding of variables such as firm costs and nudges may be ineffective compared to de-biasing. While improving individual decision making, nudges may still fail to help or even harm consumers when firms’ equilibrium responses are accounted for (Grubb, 2015; Grubb & Osborne, 2015; Handel, 2013; Mullainathan, Schwartzstein, & Congdon, 2012; Spiegler, 2014). For instance, Grubb and Osborne (2015) estimate that cellular phone users are overconfident and argue that three-part tariff calling plans are a response to exploit this. In counterfactual analysis, they show that debiasing consumers would have a large welfare increase for consumers and society. However, the more practical “nudge” adopted by the FCC of requiring bill-shock alerts (text messages warning when allowances of minutes, texts, or data are reached) is predicted to lower consumer welfare.

Grubb and Osborne’s (2015) model predicts bill-shock alerts fail to help consumers for several reasons. First when firms lose revenue from one set of fees due to regulation they may raise others, which is sometimes called the waterbed effect (Genakos & Valletti, 2011). Second, Grubb and Osborne (2015) predict that although bill-shock regulation leads to a better set of contract options for consumers, some consumers end up making worse contract choices because they misunderstand the value of the changes. Such unintended consequences of regulation are likely of general concern.
References


Holman, J., & Zaidi, F. (2010). The Economics of Prospective Memory.


