DISCUSSION PAPER SERIES

DP11786

BEHAVIORAL ECONOMICS AND THE ATHEORETICAL STYLE

Ran Spiegler

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

BEHAVIORAL ECONOMICS AND THE ATHEORETICAL STYLE

Ran Spiegler

Discussion Paper DP11786
Published 16 January 2017
Submitted 16 January 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

BEHAVIORAL ECONOMICS AND THE ATHEORETICAL STYLE

Abstract

Behavioral economics is perceived by many to be part of a general shift in the culture of economics toward a less theoretical style. I present a critical discussion of certain manifestations of this trend: a preference for an anecdotal style of exposition (illustrated by Akerlof and Shiller's Phishing for Phools), reduced-form modeling (illustrated by Campbell's Ely Lecture), and the method of capturing psychological forces using parametric modifications of conventional functional forms. I argue that the subject of "psychology and economics" is intrinsically foundational, and that a pure-theory component is essential for it to realize its transformative potential.

JEL Classification: D03

Keywords: N/A

Ran Spiegler - r.spiegler@ucl.ac.uk
Tel-Aviv University & University College London and CEPR

Behavioral Economics and the Atheoretical Style*

Ran Spiegler†

January 11, 2017

Abstract

Behavioral economics is perceived by many to be part of a general shift in the culture of economics toward a less theoretical style. I present a critical discussion of certain manifestations of this trend: a preference for an anecdotal style of exposition (illustrated by Akerlof and Shiller's *Phishing for Phools*), reduced-form modeling (illustrated by Campbell's Ely Lecture), and the method of capturing psychological forces using parametric modifications of conventional functional forms. I argue that the subject of "psychology and economics" is intrinsically foundational, and that a pure-theory component is essential for it to realize its transformative potential.

1 Introduction

In his scientific autobiography "*Misbehaving*", Richard Thaler suggests a link between his "anomalies" project and Thomas Kuhn's theory of scientific revolutions. Looking back to the 1980s, when his list of anomalies started to appear in print, he remarks:

*I am grateful to Kfir Eliaz and Ariel Rubinstein for helpful discussions.

†Tel-Aviv University, University College London, and Centre for Macroeconomics. URL: http://ww.tau.ac.il/~rani. E-mail: rani@post.tau.ac.il.

> "An important aspect of Thomas Kuhn's model of scientific revolutions...is that paradigms change only once experts believe there are a large number of anomalies that are not explained by the current paradigm...As someone who had until recently still been in the "promising" stage of his career, it would be viewed as brash, unseemly, and self-destructive to talk about my own work as something that could be part of a "revolution". My goal was much more modest: just get a few more papers published and begin to establish the case that adding some psychology to economics was an activity worth pursuing. But I had certainly read Kuhn's path-breaking book *The Structure of Scientific Revolutions*, and had secretly spent idle moments wondering whether anything like a paradigm shift could ever be possible in economics." (Thaler (2015), p. 169)

Thus, in the early days of behavioral economics (BE henceforth), it might have made sense to think (or at least daydream) about it as a movement toward a revolutionary paradigm shift, a notion that implies an overhaul of fundamental economic theory.

Times have changed. In a recent piece about teaching BE to undergraduates, Laibson and List (2015) define BE as follows:

> "Behavioral economics uses variants of traditional economic assumptions (often with a psychological motivation) to explain and predict behavior, and to provide policy prescriptions."

No Kuhnian paradigm shift here. Laibson and List's definition is methodologically conservative; it emphasizes the reliance of BE on the existing modeling frameworks of economic theory, and does not count the search for new ones as part of BE's mission. Thaler himself in his AEA Presidential Address (Thaler (2016)) says that "the rise of BE is sometimes characterized as a kind of paradigm–shifting revolution within economics, but I think that is a misreading of the history of economic thought." He goes on to describe

BE as "simply one part of the growing importance of empirical work in economics" and declares that "the best way to proceed is to stop arguing about theoretical principles and just get down to work figuring out the best way of understanding the world."

Indeed, BE is often portrayed (especially in the economics blogosphere) as a key component of the profession's general shift away from a theory-centric culture. We can get a sense of the atheoretical flavor of BE by comparing it to an earlier "transformation of the culture of economics" (to borrow a phrase from Rubinstein (2006)) brought about by Game Theory (GT henceforth). Both BE and GT were *liberating* forces: GT removed the shackles that had tied economists to competitive markets, and BE freed them from prior fixations on narrow self-interest and error-free decision makers. The difference is that unlike BE, GT not only liberated economists, but also *demanded* of them to learn a new language. Concepts like rationalizability, implementation and robustness to high-order beliefs were not "variants on traditional assumptions", but a web of new concepts, modeling tools and techniques. BE demands very little in this regard, as many of its modeling ideas are reinterpretations or formerly unutilized specifications of standard frameworks.

No sharp line separates BE from the theoretical style. Indeed, some of the theoretical arguments I will invoke in this paper were originally made by card-carrying behavioral economists. Rather, we should think of the phenomenon as a *tendency* away from abstract theorizing. This tendency takes various forms, depending on which segment of the economics spectrum we are looking at. Papers that present models with "behavioral" elements tend to follow the style of "applied" rather than "pure" theory. Psychological biases are introduced as parameters that alter familiar functional forms, rather than fundamental modifications of the underlying modeling frameworks. Policy-oriented analysis tends to employ "reduced-form" rather than "structural" modeling. And expositions that address a broader audience tend to adopt a purely anecdotal style - whereas similarly popular presentations of subjects like GT usually contain at least a modicum of theoretical abstractions, however informal.

Of course, the literature includes pure-theory approaches to "psychology and economics". Ariel Rubinstein's "procedural rationality" approach (Rubinstein (1998)) formulates decision processes in an unmistakable "pure theory" style. And then there is the subfield of "behavioral decision theory", which took off following the work of Gul and Pesendorfer (2001) on self-control preferences. This literature extends the tradition of axiomatic decision theory to new domains of choice objects, in an attempt to incorporate new behavioral elements. But are these developments *part* of BE? I don't think that Gul, Pesendorfer or Rubinstein are viewed by anyone (including themselves) as "behavioral economists". Indeed, they have written outspoken *critiques* of BE's style (see Rubinstein (2006), Gul and Pesendorfer (2008)). Their approaches seldom feature in BE conference programs or course syllabi. Given BE's strong influence on the terms in which the "psychologizing" of mainstream economics is taking place, it is important to examine its atheoretical tendency - even if there are more abstract approaches out there.

However, the key question is not whether the atheoretical tendency in BE is real, but whether it imposes any costs. I claim in this paper that it does. Given that BE deals with the very building blocks of economic behavior, it has an intrinsic "foundational" character. If BE is not going to have a high dose of "pure" or fundamental theorizing, what branch of economics would? When the subject is developed without a strong theory component, the result is a flatter discourse that robs the subject of the conceptual depth and richness it deserves, and sometimes neglects key forces that affect the analysis of economic systems with "behavioral" agents.

Full critical examination of the coevolution of BE and the general atheoretical trend in economics is a fascinating topic for historians and sociologists of economic thought; it would require a full-length book and lies beyond the scope of a paper like this. The best I can do is illustrate my thesis with a few examples, preferably ones that are current and familiar. In Section 2, I use George Akerlof and Robert Shiller's 2015 book "*Phishing for Phools*" to illustrate the limitations of the anecdotal, theory-free style that is common in popular expositions of BE. In Section 3, I use John Campbell's 2016

Ely Lecture to demonstrate the limitations of reduced-form modeling of decision errors. In Section 4, I turn to the "applied theory" style that often characterizes BE papers, with a focus on the practice of incorporating behavioral elements into economic models via parametric modifications of familiar functional forms. Here I employ a "semi-fictional" example, which conveys the point by way of an analogy to the "conjectural variations" approach to oligopoly (compared with the now-dominant GT approach). In Section 5, I come back full circle and return to Thaler's opening quote. I believe that BE does have the paradigm-shifting potential that Thaler secretly dreamed of in the 1980s. For BE to fully realize this potential, it has to put a higher premium on abstract theorizing in general, and on the creation of new modeling frameworks in particular.

2 The Anecdotal Style

The most extreme manifestation of the atheoretical tendencies in BE is expositions that shed theoretical reasoning altogether, in favor of a loose collection of anecdotes about the economic consequences of decision biases and non-standard motivations. Naturally, this style is most likely to be seen in pieces that address a broad audience. In this section I examine a recent example of this genre: Robert Akerlof and Robert Shiller's "*Phishing for Phools*" (Akerlof and Shiller (2015)). Their book explores the implications of consumer fallibility for the way we ought to think about the "free market". Its main thesis is that consumers' departure from rationality (their "phoolishness", to use the authors' neologism) makes the proliferation of exploitative transactions a necessary feature of the market system.

Akerlof and Shiller make their case with a collection of anecdotes about market exploitation of fallible consumers; their exposition is almost entirely devoid of theoretical reasoning. As one might expect from these authors, the anecdotes are illuminating and woven into an absorbing story. Nevertheless, in this section I argue that the anecdotal style has its limitations, and that incorporating *some* theorizing would have been valuable.

In the context of a popular book, I perceive the term "theorizing" in

very broad terms. In particular, I do *not* identify theorizing with formal modeling, and include verbal abstractions that do not have a formal model in the background as "legitimate" examples of theorizing. Even those are very rare in *Phishing for Phools*. However, the specific theoretical ideas that I will invoke in this section (and are missing from Akerlof and Shiller's book) are all borrowed from the existing theoretical literature on markets with less than fully rational consumers. And given that the cultural norm in academic economics is to introduce theoretical ideas by means of a formal model, these theoretical ideas were originally presented as formal models of varying degrees of sophistication. Incorporating these ideas into *Phishing for Phools* would have meant popularizing these models.

And here I must get a natural objection off the table, and that is the argument that a popular book has no room for theoretical arguments that are derived from formal models. I strongly disagree. The fact that many popular books on BE were written by psychologists and marketing researchers accounts for their "collection of biases" style. But it does not follow that the anecdotal style must carry over to discussion of the biases' *economic* implications. For example, no popular exposition of GT is complete without some kind of description of Nash equilibrium, backward induction or signaling arguments. Of course, the expositions are verbal and entertaining, but they go beyond mere anecdotes. In an age when authors like Brian Cox and Simon Singh are writing best-selling books that contain a sketch of the proof of $E = mc^2$ or an explanation of RSA encryption, readers of popular economics can survive a bit of non-technical theorizing.

Linking isolated anecdotes

One of the earliest stories in *Phishing for Phools* involves the famous empirical finding of DellaVigna and Malmendier (2006) that health-club customers appear to overestimate their future consumption when choosing a price plan. Many of those who go for monthly subscriptions (with automatic renewal) end up paying more than if they had opted for a by–the-visit plan - they "pay not to go to the gym", as DellaVigna and Malmendier put it in the title of their paper.

Remarkably, except for two sentences at the *end* of the book, Akerlof and

Shiller remain silent about a simple theoretical argument that DellaVigna and Malmendier themselves make in a companion paper (DellaVigna and Malmendier (2004)). In their model, firms in a competitive setting offer two-part tariffs to consumers with a taste for immediate gratification. In the health-club context, this means that ex-ante, consumers would like to commit to do plenty of physical exercise in the future, but as time goes by their preferences change and they become lazier. Whether or not consumers can predict this future change in their preferences, the two-part tariffs that emerge in Nash equilibrium consist of a large lump-sum payment and a per-unit price *below* marginal cost. By comparison, if consumers had dynamically consistent preferences, firms would adopt marginal-cost pricing in Nash equilibrium.

Why is the omission of this theoretical result remarkable? Because in a later chapter, Akerlof and Shiller present yet another example of market exploitation: the pricing of credit cards (see pp. 68-69). Here, common price plans are the mirror image of the health-club case; they involve no lump sum and a *high* marginal interest rate. DellaVigna and Malmendier's model offers a simple explanation. Credit cards enable the consumer to enjoy an immediate consumption benefit and defer its cost. In contrast, attending a health club is an investment that pays off in the future. According to the DellaVigna-Malmendier model, this inversion in the temporal distribution of costs and benefits explains the direction of the equilibrium departure from marginal-cost pricing.

The logic behind this result depends on whether the consumer predicts the future change in his preferences. When he does, he seeks a commitment device to counter his taste for immediate gratification. A high marginal interest rate acts as a commitment device that deters excessive use of the credit card, whereas a low per-visit price acts as a commitment device that incentivizes health-club attendance. When the consumer underestimates his future taste for immediate gratification, the equilibrium two-part tariff is an effective *bet* on the consumer's future consumption. The firm and the consumer have different prior beliefs regarding the consumer's future preferences, and therefore they have a motive to engage in speculative trade, shifting net consumer utility from the state predicted by the firm to the state predicted

by the consumer.

The DellaVigna-Malmendier model thus links two examples of exploitative pricing that otherwise appear to be unconnected. The model not only links them, but also explains the difference in their departures from marginal cost pricing. Luckily for authors of a popular book, this involves an undergraduate-level argument that can easily be conveyed to a broad audience. At the same time, it is pregnant with follow-up questions that feed "higher-level" theorizing: How would the firms set prices if they did not know the consumer's ability to predict future changes in his preferences? What kind of price plans would they offer if not confined to two-part tariffs - in particular, can we explain real-life examples of complex non-linear pricing? What is the effect of market competition on consumer welfare?[1]

The point is that some of the market exploitation anecdotes presented by Akerlof and Shiller cry out for a connecting thread (one that I have not mentioned, for the sake of brevity, is the add-on pricing example of Gabaix and Laibson (2006)). Such a connection requires some theorizing, however elementary. In the absence of theorizing, all we have is a loose collection of anecdotes. By refusing to theorize, Akerlof and Shiller water down their message.

Qualifying the main message

Another important role of theoretical reasoning - especially in the formal-modeling tradition - is to qualify sweeping verbal statements. Because the main thesis of "*Phishing for Phools*" is presented without any trace of formal modeling, it leaves the impression that "phoolishness" always harms consumers. But what if it could actually *mitigate* market failures that originate from other sources?

Ironically, Akerlof's celebrated "market for lemons" model provides a good illustration of this idea, since market failure in the lemons model is a consequence of uninformed buyers' *sophisticated understanding* of adverse selection. As Akerlof and Shiller point out, "phoolish" buyers have a lim-

[1] For a few papers that address these questions, see Eliaz and Spiegler (2006), Grubb (2009), and Kőszegi and Heidhues (2010). For more general treatments of this class of models, see Spiegler (2011), Kőszegi (2014) and Grubb (2015).

ited understanding of the incentives behind sellers' behavior, and as a result they may form a biased estimate of the quality of the products that are traded in the market (see, for example, their discussion of mortgage-backed securities in Chapter 2). A number of authors (Eyster and Rabin (2005), Jehiel and Koessler (2008), Esponda (2008)) have proposed ways to model "markets for lemons" with such buyers. These models paint a rich picture: "phoolishness" can mitigate or exacerbate the market failure due to adverse selection, depending on the nature of consumers' limited understanding and the gains from trade. Although I believe that the argument can be (at least partly) conveyed verbally to a lay audience, in the present context it would be worthwhile to do it formally.

The above-cited papers all build on a familiar reformulation of the lemons model, following Bazerman and Samuelson (1985), where a situation in which many sellers compete for a buyer is approximated by a bilateral-trade game in which the buyer has all the bargaining power. Formally, an uninformed buyer makes a take-it-or-leave-it offer p to a seller who privately learns the value v of the object he owns, where $v \sim U[0,1]$. The buyer's valuation is $v + b$, where the constant $b \in (0,1)$ represents the gain from trade. When the buyer has rational expectations, he knows that the seller will trade if and only if $p > v$. Therefore, the buyer chooses p to maximize

$$\Pr(v < p) \cdot [E(v \mid v < p) + b - p] = p \cdot [\frac{1}{2}p + b - p].$$

The solution is $p^* = b$. Thus, although trade is efficient for all v, in equilibrium it will take only place with probability b.

Eyster and Rabin (2005) used the notion of "cursedness" to model a possible departure from rational expectations. They assumed that in equilibrium, the buyer knows the marginal distributions over v and the seller's action, but does not perceive any correlation between them. Thus, the buyer has a *coarse* perception of the seller's behavior, since he fails to account for its responsiveness to v. As a result, the buyer chooses p to maximize

$$\Pr(v < p) \cdot [E(v) + b - p] = p \cdot [\frac{1}{2} + b - p].$$

Thus, the buyer's expectations completely disregard the adverse selection consideration; his forecast of the object's value conditional on trade is given by the ex-ante distribution. The solution is $p^{ER} = \frac{1}{2}b + \frac{1}{4}$. We can see that $p^{ER} < p^*$ if and only if $b > \frac{1}{2}$ - i.e., "cursedness" exacerbates the market failure due to adverse selection only if the gain from trade is large. The intuition behind this ambiguous effect is that "cursedness" has two contradictory effects. On one hand, the buyer's expected valuation is higher than in the benchmark case because he ignores adverse selection; this raises the buyer's bid relative to the benchmark. On the other hand, the buyer does not realize that a higher bid would enhance the expected quality of the traded object; this lowers the buyer's bid relative to the benchmark. When the gains from trade are small, the former consideration outweighs the latter.

This ambiguity also implies that comparative statics with respect to the buyer's degree of "phoolishness" are not monotone. Jehiel and Koessler (2008) examined an example in which the buyer has a partially coarse perception of the seller's behavior: he partitions the set of possible realizations of v into intervals (of potentially unequal size), and he believes that the seller's strategy is measurable with respect to this partition. Using the notion of "Analogy-Based Expectations Equilibrium" (Jehiel (2005)), Jehiel and Koessler show that the equilibrium probability of trade is not monotone with respect to the fineness of this partition. In other words, greater "phoolishness" does not imply a stronger market failure.

Esponda (2008) assumed that the buyer's expectation of v conditional on trade is based on naive extrapolation from the equilibrium distribution itself. In his model, the buyer learns the traded object's value from observations of past transactions - without realizing that this sample is adversely selective, such that if the price that characterized historical observations changed, so would the observed quality distribution. The equilibrium price p^E is defined as follows:

$$\begin{aligned} p^E &\in \arg\max_p \quad \Pr(v < p) \cdot [E(v \mid v < p^E) + b - p] \\ &= \arg\max_p \quad p \cdot [\frac{1}{2}p^E + b - p] \end{aligned}$$

such that $p^E = \frac{2}{3}b$. In this case the buyer's "phoolishness" unambiguously exacerbates the market failure due to adverse selection. The reason is that of the two forces identified in our discussion of "cursed" buyers, Esponda's model shares only the force that pushes the price down.

The three models described above present different ways in which the buyer's understanding deviates from the rational-expectations ideal, and they force us to ask: "When we say that buyers don't understand the seller's incentives, what is it exactly that they don't understand?" Alternatively, they suggest that the bilateral-game reformulation of the lemons market model, which is successful in the rational-buyer case, might miss a key aspect of competition among rational sellers for a "phoolish" buyer. These question marks are a valuable corrective to a sweeping message like "phoolishness leads to bad market outcomes".

What is phishing equilibrium?

Toward the end of their book, Akerlof and Shiller give an argument that may be viewed as an explanation for their atheoretical approach:

> "This general way of thinking, with its insistence of general equilibrium, has been the central nervous system for economic thinking for almost two and a half centuries. Yet behavioral economics...seems oddly divorced from it. Our two examples from behavioral economics, of DellaVigna-Malmendier and Gabaix-Laibson, illustrate. In the style required now for a journal article, their modeling and examples are very special...In accord with the standards of economics journal articles, these articles prove that phishing for phools exists. They do so by giving models and examples, where that phishing is undeniable; but the journal's demand for such undeniability comes at a cost. It means that the generality for phishing for phools cannot be conveyed." (Akerlof and Shiller (2015), pp. 169-170)

As this passage demonstrates, when Akerlof and Shiller abandon the anecdotal style, it is to advocate a "think big", general-equilibrium approach to

the subject of markets with "phoolish" consumers - compared with the piecemeal approach of analyzing small models that characterizes most of academic economic theory. (As an aside, I would have thought that Akerlof's lemons model proved once and for all the power of small models to convey big ideas.) They introduce the notion of "phishing equilibrium" and define it essentially as follows: *Every opportunity to exploit consumers is realized.*

Yet the meaning of this equilibrium concept is vague. An important feature of general equilibrium as we know it is linear-price taking. But as we saw in our discussion of DellaVigna and Malmendier (2004), endogenously complex price schemes are a hallmark of markets with non-rational consumers. Therefore, linear-price taking seem inappropriate. Another feature of general equilibrium is the no-arbitrage principle. Akerlof and Shiller rightly observe that firms seek every opportunity to exploit "phools". However, the no-arbitrage condition means that such activities should occur *off equilibrium*; in equilibrium, the profits from these opportunities have been competed away. Yet, game-theoretic models of competition for boundedly rational consumers often have the property that tougher competition does not dissipate profits because it strengthens firms' incentive to obfuscate and target erring consumers (Spiegler (2006), Chioveanu and Zhou (2013), Gabaix et al. (2016)). A "general equilibrium" model based on the assumption that competitive forces drive the gain from the "marginal phish" down to zero would exclude many interesting and potentially relevant situations.

Thus, while the call for a "general equilibrium" approach to the subject of market exploitation of "phools" is genuinely intriguing, it warrants a serious "pure theory" approach. In the absence of any attempt at formal modeling, it is hard to understand what "phishing equilibrium" could possibly mean or imply.

Summary

I have shown that key aspects of the "phishing for phools" argument could benefit from a more developed theoretical approach, even allowing for the broad-audience factor. A more theoretical style would insightfully link the anecdotes; it would qualify sweeping claims regarding the market implications of "phoolishness"; and it would impose more discipline on conceptualizations

like "phishing equilibrium". Of course, economists of Akerlof and Shiller's stature don't need a sermon about the virtues of economic theory; as the above-quoted passage indicates, they made a deliberate choice to adopt an anecdotal and atheoretical style. Their choice reflects a wider sentiment that this style is appropriate to the subject matter. Yet, as I have demonstrated, this has flattened the message of their book.

It may also have diminished the book's long-run impact. A broad audience is also a variegated one: readers of a book like *Phishing for Phools* include bright undergraduate students from various disciplines. We want such students to join our ranks and move the discipline forward. Akerlof and Shiller's absorbing anecdotal style will surely attract their attention, but a bit of abstract theorizing may also spark their imagination, by exposing them to the subject's potential depth and richness, and its open conceptual problems.

3 The Reduced-Form Style

Another aspect of the atheoretical style in BE is the tendency toward reduced-form modeling of decision biases. Rather than writing down an elaborate choice model that explicitly captures a psychological mechanism, economists often take a completely standard model in which choice follows a straightforward cost-benefit analysis, and then reinterpret or relabel some of its objects as biases or errors (e.g., Bar-Gill (2012), Mullainathan et al. (2012)). In this manner, the modeler seems to have it both ways: on one hand, he can address "behavioral" phenomena and study their implications, yet on the other hand, he can "conduct business as usual" in terms of the modeling procedure.

A recent example of this practice is John Campbell's Ely Lecture (Campbell (2016)), which was devoted to boundedly rational decision making in the context of financial products and its possible implications for market regulation. In the lecture's theoretical part, Campbell focuses on a particular regulatory intervention: imposing a tax on complex and potentially exploitative products. To evaluate this intervention, he constructs a simple model

with two products: one "simple", and the other "complex". The simple product has a fixed value, normalized to 0, which is correctly perceived by all consumers. In contrast, the complex product may be characterized by heterogeneity in consumers' valuation (denoted u). A proportion α of consumers are sophisticated and a fraction $1-\alpha$ are naive. Sophisticates' valuation of the complex product is unbiased. In contrast, when a naif values the complex product at u, its *true value* to the consumer is $u-1$. Thus, the valuation error committed by naive consumers is fixed at 1. From a descriptive point of view, this is a standard utility-maximization model and the "behavioral" element is restricted to the welfare analysis.

Campbell examines the consequences of imposing a fixed tax $b < 1$ on the complex product under various scenarios for the redistribution of tax revenues. For simplicity, I consider the case in which the revenues are *not* rebated. Consumers with $u \geq b$ ($u \leq 0$) choose the complex (simple) product both before and after the intervention. The only consumers whose behavior is affected by the intervention are those with $u \in (0, b)$. Turning to welfare analysis, all consumers with $u > b$ are harmed by the tax, whereas all consumers with $u < 0$ are unaffected by it. In the case of consumers with $u \in (0, b)$, we need to distinguish between sophisticates and naifs. The former are made unambiguously worse off since they switch to the simple product and earn a net payoff of 0, as compared to $u > 0$ prior to the intervention. In contrast, naive consumers with $u \in (0, b)$ are made better off since their true utility prior to the intervention is $u - 1 < u - b < 0$, as compared to 0 afterward. If there are a sufficient number of consumers in the latter group, then the tax improves overall consumer welfare.

Campbell's reduced-form model does not define what makes products simple or complex, and thus does not specify the origin of the naive consumers' errors. But there is something else missing: the products are offered in a *market* by interested parties who may act like "phishers", to use Akerlof and Shiller's terminology. The two features are interrelated: the type of product that firms offer is likely to be responsive to the procedures consumers use to cope with product complexity. We will see that taking both features into account can dramatically change the analysis. (Campbell acknowledges

that his simple model neglects various considerations, including interaction between "behavioral" effects and other market failures or firms' political lobbying. However, one would have thought that the market context in which products are offered would be considered to be a key feature that should be at the fore.)

I describe a simple model in the spirit of Spiegler (2006), which mimics Campbell's reduced-form model as closely as possible while incorporating the two missing features. In the model, a single firm offers the complex product, which I interpret to be a state-contingent service contract. The state of nature is uniformly distributed over $[0, 1]$. The service is offered with two possible quality levels, 0 or 1. When the firm offers quality $q \in \{0, 1\}$ in some state, it incurs a cost of cq and the consumer earns a payoff of $q - b$, where $c \in (0, 1)$ is the cost of offering a high level of quality, and $b \in [0, 1 - c)$. The firm's strategy has two components: a price T, and a function $f : [0, 1] \to \{0, 1\}$ which determines the quality of service in every state. Let p denote the frequency of the states in which it offers a high level of quality - that is,

$$p = \int sf(s)ds$$

The complexity of the firm's product has a concrete meaning in this model: the product is a state-contingent contract with a rich state space.

A fraction $1 - \alpha$ of consumers are naive and find it difficult to evaluate the contract. Every naive consumer follows a simplifying heuristic: he draws a state s at random, learns the value of $f(s)$, and regards it as a prediction of the level of quality he will receive if he chooses the firm's product. There is no correlation between the state the consumer draws and the state that will actually be realized. The interpretation is that the consumer, unable to fully digest the contract with its many clauses, examines a random clause and treats it as being "representative". His error lies in the fact that he exaggerates the informativeness of a very small sample - a stylized version of the phenomenon that Tversky and Kahneman (1971) called "the law of small numbers".

The remaining fraction α of the consumer population are sophisticated,

in the sense that their belief regarding the level of quality they will receive is correct given their information. In order to mimic Campbell's assumption that the distribution over subjective valuations is the same for both the naive and sophisticated consumers, I assume that the latter are *perfectly* informed of the state of nature, and therefore know the level of quality they will receive if they choose the complex product. Thus, they also have an *informational* advantage over the naive consumers. Note that by paying attention to the procedural origins of the naive consumers' error, we get a better understanding of what lies behind Campbell's stark assumption. Finally, the terms of the simple product are exogenous; i.e. quality 0 is offered in *all* states free of charge, and therefore, both of consumer types value it at zero. The simplicity of the simple product stems from the lack of quality variation across states.

A consumer's gross valuation of the complex product takes two possible values, 0 or 1. It follows that the firm will necessarily choose the price $T = 1-b$, such that a consumer's net subjective valuation of the complex product is either 0 (in which case he breaks the tie in favor of the complex product) or -1 (in which case he chooses the simple product). As in Campbell's model, the sophisticated consumer is always right. Unlike Campbell's model, the naive consumer's valuation is unbiased on average since it is generated by an unbiased signal. However, because the consumer will only choose the complex product when he has a high assessment of its quality, his valuation of the complex product is biased upward *conditional on choosing it.* The size of the bias is $1-p$, since the product's true expected quality is p whereas the conditional perceived quality is only 1.

The firm's problem is reduced to choosing $p \in [0,1]$ in order to maximize

$$\alpha p(1-b-c) + (1-\alpha)p(1-b-pc)$$

The first (second) term represents the firm's profit from a sophisticated (naive) consumer. Every consumer chooses the firm with probability p. The firm's net profit conditional on being chosen by a sophisticated consumer is $1-b-c$ since he chooses the complex product knowing that it will provide

a high level of quality. (Our assumption that $b < 1 - c$ implies that the firm does not incur a loss on sophisticated consumers.) The firm's net expected profit conditional on being chosen by a naive consumer is $1 - b - pc$ since the actual level of quality he will obtain is independent of the level of quality in the state he sampled.

As long as α is not too large, the solution p^* to the firm's maximization problem is interior:

$$p^* = \frac{1 - b - \alpha c}{2c(1 - \alpha)}$$

By the assumption that $b < 1 - c$, $p^* > \frac{1}{2}$. This property will be instrumental in the welfare analysis presented below. Note that p^* decreases with b, i.e., *the firm responds to the tax with a lower frequency of offering a high level of quality.* Intuitively, transactions with naive consumers have an exploitative "bait and switch" flavor: with probability $p(1 - p)$, the firm attracts a consumer who sampled a high level of quality and ends up providing him with a low level, thus saving the cost. As b rises, the firm's profit margin shrinks, and its incentive to adopt the cost-saving bait-and-switch tactic becomes stronger.

Now turn to a calculation of consumer welfare as a function of b. Sophisticated consumers earn a true payoff of zero both before and after the intervention. Therefore, consumer welfare is driven by the naifs. A fraction p^* of them choose the complex product and earn a true expected payoff of $p^* - (1 - b) - b = p^* - 1$, whereas their subjective payoff is 0. Thus, the valuation error of naive consumers who choose the complex product is $1 - p^*$. Unlike in Campbell's reduced-form model, the magnitude of a naive consumer's error *increases* with b due to the firm's *endogenous* response to the tax. When b increases, fewer naive consumers end up being exploited, but those who are get exploited to a greater degree. The latter effect is a regulatory cost that is missing from Campbell's model. Total consumer welfare is $-(1 - \alpha)p^*(1 - p^*)$, and because p^* is greater than $\frac{1}{2}$ and decreasing in b, consumer welfare unambiguously *decreases* with b. That is, the intervention's adverse effect due to greater exploitation of naive consumers who demand the complex product outweighs the positive effect of reducing their

numbers. This equilibrium effect thus turns out to be crucial but is missed by the reduced-form model.

The economic lesson from this exercise is that using taxes or subsidies to make a complex product objectively less attractive may impel firms to magnify its role as a vehicle for exploiting naive consumers. Although the example was "cooked" to mimic as many features of Campbell's model as possible, the aspect it highlights would appear in certain competitive variations of the market model (which would be technically more intricate), as well as under different conceptualizations of product complexity and consumers' response to it.

I do not wish to push the regulatory lesson too far - that is not my objective here. Rather, the point is that when we analyze the effect of regulating "complex" products, it helps to have *some* model of what this complexity consists of and how consumers deal with it, since this may provide a clue as to the endogenous market response to the regulatory intervention. In Spiegler (2015), I apply this methodology to regulatory interventions known as "nudges" (default architecture, disclosure). To my mind, the promise of "psychology and economics" lies precisely in the ability to *enrich* economic analysis in such directions, rather than in giving us permission to use the same old models while relabeling some of their components as errors.

4 The Applied Style

Perhaps the subtlest manifestation of BE's turn away from theoretical abstraction involves what I call "the applied style": economic models that incorporate behavioral elements in a way that retains the appearance of "applied theory" (while going beyond the reduced-form approach described in the previous section, with its "behavioral" relabeling of terms in an otherwise utterly conventional cost-benefit analysis).

It is difficult, and probably futile, to draw a precise line between "pure" and "applied" theory. Nevertheless, the following two distinctions are relevant to the present discussion: First, in the pure style a new model is presented using a simple, bare-bones example or a general abstract formulation,

whereas in the applied style an elaborate specification of the model is developed in an attempt to capture salient features of the environments to which the model is meant to be applied. As a by-product, an applied-style piece tends to put a small weight on choice-theoretic or revealed-preference considerations. Second, in the applied style a model often introduces novel behavioral elements in the form of parametric modifications of familiar functional forms, whereas in the pure style a model targets the conceptual framework to which they belong, by modifying some of its fundamental assumptions or introducing new primitives.

What is the desirable mix between the two styles? In view of the intrinsically "foundational" aspect of BE, a minimal dose of the pure style in the early stages of development of a non-trivial behavioral model is essential in my opinion. In this context, the atheoretical shift that is the subject of this paper can be described as an insufficient dose. I will try to substantiate this claim using two examples.

The case of optimal expectations

The question of the appropriate mix between the pure and applied styles in BE modeling was on my mind in Spiegler (2008), where I examined the model of "optimal expectations" due to Brunnermeier and Parker (2005) (BP henceforth). This example is particularly appropriate since it illustrates the first of the two distinctions made above. The BP model is based on the idea that decision makers deliberately distort their beliefs in order to enjoy "anticipatory utility" (in addition to standard material utility). The distortion is not arbitrary and is subjected to a "no cognitive dissonance" constraint, according to which the decision maker's action maximizes his expected material utility given his chosen belief.

BP define their model in the context of an intertemporal consumption problem since they are investigating macro and finance applications. It is worthwhile presenting the decision maker's objective function in their model (with a slight change in notation) in order to get an impression of its intim-

idating complexity:

$$E_\pi \left\{ \frac{1}{T} \sum_{t=1}^{T} \left[\beta^{t-1} \left(\sum_{\tau=1}^{t-1} \beta^{-\tau} u(c_{t-\tau}) + u(c_t) + E_{\hat{\pi}} \left(\sum_{\tau=1}^{T-\tau} \beta^{\tau} u(c_{t+\tau}) \mid s_1, ..., s_t \right) \right) \right] \right\}$$

where c_t is consumption in period t; u is the material utility from periodic consumption; s_t is the realization of an exogenous state variable in period t; π is the objective distribution over $(s_1, ..., s_T)$; and $\hat{\pi}$ is the decision maker's chosen belief over $(s_1, ..., s_T)$. In one of BP's applications, an investor chooses between two financial assets, one safe and the other risky and they characterize the investor's behavior in this class of binary choice problems.

I have always found the idea of the BP model very interesting. At the same time, I felt that a model in which decision makers *choose* what to believe is a major departure from the basic principles of rationality, and thus should be treated with extra caution. The above expression's complexity makes it hard to gauge the model's departure from rational choice. In Spiegler (2008), I tried to get a better understanding of the BP model using a much simplified *static* version, where the decision maker chooses an action $a \in A$ and a belief $\hat{\pi} \in \Delta(S)$ (where S is a finite set of states of nature) in order to maximize the objective function $\alpha E_\pi(a) + (1-\alpha) E_{\hat{\pi}}(a)$ subject to the constraint that $a \in \arg\max_{a'} E_{\hat{\pi}}(a)$, where $\alpha \in (0,1)$ is constant and $\pi \in \Delta(S)$ is assumed to have full support.

I posed the following question: Can the observed choice correspondence induced by this simplified BP model be rationalized? The answer turns out to be *no*. When we take a choice set like the one that BP examined, we can generate examples that exhibit the following pattern: the decision maker selects the risky option (and optimistically distorted beliefs) in the binary-choice case, but when a third, ultra-risky alternative is added to the choice set, he will revert to the safe option (and realistic expectations), thus violating the Independent-of-Irrelevant-Alternatives (IIA) axiom. Intuitively, the decision maker must choose a very optimistic belief if he wants to enjoy an anticipatory utility from the moderately risky action. However, in the expanded choice set, the no-cognitive-dissonance constraint requires him to

react to this belief by choosing the ultra-risky action, which generates lower overall utility due to its extreme downside.

This finding has several implications. In terms of economic substance, it shows that the BP predictions regarding the shift in investors' choices due to optimal expectations are not robust, since they can be overturned if we expand the choice set. At the psychological and choice-theoretic level, the violation of the IIA axiom is not arbitrary, but appears to capture an interesting and possibly general insight: people with access to more options (specifically, riskier ones) are less likely to delude themselves.

On the methodological level, this exercise demonstrates the benefit of the pure style. By taking the simplest possible version of the BP model and thinking about its basic choice-theoretic aspects, we obtained an interesting finding that is crucial to the interpretation of BP's results. In general, principles like IIA in choice theory, the Revelation Principle in mechanism design or the Single Deviation Property in dynamic games are extremely useful in simplifying the analysis of economic models. However, they are sometimes violated by BE models. Checking whether they continue to hold is a basic exercise that cannot be "outsourced" to pure-theory specialists - it is essential to a modeler's basic understanding of his own model.

Parametric modification of functional forms

I now turn to the second distinctive feature of the applied style: the tendency to capture behavioral elements by distorting familiar formulas with additional parameters. Rabin (2013) presents an eloquent guide to the "parametric modification" approach. The virtues of this approach that he emphasizes - e.g., enabling empirical tests of a null hypothesis that the behavioral effect does not exist, and quantifying the departure from this hypothesis when it is rejected - are "applied" in nature. The following example illustrates the limitations of the parametric approach. For the sake of variety, I use a different line of argumentation than in the rest of the paper. Rather than tackling a particular BE paper, I present a partly fictional example that relates to BE only by way of analogy.

Imagine that we live in a world in which GT had not been invented; moreover, the only familiar models of market structure are standard monopoly and

perfect competition. Now comes along Professor X and proposes a modeling approach to oligopolistic behavior. He considers a market for a homogenous product with n firms and constant marginal cost c. The inverse demand function is $P(Q)$, where Q is the aggregate supplied quantity. Each firm chooses its production quantity q in order to maximize the following expression:

$$q \cdot [\alpha \cdot P(nq) + (1 - \alpha) \cdot p^* - c]$$

where $\alpha \in [0, 1]$ is an exogenous parameter that is allowed to vary with n, and p^* is the market equilibrium price. In equilibrium, the firms' optimal quantity q^* satisfies

$$P(nq^*) = p^*.$$

This model employs the parametric approach to capture equilibrium behavior in an oligopoly. When $\alpha = 1$, the firm plays as if it is part of a cartel that maximizes industry profits and allocates production evenly among its members. When $\alpha = 0$, the firm acts as a price taker, and the model collapses to competitive equilibrium. An interior value of α captures the intermediate case of oligopoly. Moreover, we can capture the intuition that a market with more competitors is more competitive, by assuming that α is some decreasing function of n. When we assume linear demand $P(Q) = 1 - Q$ and $c \in (0, 1)$, the equilibrium price is

$$p^* = \frac{\alpha + c}{\alpha + 1}.$$

This result is intuitively appealing: a higher value of α (which corresponds to a greater departure from perfect competition) results in a higher equilibrium price. Moreover, Professor X can make assumptions about the speed with which α decreases with n in order to derive quantitative predictions of equilibrium mark-ups and industry profits.

I said earlier that this example is partly fictional. In fact, it is very close in spirit to the actual model of "*conjectural variations*", which was a popular approach to oligopoly before the advent of GT (see Tirole (1988), p. 244). That model, too, had a free parameter, which captured the firm's belief regarding the reaction of its opponent to changes in its own behavior.

With the benefit of hindsight, it is clear that the game-theoretic approach to oligopoly has given us a *language* for studying many aspects of oligopolistic behavior - tacit collusion, the value of commitment, entry deterrence, etc - that seem to lie beyond the scope of the parameteric approach (whether it takes the form of conjectural variations or the present example). The latter could continue to offer useful "reduced form" models for applied work, but its status as a Theory of Principle has clearly been diminished by the rise of the game-theoretic approach.

By way of analogy, I believe that a similar diagnosis applies to BE: behavioral phenomena possess intrinsic depth that calls for a pure-style modeling approach that will enrich our analytical vocabulary in ways that lie beyond the reach of the applied-style parametric approach. In fact, the example that Rabin (2013) adduces as the biggest success of the parametric approach - the (β, δ) representation of intertemporal preferences with present bias - strengthens my point. As Rabin himself points out, the (β, δ) model is a specification of the multi-selves approach to analyzing behavior under dynamically inconsistent preferences. This approach was well in place before the surge in the popularity of the (β, δ) model. And the fact that researchers were able to place the (β, δ) parametrization firmly within the multi-selves framework facilitated coherent analysis of its implications as well as the realization that a key issue is the solution concept one employs to analyze the resolution of the conflict among selves. I believe that the same holds for other successful examples of the parametric approach: their power grows with our ability to relate them to more abstract modeling frameworks, whether existing ones or ones that remain to be developed.

5 Conclusion

I hope that this journey has not left the impression that I am some kind of a "theory fanatic". In fact, I am as uncomfortable as the next person with superfluous formalism. The range of desirable styles of theorizing varies across subjects, and not every subject requires sophisticated theorizing of the "pure", foundational variety. However, "psychology and economics" is surely

one that does, since it deals with the very building blocks of economic models. The fact that BE has sung its music with a low-volume theory register is one of the reasons for its popularity; and it is undoubtedly a sound approach in many contexts. Nonetheless, the approach has its limitations, as I hope to have demonstrated in this paper.

Many of the examples I have looked at involve the implications of BE for market interactions and their regulation. This reflects the centrality and topicality of this particular question, as well as my own prior preoccupation with it. But it also highlights a more general point: the need for a "high-volume theory register" is especially acute in the analysis of *interactions* with or between "behavioral" agents.

Richard Thaler's call to "stop arguing about theoretical principles and just get down to work", quoted in the Introduction, was made as part of a passage that endorsed another Ely Lecture devoted to BE, this one given by Raj Chetty, who advocates a "pragmatic" approach to BE. The following paragraph from Chetty's paper summarizes his approach well:

> "The decision about whether to incorporate behavioral features into a model should be treated like other standard modeling decisions, such as assumptions about time-separable utility or price-taking behavior by firms. In some applications, a simpler model may yield sufficiently accurate predictions; in others, it may be useful to incorporate behavioral factors, just like it may be useful to allow for time non-separable utility functions. This pragmatic, application-specific approach to behavioral economics may ultimately be more productive than attempting to resolve whether the assumptions of neoclassical or behavioral models are correct at a general level." (Chetty (2015), p. 3)

In some sense, I share the sentiment expressed in this passage. Like Chetty, I am often impatient with debates about the general validity of behavioral assumptions. But the thing I find striking is the analogy between incorporating behavioral elements in economic analysis and the rather minor decision of whether to assume time-separable utility functions. In his

attempt to make BE more palatable to a general audience, Chetty has also made it seem *harmless*. We have thus come back full circle: the atheoretical trajectory of BE is coupled to what is effectively a denial of its revolutionary potential.

BE is *not* harmless. When one reads the works of Tversky and Kahneman from the 1970s, or the early works of Thaler on mental accounting, one encounters insights that undermine conventional economic modeling. They attack Bayesian probabilistic sophistication as an unrealistic description of how people reason about uncertainty. And they claim that preferences are so malleable, context-specific and prone to mental accounting as to render the notion of stable preferences meaningless. Reading these impressive works, the message that I perceive is that a powerful reimagining of economic theory at the foundational level is needed, one that is comparable to and perhaps exceeds that brought about by GT. Although the pragmatic approach to BE has its place, "arguing about theoretical principles" is necessary in order to fully realize its transformative potential. Ultimately, the question is whether future graduate-level textbooks on economic theory will have to be rewritten from scratch - as they were in the 1990s, in the wake of the GT revolution - or will a couple of new chapters suffice.

References

[1] Akerlof, G. A., & Shiller, R. J. (2015). Phishing for phools: The economics of manipulation and deception. Princeton University Press.

[2] Samuelson, W. F., & Bazerman, M. H. (1985). Negotiation under the winner's curse. Research in experimental economics, 3, 105-38.

[3] Bar-Gill, O. (2012). Seduction by contract: Law, economics, and psychology in consumer markets. Oxford University Press.

[4] Campbell, J. Y. (2016). Restoring Rational Choice: The Challenge of Consumer Financial Regulation. The American Economic Review, 106(5), 1-30.

[5] Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. The American Economic Review, 105(5), 1-33.

[6] DellaVigna, S., and U. Malmendier (2004). Contract design and self-control: Theory and evidence. The Quarterly Journal of Economics (2004): 353-402.

[7] DellaVigna, S., and U. Malmendier. (2006). Paying not to go to the gym. The American Economic Review, 96(3), 694-719.

[8] Eliaz, K., & Spiegler, R. (2006). Contracting with diversely naive agents. The Review of Economic Studies, 73(3), 689-714.

[9] Esponda, I. (2008). Behavioral equilibrium in economies with adverse selection. The American Economic Review, 98(4), 1269-1291.

[10] Eyster, E., & Rabin, M. (2005). Cursed equilibrium. Econometrica, 73(5), 1623-1672.

[11] Gabaix, X., & Laibson, D. (2005). Shrouded attributes, consumer myopia, and information suppression in competitive markets. Quarterly Journal of Economics, 121 (2), 505-540.

[12] Gabaix, X., Laibson, D., Li, D., Li, H., Resnick, S., & de Vries, C. G. (2016). The impact of competition on prices with numerous firms. Journal of Economic Theory, 165, 1-24.

[13] Grubb, M. D. (2009). Selling to overconfident consumers. The American Economic Review, 99(5), 1770-1807.

[14] Grubb, M. D. (2015). Overconfident consumers in the marketplace. The Journal of Economic Perspectives, 29(4), 9-35.

[15] Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. Econometrica, 69(6), 1403-1435.

[16] Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. The foundations of Positive and normative Economics: A handbook, 1.

[17] Jehiel, P. (2005). Analogy-based expectation equilibrium. Journal of Economic theory, 123(2), 81-104.

[18] Jehiel, P., & Koessler, F. (2008). Revisiting games of incomplete information with analogy-based expectations. Games and Economic Behavior, 62(2), 533-557.

[19] Kőszegi, B. (2014). Behavioral contract theory. Journal of Economic Literature, 52(4), 1075-1118.

[20] Heidhues, P., & Koszegi, B. (2010). Exploiting naivete about self-control in the credit market. The American Economic Review, 100(5), 2279-2303.

[21] Laibson, D., & List, J. A. (2015). Principles of (behavioral) economics. American Economic Review, 105(5), 385-90.

[22] Mullainathan, S., Schwartzstein, J., & Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. Economics, 4.

[23] Rabin, M. (2013). An approach to incorporating psychology into economics. The American Economic Review, 103(3), 617-622.

[24] Rubinstein, A. (1998). Modeling bounded rationality. MIT press.

[25] Rubinstein, A. (2006). Discussion of "Behavioral Economics". Econometric Society Monographs, 42, 246.

[26] Spiegler, R. (2006). Competition over agents with boundedly rational expectations. Theoretical Economics, 1(2), 207-231.

[27] Spiegler, R. (2008). On two points of view regarding revealed preference and behavioral economics. The foundations of positive and normative economics: A handbook, 95.

[28] Spiegler, R. (2011). Bounded rationality and industrial organization. Oxford University Press.

[29] Spiegler, R. (2015). On the equilibrium effects of nudging. The Journal of Legal Studies, 44(2), 389-416.

[30] Thaler, R. H. (2015). Misbehaving: The making of behavioral economics. WW Norton & Company.

[31] Thaler, R. H. (2016). Behavioral economics: Past, present and future. Present and Future. Mimeo, University of Chicago.

[32] Tirole, J. (1988). The theory of industrial organization. MIT press.

[33] Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psychological bulletin, 76(2), 105.