



HAL
open science

A controversy about modeling practices: the case of inequity aversion

Alexandre Truc, Dorian Jullien

► **To cite this version:**

Alexandre Truc, Dorian Jullien. A controversy about modeling practices: the case of inequity aversion. *Journal of Economic Methodology*, 2023, 30 (3), pp.203-227. 10.1080/1350178X.2023.2180153 . hal-04719263

HAL Id: hal-04719263

<https://hal.science/hal-04719263v1>

Submitted on 3 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A controversy about modelling practices: the case of inequity aversion

Alexandre Truc*, Dorian Jullien†

October 3, 2024

Abstract

Abstract: This paper studies the controversy on Fehr and Schmidt's model of inequity aversion. It borrows insights from disciplines such as philosophy and the sociology of science that have specialized in studying scientific controversies. Our goal is to contribute to the historical and methodological literature on behavioral economics, which happens to have neglected behavioral economists' research on social preferences. Our analysis of the controversy reveals some new insights about the relation of behavioral economics with other sub-fields in economics, as well as with other disciplines.

Keywords: Controversies; Behavioral Economics; Rhetoric; Social Preferences; Norms; Inequity Aversion

*Atruc@unice.fr. Université Côte d'Azur, CNRS (Centre national de la recherche scientifique) – GREDEG (Groupe de Recherche en Droit, Economie, Gestion), France.

†Centre d'Economie de la Sorbonne, Université Paris I - Panthéon Sorbonne

1 Introduction

The study of social preferences, or pro-social behavior, is the biggest sub-field of research in behavioral economics. It represents about 35% of the publications from behavioral economists after 2010 (Truc, 2022b, Figure 4b) and its most cited model, which is the inequity aversion model of Fehr and Schmidt (1999), is also the second most cited article in behavioral economics after Kahneman and Tversky (1979) according to Angrist et al. (2020, Table 1, 31). Yet, the historical and methodological literature on behavioral economics has rarely focused on this sub-field.¹ Some exceptions address various methodological tensions underlying the study of pro-social behavior in economics from a philosophical and theoretical perspective (Guala and Mittone, 2010; Guala, 2012; Paternotte and Grose, 2013; Lisciandra, 2018; Nagatsu, 2021; Vromen, 2022). However, the socio-historical dynamics underlying the way these tensions are addressed by economists remains understudied.

This absence of interest is even more surprising as the sub-field of social preferences has witnessed one of the longest and most virulent *methodological* controversies involving behavioral economics. This controversy started with an attack by Shaked (2005) on Fehr and Schmidt's (1999) work on inequity aversion, to which they responded (Fehr and Schmidt, 2005). Binmore and Shaked (2010a) then reiterated and amplified Shaked's (2005) criticisms, to which both Fehr and Schmidt (2010) and Eckel and Gintis (2010) responded, before Binmore and Shaked (2010b) concluded these exchanges by accusing Fehr and Schmidt of scientific malpractices representative of a larger part of behavioral economics. Our primary goal in this paper is to study this controversy to see what we can learn from it that could contribute to the historical and methodological literature on behavioral economics.

The secondary goal of this paper is to propose a toolbox for the study of controversies that can be useful for historians and methodologists of economics. It can be argued that most studies of controversies by historians and methodologists of economics are conducted rather intuitively. By "intuitively", we mean that the questions that are asked and the perspec-

¹Indeed, this subfield is not the focus of the contributions of Sent (2004); Heukelom (2014); Grüne-Yanoff (2016); Angner (2019); Heidl (2016); Hands (2010); Geiger (2017); Braesemann (2019); Davis (2010); Jullien and Vallois (2014), among others.

tives that are taken on the controversies under study are not inspired by the research on the study of scientific controversies in other disciplines. Rather, the questions and perspectives taken on economic controversies are derived from the economic knowledge and interests of the historians or methodologists.² We believe that this intuitive way of analyzing economic controversies can be nicely enriched by insights from decades of empirical and theoretical work on scientific controversies in some sub-fields of philosophy and sociology, among other disciplines. It is understandable that contributions in the history and methodology of economics do not draw on insights from these sub-fields because the corresponding literature is quite large and can seem to be conceptually messy at first sight, especially across disciplines. The toolbox that we propose consists of inductively extracted regularities from studies of scientific controversies that we systematically present as a set of insights. In other words, the purpose of this toolbox is to present a set of concepts and questions that allows taking a richer perspective on controversies than the one that is only based on intuitions and knowledge of economics. We want to stress that these tools are useful to study the history and methodology of economics, without implying normative claims about how economists should debate among themselves.

The paper is organized as follows. We introduce the above-mentioned toolbox for the study of scientific controversy in Section 2. We then present the controversy on inequity aversion in Section 3. Finally, we analyze that controversy in Section 4 with our toolbox.

2 A toolbox for the study of controversies

Scholars holding different epistemological positions may disagree over why and how to study scientific controversies. Raynaud (2015, chap.1) reviews what he calls a “metacontroversy”

²This is a general impression that we have, which we quickly checked by searching for the keyword “controversy” in the *Journal of Economic Methodology*, taking the first ten papers that were found and then proceeding to a quick reading of the papers and a much more attentive reading of the bibliographies. None of these papers used any work from other sub-disciplines that specialize in the study of scientific controversies. To the best of our knowledge, the only methodological paper referencing (very swiftly) one such reference is Mongin (2006). There are no doubt others like this one, but our point is that we aim to build a much larger and more systematic bridge with this literature.

between the “relativists” and the “rationalists” on this point – while acknowledging that more subtle oppositions can be drawn both between and within these positions (see also Kitcher, 2000). To put it bluntly, rationalists are interested in controversies as they constitute the empirical locus of criticisms from peers that are theoretically supposed to be responsible for scientific progress (see, e.g., the contributions in Machamer et al., 2000; Dascal, 1998; Fritz, 2019; or Raynaud, 2015 himself). From this perspective, the study of scientific controversies allows constructing accounts of scientific progress that are richer than the linear ones proposed by Popper (1934) or Lakatos (1978), while retaining their emphasis on the key role played by rational criticisms.³ Relativists, on the other hand, are interested in controversies as they make more visible the political, cultural, or material factors that are non-trivially present in the daily practice of science, but often absent from both rationalists’ and scientists’ accounts of scientific progress (see, e.g., the contributions cited by Pinch, 2015; Lemieux, 2007; Pestre, 2007; or Debaz, 2017). From this perspective, understanding how scientific knowledge is socially constructed is more important than giving an account of whether and how it is progressive or cumulative.

Lemieux (2007) argues that insights from these two ideal-typical perspectives should be combined to resist the two forms of reductionism that each of them tends to lead to: an “argumentative reductionism” from the rationalist perspective which focuses too much on theoretical demonstrations or empirical results and a “strategic reductionism” from the relativist perspective which focuses too much on the institutional and material context in which the scientists involved in a controversy evolve. Our goal is to present the insights from studies of scientific controversies in a balanced manner so as to avoid falling into these two forms of reductionism. This task is easier than it appears because most insights from the two camps are actually either similar or quite compatible when we abstract from foundational questions about the nature of scientific knowledge to focus on the details of what to look for in, and which questions to ask about, a scientific controversy. Contrary to some

³It should be noted that, contrary to what the sociological on scientific controversy might lead one to believe, philosophers of science have recently offered a rich set of approaches that go beyond the rationalist *vs.* relativist dichotomy (see e.g., Douglas, 2000; Longino, 2013; Chang, 2014; Chang, 2014; Chang, 2017). Małecka (2021) has recently promoted the use of such approaches for economic methodologists.

recent philosophical contributions that go beyond the rationalist *vs.* relativist dichotomy (see footnote 3), we focus on contributions that take scientific controversies not only as their main object of study, but also as the starting point of their analyses.

It is rather intuitive to conceive a scientific controversy as an exchange of criticisms and replies between (at least) *two* camps of scientists. However, most scientific controversies are public and it is easy to forget that a wider heterogeneous scientific community plays a non-trivial third role in the exchange. Indeed, the criticisms and replies explicitly exchanged between the two camps are usually tailored to convince that wider community more than the other camp (see, e.g., Machamer et al, 2000, 12; Lemieux, 2007, 195; Raynaud, 2015, 8). It can therefore be interesting, when studying a scientific controversy, to account for the audience targeted by the scientists in dispute. How are they targeting their audiences (e.g., explicit bibliographic references or implicit references to beliefs known to be common in a community)? Are they targeting the same audience? If not, what can be considered as relevant differences between the targeted audiences? Is there some reactions of approval or disapproval from the targeted audiences? Collins (1981b) proposed the concept of the “core-set” to refer to the two camps at the center of a scientific controversy and to the way they address a wider community:

“A core-set has been defined as being made up of those scientists deeply involved in experimentation or theorization which is directly relevant to a scientific controversy or debate [...] It is not always easy to define the boundaries of a core-set, because disputes within core-sets often involve the ‘boundary-work’ of trying to define people in or out – that is, defining them as legitimate or illegitimate commentators” (Collins and Evans, 2002, 242)

When the core-set and the wider community of a scientific controversy are identified, it may be relevant to study how the wider community deals with the stakes of the controversy, because it is not rare that a “debate closes down in the wider community before it closes in the core-set” (ibid, 246). As Lemieux (2007, 2005) puts it, controversies are less often settled (in the core-set) than “progressively *normalized*” (in the wider community).

That last remark is especially useful if one wants to achieve what could be taken as the most intuitive goal to be set when studying a scientific controversy: to identify how the controversy has achieved closure, i.e., how it has been resolved. There are some disagreements as to whether a controversy needs to have led to a clear solution to the initial problem to be worth studying (see Kovaka, 2021). Indeed, it is not always easy to be able to identify a clear winner in, or a clear epistemic gain from, a scientific controversy; but that does not prevent one from identifying relevant conceptual or methodological differences between the two camps (Dascal, 1998). Furthermore, even if a given scientific controversy may give the impression to achieve closure, several authors have pointed out that scientific controversies tend to be cyclical, i.e., to re-open under different forms at different points in time (see, e.g., Hacking, 2000; or Lemieux, 2007, 208).

Nevertheless, it certainly remains legitimate to study a controversy in order to identify the path-breaking theoretical or empirical results that may emerge from it (see, e.g., Dascal, 1998; Machamer et al., 2000). And while it is intuitive to focus on such results as they directly relate to the explicit initial object of a controversy, a broader perspective might often be desirable. Indeed, several scholars have insisted on the relevance to account for:

- How each camp tries to redefine constantly what this object is (Dascal, 1995);
- How such “*topic shift*” often leads to “*generalized questioning ... [about] factual, methodological and conceptual presuppositions*” (Dascal, 1998);
- How a controversy can be about diverse and interlinked objects (Fritz, 2019);
- How a controversy is not necessarily about what the two camps say it is about (Kovaka, 2021, 12-3);
- How “what the controversy is really about is a topic for controversy, often the most important one” (Freudenthal, 1998, 158).

Following these insights can broaden the scope of theoretical or empirical results that can be taken as consequences of the controversy under study. It should also be noted that, besides

theoretical or empirical matters, controversies can be (fully or partly) about methodology, e.g., over what counts as a sound inference or proper “methodological canons” (Kitcher, 2000, 29; see also Machamer et al, 2000, 8). When this is the case, they offer a nice opportunity to characterize opposing demarcation criteria used by scientists to differentiate science from non-science (Graham and Walsh, 2019).

All these insights can justify why one’s account of a controversy can be different from the accounts of those who are involved in it (as part of either the core-set or the wider community). Such a difference is also often justified by the potential of controversies to *reveal* elements that are of interest for the observer but not necessarily for the scientists involved. For instance, a clash between different interpretations of a given set of empirical results may reveal relevant differences in the scientific cultures of the two camps. Indeed, Collins (1981a) argues that scientific controversies are especially interesting as moments during which the “interpretative flexibility” of facts is most visible. Similarly, a clash between different interpretations of a given theory may reveal new information about the foundations of that theory (Freudenthal, 1998). More generally, controversies are often studied because they reveal presuppositions and background assumptions held more or less consciously by the scientists involved. For instance, it has been remarked that controversies in which the two camps put symmetrically different weights on theoretical and empirical results reveal different “styles of reasoning” that can be useful to update or contextualize the well-known opposition between rationalism and empiricism (Baltas, 2000, 43; see also Dascal, 1998, 151-152). From a different perspective, studying a controversy can reveal relevant information to characterize the structure of a scientific community (Freudenthal, 1998, 157), i.e., scientists from the core-set are pointing at others from the wider community as belonging to their camps and scientists from the wider community can express sympathy for a given camp. In some instances, an increasing number of scientists joining both camps and using different labels to designate their respective positions can reveal the definition of new “(sub)disciplinary boundaries” (Gavroglu, 2000, 177).

Finally, a last word should be said about one of the heaviest trends in studies of scientific

controversies, which consists in accounting for how they can overflow outside of academia and involve many non-scientists from ordinary citizens on social media to policymakers (see Hacking, 2000; Collins and Evans 2002; Pestre, 2007; Lemieux, 2007; Pinch, 2015; Dhermy-Mairal and Fages, 2017; Debaz, 2017). We do not focus on the insights that can be extracted from the contributions underlying that trend because it is not clear how they would be useful for the controversy we study below. Still, it is worth noting that the more a controversy is confined in a “*constitutive forum*” among peers – as opposed to a “*contingent forum*” that includes the public sphere (Raynaud, 2015, 14) – the more the rhetoric of the exchanges are likely to be civilized and polite and the dignity of the participants to be respected (Fabiani, 1997, 15; Lemieux, 2007, 195).

3 The controversy

3.1 The inequity aversion model

The model of inequity aversion that is at the center of the controversy we study was published in 1999 by Fehr and Schmidt, two economists respectively based in Zurich and Munich (in 1999 and still now) whose academic careers have unfolded mostly in Europe. By 1999, the main theoretical models in behavioral economics were Kahneman and Tversky’s (1979; Tversky and Kahneman, 1992) prospect theory for decisions under risk, Laibson’s (1997) model of quasi-hyperbolic discounting for inter-temporal decisions and Rabin’s (1993) incorporation of fairness in game theory for inter-personal decisions. All these models tried to account for some empirical regularities that are incompatible with standard models. For decisions under risk and inter-temporal decisions, the standard models are easily identified with expected utility theory and models with exponential discounting, respectively. Furthermore, the empirical regularities in these two domains tended to be interpreted by behavioral economists as instances of non-rational behavior and the standard models as providing relevant normative guidance regarding how to make rational decisions.

For inter-personal decisions, things are more complicated. Firstly, the standard model

here is not a clearly identifiable formal model that is applied differently in different situations (as are models with expected utility or exponential discounting). It is rather an assumption made more or less explicitly and more or less frequently in economic analysis, namely that individuals are guided only by their self-interest and do not derive (dis-)utility from what happens to other people. Secondly, the empirical regularities that are incompatible with this assumption tend to be interpreted by behavioral economists as instances of rational behavior. We can illustrate this point with one of the main regularities in that field that can be observed in the so-called ultimatum game (Güth et al., 1982). In that game, one person, the proposer, is given a sum of money and asked to divide it with another person, the responder, who then either accepts the offer so that both get some money, or refuses so that both get no money. That some people tend to reject “unfair” offers (e.g., 80% of the money for the proposer and 20% for the responder) is not interpreted as an instance of non-rational behavior by behavioral economists, while the self-interest assumption is arguably incompatible with preferring no money for oneself to a bit of money for oneself (which is the preference expressed by a rejection of unfair offers in the ultimatum game). This is the type of empirical regularities, interpreted as instances of rational behavior, that both Rabin’s (1993) and Fehr and Schmidt’s (1999) models try to account for.

Furthermore, while Rabin’s (1993) model was constructed to represent games between two players, such as the ultimatum game, Fehr and Schmidt’s (1999) is constructed to represent games with many players. This makes their model potentially applicable to so-called “experimental markets”, which are games involving competition among many players (see Smith and Williams, 1992, for a pedagogical introduction). Contrary to the ultimatum and related games, it is widely acknowledged (notably by all the participants to the controversy we study) that models using the assumption of self-interest account for the regularities observed in market games (which are often that a competitive equilibrium is attained). The main theoretical challenge faced by Fehr and Schmidt (1999) is therefore to account for regularities that are both compatible and incompatible with the assumption of self-interest (and to interpret all the underlying behavior as rational). To do so, their model consists in

the following relatively small departure from a traditional utility function. Consider a set of n individuals with respective incomes x_1, x_2, \dots, x_n . The utility function of any individual i is given by:

$$U_i = x_i - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max[x_j - x_i, 0] - \frac{\beta_i}{n-1} \sum_{j \neq i} \max[x_i - x_j, 0] \quad (1)$$

Where $\alpha_i \geq \beta_i$ and $0 \leq \beta_i \leq 1$. When $\alpha_i = \beta_i = 0$, the model reduces to the first term and is therefore equivalent to the assumption of self-interest because the individual only derives utility from his or her own income x_i . When $\alpha_i \neq 0$ or $\beta_i \neq 0$, the second term represents a loss of utility from an inequitable outcome that makes the individual worse off compared to the other individuals and the third term represents a loss of utility from an inequitable outcome that makes the individual better off compared to the other individuals. For instance, in an ultimatum game (hence with $n = 2$) with 10 euros to be divided, a proposer i who offers to keep 8 euros and to give 2 euros will obtain a utility of $U_i = 8 - 6\beta_i$ if the offer is accepted (and $U_i = 0$ if it is refused) and the responder j will obtain a utility of $U_j = 2 - 6\alpha_j$ if he or she accepts the offer (and $U_j = 0$ if he or she refuses it). As individuals are modeled as rational, they choose what maximizes their utility. Therefore, fixing the values of the α s and the β s is necessary to predict choices. Furthermore, Fehr and Schmidt (1999) model individuals as heterogeneous: the population of individuals playing a game is characterized by a joint distribution of different types (α, β) , and their predictions of the outcomes of games are derived from the interaction between different types (and each individual believes the correct distribution of types).

3.2 Part I: the first exchange

3.2.1 The pamphlet

The controversy that we study starts with a paper attacking the work of Fehr and Schmidt by Shaked (2005), an economist based at Bonn since 1989, where Schmidt happened to defend his thesis in 1991 – Shaked knew Schmidt and remembers him as one of the “best theory

students in Bonn”⁴. Shaked calls his 2005 paper a “pamphlet”. It is the result of an earlier version that he sent to Fehr and Schmidt, who did not respond, and to ten other colleagues who advised him to publish the paper online, but to tone down the language. This is what he did, along with sending this new version to 400 economists.

Shaked (2005) presents his original motivation as a problematic discrepancy he remarks about the “rhetoric” that Fehr and Schmidt use to describe the empirical applications of their model between, on the one hand, their 1999 seminal paper, and, on the other hand, the presentation of that seminal paper in a 2003 survey on social preferences that they were “invited to present ... in the distinguished 8th World Congress of the Econometric Society” (Shaked, 2005, 2). The main passages that bother Shaked are the following ones, in which he adds the bold characters:

“[From Fehr and Schmidt 1999]: ‘The objective is rather to offer a **first test** for whether there is a chance that our theory is consistent with the quantitative evidence from **different games**. Admittedly, this test is rather **crude**.’ [843], and again: ‘Clearly, the above computations provide only **rough** evidence in favor of our model’ [846].

[From Fehr and Schmidt 2003]: ‘Using the data that is available from many experiments on the ultimatum game, Fehr and Schmidt (1999) **calibrate** the distribution of α and β in the population. Keeping this distribution constant, they show that their model yields quantitatively **accurate** predictions across **many** bargaining, market, and co-operation games.’ [222]” (Shaked, 2005, 5-6)

Shaked (2005, 6) presents the main motivation for his criticisms by commenting on these passages as follows:

“It is unfair to criticize a crude and rough calculation which is meant to be only a first test, but now that [Fehr and Schmidt] see this calculation as a calibration

⁴Avner Shaked, e-mail to author, January 27, 2020 (see Appendix 6.1).

which yields accurate predictions, we may be justified in applying stringent measures to test their claims. A careful study of their computation will show that their first description of it is by far the more suitable one.”

He then proceeds to point out a very large number of problems in Fehr and Schmidt’s (1999) applications of their model to experimental games, which he regroups in two categories, that can be summarized as follows. The first category concerns Fehr and Schmidt’s (1999) calibration of their model *and* their use of the calibrated model to derive predictions for the outcomes of experimental games, both of which contain many methodological flaws according to Shaked. The calibration of Fehr and Schmidt’s (1999) model consists in fixing a joint distribution of types of players of the form “ $x\%$ of players are ‘ $(\alpha = y, \beta = z)$ type’”, i.e., the percentages of players for whom their α and β can be characterized by certain couples of values. Fehr and Schmidt (1999) decided to use data from ultimatum games for this calibration. However, as players express *either* their α (as responders) *or* their β (as proposers) in an ultimatum game, the data cannot be used to derive a *joint* distribution of types of players, but only separate distributions of the parameters α and β in the population (Shaked, 2005, 6). The distributions that they use for their calibration are the following ones (Fehr and Schmidt, 1999, 844):

- 30% of $\alpha = 0$, 30% of $\alpha = 0.5$, 30% of $\alpha = 1$, 10% of $\alpha = 4$
- 30% of $\beta = 0$, 30% of $\beta = 0.25$, 40% of $\beta = 0.6$

Shaked (2005, 7-8) points out that, since the data from ultimatum games presented by Fehr and Schmidt warrant the inference of only some parts of these distributions, Fehr and Schmidt should have explained how they chose the missing parts. In particular, their model implies that proposers who offer to share half of the money have a $\beta \geq 0.5$. Shaked claims that the arbitrary choice of $\beta = 0.6$ is motivated by the results that Fehr and Schmidt aim to predict for other games, which should therefore be banned as possible targets for the predictions of the calibrated model. In the same spirit, Shaked (*ibid*, 12-13) also points out that Fehr and Schmidt assume a perfect correlation between these two distributions (i.e.,

30% for $\alpha = \beta = 0$, 30% for $\alpha = 0.5$ and $\beta = 0.25$, 30% for $\alpha = 1$ and $\beta = 0.6$, 10% for $\alpha = 4$ and $\beta = 0.6$) based on the outcomes they aim to predict for a public good game with punishment. More generally, Shaked's point is that Fehr and Schmidt have too many degrees of freedom in the way they calibrated and hypothesized their parameters. Therefore, they can choose arbitrary values that make the model's predictions look good rather than using values imposed more strictly by the data.

The second category of problems pointed out by Shaked concern Fehr and Schmidt's (1999) claim that, even with a population that is highly averse to inequity, their model is compatible with many results from market experiments conducted by Smith and others (though the experiments involving market interactions of which they aim to predict the outcomes are not conducted by Smith but by other market experimentalists). In these experiments, the competitive equilibrium often observed empirically is rather inequitable and traditionally explained with models that assume that players are self-interested. Shaked claims that many simplifications made by Fehr and Schmidt (1999) to predict competitive equilibria are not warranted and that without these simplifications their model implies that a single inequity-averse player can enforce an equitable equilibrium, which is precisely contrary to what Fehr and Schmidt (1999) aimed to prove. For instance, the feature of the model that allows the competitive equilibrium to be attained in one market game (ultimatum game with competition among proposers) is the normalization of the function by $n - 1$. In Fehr and Schmidt (1999) this normalization is justified by the simplification that an individual player compares himself to the rest of the group but does not care for the inequality within the group as a whole. As the number of players increases so that competition among players intensifies, the probability that the competitive equilibrium will be attained increases. However, Shaked is critical of the very idea of normalization that he sees as a way to make inequity aversion compatible with competitive equilibrium rather than an empirically justified assumption: "The normalization assumption was introduced with no attempt to justify it [...], and there seems to be no compelling reason why this consequence should be true. Is it really reasonable that an individual is less likely to contribute to a charity, helping the victims of a natural

disaster, just because the number of victims increased?” (Shaked, 2005, 15).

To summarize, for Shaked, Fehr and Schmidt calibrate their model and establish its hypotheses in an arbitrary way, unconstrained by empirical results, to give the impression that their model is better than standard models using the assumption of self-interest. It should be noted that Shaked (2005) discusses many more methodological problems that are more specific to each prediction derived by Fehr and Schmidt (1999) for particular games.

3.2.2 The response

When they received Shaked’s paper, Fehr and Schmidt found it “unusual because of its aggressive and hostile tone” but also “poorly researched”.⁵ They decided to respond because Shaked had sent out the paper by email to many economists despite finding it unusual that the paper was not submitted to a journal to be peer-reviewed.⁶

Fehr and Schmidt’s (2005) reply, which is also an online paper, addresses three of Shaked’s points. The first, which we did not discuss in the previous section, is quite straightforward: Shaked claims that a proof of Fehr and Schmidt is false. Fehr and Schmidt acknowledge a mistake, but they point out that it is only a typo in the paper that does not invalidate the proof.

Second, they tackle the issue of the distribution of the parameters in the model inferred from the ultimatum game. The questions addressed by Fehr and Schmidt are numerous and quite specific, though they do not address Shaked’s general remark that their theory supposes to use a joint distribution of the parameters rather than separate distributions. They do however address their inability to find the exact value of β for equal splits in the ultimatum game. They reject Shaked’s assertion that any value above 0.5 could be chosen because they have a constraint that $\beta < 1$ (Shaked does state this constraint when he presents their model but not when he criticizes their arbitrary choice of $\beta = 0.6$). Shaked (2005) showed that with a value of 0.84, the accuracy of their prediction of a competitive equilibrium for a game dropped from 80% to 50%. Fehr and Schmidt respond that a value of 0.8 would reflect

⁵Klaus Schmidt, e-mail to author, January 24, 2020.

⁶*Ibid.*

quite extreme behavior when compared to results from the experiments. It would imply that players are ready to sacrifice one dollar of their payoff to improve the payoff of the other by 25 cents (Fehr and Schmidt, 2005, 5). Fehr and Schmidt then justify that they pick the value of 0.6 because it “implies that such a player is willing to give away one dollar if this increases his opponent’s payoff by at least 66 cents, which seemed more *plausible* to us” (2005, 5, our emphasis). Furthermore, they judge that getting the correct prediction 50% of the time with Shaked’s extreme suggestion is rather a point in favor of their model (ibid, 6). Following a similar logic, they also justify the perfect correlation between β and α by saying that it is “assumed for simplicity” but also that it is “plausible” (Fehr and Schmidt, 2005, 7).

Furthermore, they address the discrepancy between the rhetoric of their 1999 seminal paper and their 2003 survey by arguing that space constraints prevented them to go into more detail about their methodology so that ““calibration” seems to us to be the best description of our procedure in such a short summary” (ibid, 8). They also argued that some of their experiments on contract games cited in their 2003 survey but not in their 1999 paper (and not discussed by Shaked) “are not only consistent with our model of inequity aversion, but also with the calibration that we used in QJE” which “gives us the confidence to claim that our model ‘yields quantitatively accurate predictions’” (ibid).

More generally, they argue that their methodology is commonplace in the economic literature as every model faces the same challenges and employs simplifications (Fehr and Schmidt, 2005, 8). While they admit that some of the calibration and hypothesis are based on simplified assumptions or plausible behavior that is consistent with observed behavior rather than precise calibration, for them it is the challenge that any economic model faces:

We have to admit that the terms “fine calibration” and “full calibration” are new to us. But we never claimed that we “finely” or “fully” calibrated our model. We just said that we calibrated it, which involves the same problems that many other calibrations in the literature face as well. (Fehr and Schmidt 2005, 8)

Third, they address the issue of the model’s compatibility with market experiments by dismissing Shaked’s statements on a similar basis. For example, they tested the hypothesis

regarding the normalization of the utility function in a subsequent experiment:

“This is exactly what we observe in the Fischbacher et al. (2003) experiment. The introduction of one competing proper (going from $n=2$ to $n=3$) yields a large increase (away from the equal split) in the proposed offer. Shaked cites this paper but fails to mention that one of us did exactly the experiment that he would like to have seen” (Fehr and Schmidt, 2005, 10).

They also explicitly address the normalization argument we presented in the previous section as containing a logical flaw:

We do not want to speculate about the motives of people giving to tsunami victims here, but it is worthwhile to point out the logical flaw in Shaked’s argument. Even if one applies our theory to charitable giving, it does not say that a person is less likely to give if the number of victims increases, but rather if the reference group increases (Fehr and Schmidt 2005, 10, original emphasis).

At this point of the controversy, we can make the following two remarks. First, the controversy is very specific and does not spill over into more general epistemological problems. While there is a clear methodological divide between Shaked and Fehr and Schmidt in the way empirical and theoretical work should be articulated, the controversy is dedicated to very specific points concerning the seminal paper of Fehr and Schmidt (1999). Second, both sides were dissatisfied with the controversy. Shaked sees Fehr and Schmidt’s methodology as fundamentally flawed and thinks their answer failed to address his arguments. Fehr and Schmidt, on their side, considered Shaked’s main points refuted and the controversy unfounded.

3.3 Part II: the second exchange, back to academia

After this informal exchange outside of traditional academic boundaries, the controversy was integrated back into academia five years later as a special issue of the *Journal of Eco-*

conomic Behavior & Organization on “Issues in the Methodology of Experimental Economics” dedicated a whole part to it.

3.3.1 Binmore joins Shaked

Shaked felt pressure from his colleagues to move the controversies to an academic journal so that the details of the criticism could be published.⁷ He did so with a paper entitled “The Explanatory Value of the Theory of Inequity Aversion” (Shaked, 2006), a detailed version of the paper discussed in Part I. The paper was rejected in two journals⁸. This is the point when Binmore joins the controversy. Binmore was shown the report by Shaked and decided to get involved and help rewrite the paper in a milder style. The paper also proved difficult to publish because of a lack of interest, until Barkley Rosser, one of JEBO’s editors proposed to devote a whole issue to the methodology of experimental economics with a part dedicated to the controversy. The first notable element of Binmore and Shaked’s (2010) paper is that it expanded Shaked’s (2005) scope on two dimensions.

Firstly, they addressed methodological and epistemological issues at a more general level, independently of Fehr and Schmidt’s work. Secondly, they present their criticisms of the work of Fehr and Schmidt as a case study representative of unscientific practices shared among most behavioral economists.

The introduction of the paper frames part of the problem as being about the opposition between science and ideology. Binmore and Shaked contrast the good experimental economists who seek recognition by “adopting the scientific standards” of neighboring disciplines with the bad tradition of policy-orientated economists who treat “experimental results as just another rhetorical tool to be quoted when convenient in seeking to convert others to whatever your own point of view may be” (Binmore and Shaked, 2010a, 87).

The first two sections of the paper are indeed dedicated to framing the controversy using very broad philosophical and methodological problems in economics and science.

The first section is a general characterization of economics as a scientific discipline. They

⁷Avner Shaked, e-mail to author, January 27, 2020.

⁸Avner Shaked, e-mail to author, January 27, 2020.

start by laying out neoclassical economics as an “optimizing paradigm”. This includes the point that economists do not account for taste (and therefore do not claim that people are selfish as the framework can be adapted to include selfish and non-selfish agents alike), they are not committed to any psychological interpretation of utility, and the maximization of money in games does not imply selfish maximization of well-being. As they point out, one may wish to maximize money on certain occasions (such as when playing experimental games in a laboratory) only to redistribute that money to other people on other occasions. For Binmore and Shaked, by identifying the maximization of money in games with self-interested behavior, and by taking backward induction as a standard of neoclassical economics, behavioral economists are attacking a straw man that is not neoclassical economics.

The second section offers a framework for “Testing theories scientifically”. According to them, economists should “follow what is regarded as best practice in other sciences” and focus on prediction, which they see as “the gold standard” in science (Binmore and Shaked, 2010a, 89). Binmore and Shaked insist on differentiating predicting from fitting. Economists should clearly define the domain of application of their theory and test it with data that was not used to formulate and calibrate the theory. The example of Ptolemy’s epicycle movements is mobilized, and Binmore and Shaked note: “It is, therefore, necessary to be very careful when parameters are left floating and so are available to be fitted to new data that is supposedly being predicted” (Binmore and Shaked, 2010a, 89). This section also includes general advice such as “Respecting logic”, replicating experiments before accepting their results, and “Serious refereeing”: “Our claims in this paper have been checked down to the last rounding error by a referee who took his duty very seriously. If the same had been true of the referees of the papers examined in the rest of this paper, we would have no grounds for complaint.” (Binmore and Shaked, 2010a, 90).

After establishing what they see as uncontroversial good practices, Binmore and Shaked move on to evaluate the scientificity of Fehr and Schmidt’s work on inequity aversion. They motivate the choice of their work as a case study by the popularity of the model, as the profession seems to have accepted it uncritically (Binmore and Shaked, 2010a, 90)⁹. Bin-

⁹Binmore and Shaked (2010a) mention the 2 390 citations for Fehr and Schmidt (1999) as a measurement

more and Shaked (2010a) reiterate the problems pointed out by Shaked (2005) about the calibration of the model to stress that, although Fehr and Schmidt have made methodological declarations that seem to agree with the view that prediction is the gold standard in science, their practices contradict these declarations. For Binmore and Shaked (2010a, 91), what Fehr and Schmidt are doing could at best be called a “parametrization” of their model. Still, in the same spirit as Shaked (2005), they argue that Fehr and Schmidt’s parametrized model does not predict the outcomes of their targeted experimental games better than the money-maximizing model, and when it does it is because of the high degree of freedom. They also criticize Fehr and Schmidt’s (2005) claim that the post-1999 contract games provide further support for their parametrized model. Among the many methodological problems they address regarding Fehr and Schmidt’s application of their model to predict the outcomes of these games, they notably point out that Fehr and Schmidt changed the initial 1999 distribution of the parameters so that the same distribution is not anymore constant across games. More generally, for Binmore and Shaked (2010a), Fehr and Schmidt “echo Milton Friedman’s (1953) defense of the Chicago ethos by observing that the subjects behave ‘as if’ motivated by inequity aversion and that their theory ‘helps to organize and interpret the data’.” (Binmore and Shaked, 2010a, 97)

Binmore and Shaked (2010a) also present their favored alternative approach to social preferences in order to explain pro-social behavior:

“The alternative approach we favor explains the behavior of subjects in terms of social norms. We think it likely that people enter laboratories primed with a variety of social norms, one of which is triggered by the manner in which the experiment is framed. If the resulting behavior is close to a Nash equilibrium of the game (as in the Ultimatum Game), then the social norm is stabilized in the laboratory environment. If it is not (as in the Prisoners’ Dilemma), then the subjects’ behavior moves towards a Nash equilibrium.” (Binmore and Shaked, 2010a, 98)

of its impact, which reached more than 10 000 citations in 2018 according to Google scholar.

Overall, as in Shaked’s first paper, Binmore and Shaked’s argument is that inequity aversion has too many degrees of freedom in its parameters, thus allowing Fehr and Schmidt to adapt the model as they wish between papers and between experiments. While the paper starts by putting aside the controversial aspect of the first exchange, the paper is still full of strong rhetoric. Most notably, it suggests a strong divide between the two communities as it represents a danger to Science: “Our aim was simply to draw the attention of mainstream economists to the danger of tolerating practices that would be regarded as unscientific in other disciplines.” (Binmore and Shaked, 2010a, 99)

3.3.2 The response: Eckel and Gintis join Fehr and Schmidt

The response from Fehr and Schmidt (2010) focuses on the four claims by Binmore and Shaked (2010a) and answers them with quotes from their 1999 article to show, for example, that they clearly state that the distribution of parameters consistent with the ultimatum game they used initially was not intended to be unique or definite. They argue that the different distributions they used in subsequent experiments were just a simplified version of the one identified in Fehr and Schmidt (1999). They accept Binmore and Shaked’s characterization of what they do as “parameterization” rather than “calibration” with slight contempt: “If Binmore and Shaked prefer to call our exercise a ‘parameterization’ then this is fine with us” (Fehr and Schmidt, 2010, 104).

However, Fehr and Schmidt (2010) also engage Binmore and Shaked on larger problems and use the controversy as a platform to clarify their epistemological commitments. For Fehr and Schmidt, there is a wider context to take into account to understand the contribution of inequity aversion. By putting “inequity aversion into perspective” they emphasize once again that their model is simple and does not have all the ambitions that Binmore and Shaked suggest it has. Fehr and Schmidt state that their main priority is to contrast selfish and fair behavior among individuals and use the model as a starting point for more experiments.

Moreover, Fehr and Schmidt go on to attack Binmore and Shaked on their claim that a theory of social norm could provide explanations of the pro-social behavior observed in

the lab that would be just as good, if not better, than the ones provided by a theory of social preferences. For Fehr and Schmidt (2010, 103) their approach could be characterized as unscientific by using Binmore and Shaked's falsification arguments they just mobilized: they "have complete freedom to rationalize any result *ex post* that is compatible with a Nash equilibrium by choosing the appropriate social norm that explains the behavior". Similarly, they can explain differences in behavior between market and ultimatum games by "arbitrarily invoking different social norms".

The end of their reply is dedicated to precisely that question. When should a social theory be deemed refuted? For Fehr and Schmidt, using the falsificationist perspective of Binmore and Shaked, "any theory of social behavior is refuted" (Fehr and Schmidt, 2010, 105). Theories of human behavior will always be incomplete and imperfect, especially when focused on heterogeneous behavior. But they are still useful if they help to interpret and better understand data patterns. Inequity aversion is a model oriented towards making empirical investigation easier and suggesting future experiments.

Despite the formal format of the controversy, Fehr and Schmidt still consider it a relatively unproductive enterprise (Fehr and Schmidt, 2010, 102) and an "unproductive debate" (Fehr and Schmidt, 2010, 107). They contrast the controversy by setting two future empirical tasks for behavioral economists. First, in line with what Bruhin et al. (2010) did with prospect theory, economists should offer a parsimonious characterization of the distribution of social preference types. Second, while economists should look into improving theoretical models, they should remain parsimonious and deliberately abstracted from empirical facts because they are just "tools" to help economists think about empirical data and motivate new experiments and hypotheses.

A noteworthy aspect of the response from Fehr and Schmidt is that it is accompanied by another response to Binmore and Shaked, by Eckel and Gintis (2010). In a nutshell, this additional response emphasizes the large scope of experimental data that contradicts the assumption of self-interest and argues that theoreticians such as Binmore and Shaked should not criticize experimentalists such as Fehr and Schmidt for not having proposed a satisfying

theoretical model.

Overall, on both sides, similar core arguments to the ones in the first exchange are repeated, but this time both sides put their own perspective in a larger framework. Binmore and Shaked emphasized respecting logic, good scientific practices, and the future of economic theory. For Fehr, Schmidt, Eckel and Gintis, the main point was about the changing relationship between theory and empirical work in behavioral economics, setting up the ground for future endeavors in BE, and following reasonable and looser epistemological principles than the strict falsificationist perspective laid out by Binmore and Shaked.

The controversy concluded with a last short rejoinder by Binmore and Shaked (2010b) who explicitly accuses Fehr and Schmidt of being unscientific.

4 Analysis

4.1 The controversy

4.1.1 The rhetoric and frame

Drawing on the insights from Section 2, a first element of context to note about our controversy is that it unfolds within an unconventional forum. The main specificity comes from the first exchange, which is addressed to scientific peers but does not take place within the traditional institutional boundaries of peer-reviewed publication. As such, the scientific community is unable to exert control over the rhetorical devices used by the participants. The second exchange is subjected to peer-reviewing, thus our controversy offers a nice opportunity to study how this change of context impacts the rhetorical behavior of the participants.

Shaked (2005) was drawn to publish his comments online rather than in a traditional journal because he considered the editorial process of economics publications to be flawed by a lack of published discussions, comments, and responses to published papers within the discipline, especially regarding methodology.¹⁰ Shaked's very first version (sent to Fehr and Schmidt and ten colleagues) initially used a very strong rhetoric, that can be decomposed

¹⁰Avner Shaked, e-mail to author, June 11, 2021.

into two dimensions: one involves irony, sarcasm, and satire directed at Fehr and Schmidt, while the other involves aggrandizing the scientific stakes of the controversy. These strong rhetorical elements served a strategic purpose. Despite some of the feedback from colleagues to tone down the language, Shaked “felt strongly about the issue and believed that this way it will make people pay more attention”.¹¹ After 2005 and the successful rise of behavioral economics and the Nobel of Kahneman, it was more generally perceived that behavioral economics had the favor of the economics profession and the Nobel world, making critics harder to publish.¹²

In the non-peer-reviewed pamphlet, Shaked criticized the “aggressive marketing techniques” of Fehr and Schmidt, which “cast a dark shadow” on the theory (Shaked, 2005, 3). Shaked’s somber tone is present throughout as he describes his accusations as “grave charges” (Shaked, 2005, 3). Similarly, the pamphlet emphasizes the separateness of BE from the mainstream of the profession. Shaked considers himself unfamiliar with the social preferences’ literature, and he asserts the separation between the work of “main-line” economics and the “subculture” of BE:

For my shame, I have not read this literature until recently. When I finally came to read the two papers, I was introduced to a *subculture hitherto unknown to me*, a subculture that apparently *coexists parallel to main-line economics* and in which *different rules of logic and different laws of proof* apply (Shaked 2005, 2, emphasis added)

The issue for Shaked is also larger than just the theory. It is more largely a failure of the economic profession and of the peer-review process, thus the strong rhetoric and the choice of a pamphlet:

But the sad point is not only that two distinguished economists allow themselves to write in this style, it is also that they managed to get by, that these papers have caused no outcry or protest in the experimentalist community. Could it be

¹¹Avner Shaked, e-mail to author, June 11, 2021

¹²Avner Shaked, e-mail to author, January 27, 2020.

that, with the exception of [*Quarterly Journal of Economics*]'s proofreader, no one has ever carefully read the details of the paper until now? (Shaked 2005, 18)

In their response, Fehr and Schmidt also attack Shaked's pamphlet with a rhetoric that does not abide by traditional academic norms. Starting their response to Shaked by focusing on the typo issue, they tease Shaked by accusing him of bad faith and pointing out typos in his own pamphlet:

Shaked did not look at the proof, but had he made any effort to understand the argument behind Proposition 4b, he would have quickly discovered the typo and not made a false accusation.

The typo in Proposition 4b is embarrassing and we apologize to our readers. However, we find some comfort in the fact that even A. Shaked is not immune to similar mistakes. When he summarizes our model he states our parameter restrictions as $0 < \beta_i < 1$ and $\beta_i < \alpha_i \dots$, while the correct statement would have been $0 \leq \beta_i < 1$ and $\beta_i \leq \alpha_i$. (Fehr and Schmidt 2005, 4)

In their conclusion, they return Shaked's initial statement, but this time by putting him in the "literary criticism" discourse rather than the scientific one:

Shaked has introduced us to a subculture hitherto unknown to us, a world of literary criticism in which different rules of logic and different rules of proof apply. In this sphere it seems permitted to misquote other people, to ignore crucial information that is provided in footnotes, appendices and even in the body of the text, to treat quotations in a casual and nonchalant way, and it is allowed to make grave charges without substantiating them.

We very much hope that he becomes an economist again and contributes to the exciting scientific discussion of the insights of experimental economics. (Fehr and Schmidt 2005, 11)

The strong rhetoric of these first exchanges was an obstacle in the process of moving the controversy into a more traditional peer-reviewed institutional context. The part of the special issue on the controversy was actually quite difficult to publish.¹³ Barkley Rosser considers it “probably the single most difficult episode in my time as a journal editor”. The first version of the paper by Binmore and Shaked required a lot of revisions to reduce “over-the-top personal attacks”, and similar problems emerged with Fehr and Schmidt’s response. An earlier working version of the paper of Binmore and Shaked is available online, and is instead - more provocatively - titled “Experimental economics: science or what?” instead of “Experimental economics: Where next?”.

It was also difficult for Rosser to convince Fehr to participate, and conditions were laid down so that the reply could be made. For example, according to Shaked the question of a potential Nobel Prize for Fehr and Schmidt should not be discussed.¹⁴ Obviously, such a controversy could impact the chance of Fehr and Schmidt to receive the Nobel, but the necessity to impose explicit rules in the discussion is a testimony of some of the difficulty encountered by any attempts to bring the controversy back into traditional academic discourse.

By the end of the editorial process, the ironical, sarcastic, and satirical dimension of the controversy disappeared (i.e., the personal attacks), but the aggrandizing dimension with respect to the scientific stakes remained and even amplified as the controversy actually grew in scope. For instance, Binmore and Shaked concluded the controversy as follows:

Is it scientific to proceed like Fehr and Schmidt or is it not? Fehr and Schmidt say yes. So do Eckel and Gintis. The implications are quite far-reaching for those like us who think it is obvious that the answer is no. (Binmore and Shaked 2010b, 120)

In a sense, it can be argued that the strong rhetoric of the controversy makes it explicit that proponents are not aiming to achieve a consensus, and dropping the personal dimension of the rhetoric was not sufficient to find a path toward closure.

¹³Barkley Rosser, e-mail to author, January 28, 2020.

¹⁴Avner Shaked, e-mail to author, January 27, 2020.

4.1.2 Scope of the controversy and its surroundings

Using the toolbox presented in section 2, we can remark that, as the controversy unfolds, it becomes broader in at least two senses: (1) the core-set of scientists involved expands, and (2) the object of the controversy becomes larger.

Regarding the expansion of the core-set, the first exchange is very local and only involves Shaked, Fehr and Schmidt (with two having shared the same institutional affiliation). After the first exchange, the opposition between the two camps becomes explicitly a larger issue between two scientific communities rather than a personal/local issue involving three or four specialists of the question at stake. This expansion is made most visible by the fact that Binmore joined Shaked as a co-author and Eckel and Gintis joined Fehr and Schmidt. The intervention of Eckel and Gintis illustrates rather well the boundary work done in a core set of trying to define who is legitimate to comment on the issue of the controversy. They suggest that Binmore and Shaked's criticism might be misplaced as the roles of the theorist and the experimentalist differ: "‘You broke it [the theory], so you go fix it’ may be appropriate in some walks of life, but not for the halls of science." (Eckel and Gintis, 2010, 117)

What we can learn from this controversy that is not often (if at all) stressed by the literature discussed in section 2 is that this expansion of the core-set also has a less visible part. On the one hand, Rosser, Binmore, and Shaked all acknowledge that Binmore and Shaked's paper faced serious opposition from a dozen of referees.¹⁵ While Rosser emphasizes that the tone of the paper was one big problem, for Binmore these reports were also related to Fehr and Schmidt's supporters playing against them. On the other hand, Binmore also points out the (rather invisible) role of the journal's editorial board in another paper, as he describes threats against the editor after receiving the reports, which were "so hostile that the editor was apparently threatened by a group of his editorial board with being fired if JEBO published our paper" (Binmore, 2011, 180). Taken together, this relatively strong but not publicly visible expansion of the core-set in favor of Fehr and Schmidt's camp can be explained by the fact that the JEBO is a central journal for behavioral economists. In other

¹⁵Barkley Rosser, e-mail to author, January 28, 2020.

words, Binmore and Shaked published their criticism in a rather hostile territory.

Regarding the enlargement of the object of the controversy, while the first exchange was focused on the empirical application of Fehr and Schmidt’s model, which could arguably be settled (at least in principle), the second exchange in the JEBO embeds this issue in more general epistemological issues. Binmore and Shaked reminded readers of the rule of logic and scientific reasoning, and Fehr and Schmidt reflected on the methodology of behavioral and experimental economics. This is clearly an instance of a *topic shift* that leads to *general questioning* (Dascal, 1998). While the first exchange has something very personal in the rhetoric, the second one leads the authors to draw strict boundaries between behavioral economics and more standard economics in a demarcationist perspective (Binmore and Shaked, 2010b, 120).

This fits nicely with Walsh and Graham (2019) for whom most *scientific* controversies feature opposing demarcation criteria to differentiate science from non-science. We can add that this leads to issues that are arguably much less likely to be settled (than the technical ones of the first exchange). In this case, the participants fail to make their opposing demarcation criteria explicitly enough to at least agree to disagree.

4.1.3 After the controversy

A natural question to ask is whether our controversy has achieved any kind of closure (see Section 2). In our case, it is clear that there is no emerging consensus between the two camps or a clear epistemic gain, as both camps recognize that the controversy was mostly unproductive (Fehr and Schmidt, 2010, 102, fn1; Binmore and Shaked, 2010b)¹⁶. However, as is claimed in most references used in Section 2, the way in which the controversy does not achieve closure can be interesting in itself. In the words of Kitcher (2000, 22): “Each resolution of a scientific controversy proceeds in much the same way; all unresolved scientific controversies are unresolved in their own way”.

How exactly, then, is our controversy unresolved? First of all, one year after the con-

¹⁶The e-mails referenced in the previous footnotes all showed that this impression still held ten years after the JEBO exchange for Shaked, Binmore, Fehr and Schmidt

troversty, we can find remnants of it. For instance, Binmore, in a 2011 review of Bardsley (2010) with no relation to the controversy, renewed his criticisms against the “propaganda of behavioral enthusiasts” (Binmore, 2011, 178) and “the flouting of the most basic of scientific principles” (Binmore, 2011, 179). Binmore further argues in that paper that the controversy affected the way the community perceived the theory of inequity aversion:

Experimental economists seem to have accepted, for example, that Fehr and Schmidt’s claims for inequity aversion are unfounded. A typical comment is: “I never believed in inequity aversion anyway.” However, nobody seems willing to take the point that such bad science is not confined to the theory of inequity aversion, but is endemic in much of our current practice (Binmore 2011, 180).

In 2020, Fehr and Schmidt did not share this view and believed that the controversy did not have a visible impact.¹⁷

From a quantitative point of view, when we look at the citations of Fehr and Schmidt (1999), both camps seem to be right if we take into account that they expressed their opinions at different periods. We see in Figure 1 that the controversy seems to have affected the popularity of Fehr and Schmidt’s (1999) model, as we do find a dip in yearly citations a couple of years after the second part of the controversy (the citation pattern of Rabin’s 2003 model suggests that this dip is indeed proper to Fehr and Schmidt’s model). However, the yearly citations of Fehr and Schmidt’s (1999) recovered rather quickly. Even if Binmore’s statement was to remain a true description of most economists’ attitudes towards Fehr and Schmidt’s model, most of them may still continue to cite Fehr and Schmidt (1999) as a seminal contribution in BE’s study of social preferences.¹⁸

Beyond the effect of the controversy on the reception of Fehr and Schmidt’s model, the effect of the controversy in economics as a whole seems to be even less considerable. From a quantitative perspective, at the beginning of 2022 and according to Google Scholar, Binmore and Shaked’s (2010a) paper is cited 305 times, Fehr and Schmidt’s (2010) paper

¹⁷Schmidt Klaus, e-mail to author, January 24, 2020.

¹⁸It should be noted that we do not equate the citation patterns presented here as an indication of the quality of the content of the cited papers, but only as an indication of the reactions of the community.

is cited 115 times, which are both relatively small as compared to the 13,700 citations of Fehr and Schmidt (1999). From a qualitative perspective, the papers that do cite Binmore and Shaked (2010a) or Fehr and Schmidt (2010), tend to just acknowledge very rapidly the existence of the controversy without taking sides (regardless of whether their work are in line with Fehr and Schmidt’s work or an alternative to it). Here are two representative examples:

“There is some debate about just how widely and accurately the FS model applies; see Binmore and Shaked (2010) and Fehr and Schmidt (2010).” (Dreber et al., 2014, 42, f2)

“Importantly, we do not intend to ‘take sides’ in the specific criticisms that Binmore and Shaked put forward with respect to the Fehr-Schmidt model. See Fehr and Schmidt (2010) for a concise reply to these critiques.” (Yang et al., 2016, 2, f2)

Overall, we can consider that this controversy is “dissolved” rather than “solved” (Dascal, 1998, 150) or that it has been progressively normalized in the community (Lemieux, 2005).

4.2 The controversy in perspective

4.2.1 Cycles of controversies and delimitation of sub-disciplinary boundaries

Most insights that we derived from the study of controversies in Section 2 suggest that a historical perspective is often necessary to appreciate the full meaning of a given controversy. We propose to focus on two of these insights, namely the extent to which our controversy can be interpreted (1) as one occurrence of a broader cyclical controversy that (2) reveals the structure of a scientific community and notably how sub-disciplinary boundaries are progressively delimited.

The controversy we studied is not the first one in the history of behavioral economics. The literature on the history of behavioral economics shows that its early developments were accompanied by several controversial episodes with the experimental tradition of Smith

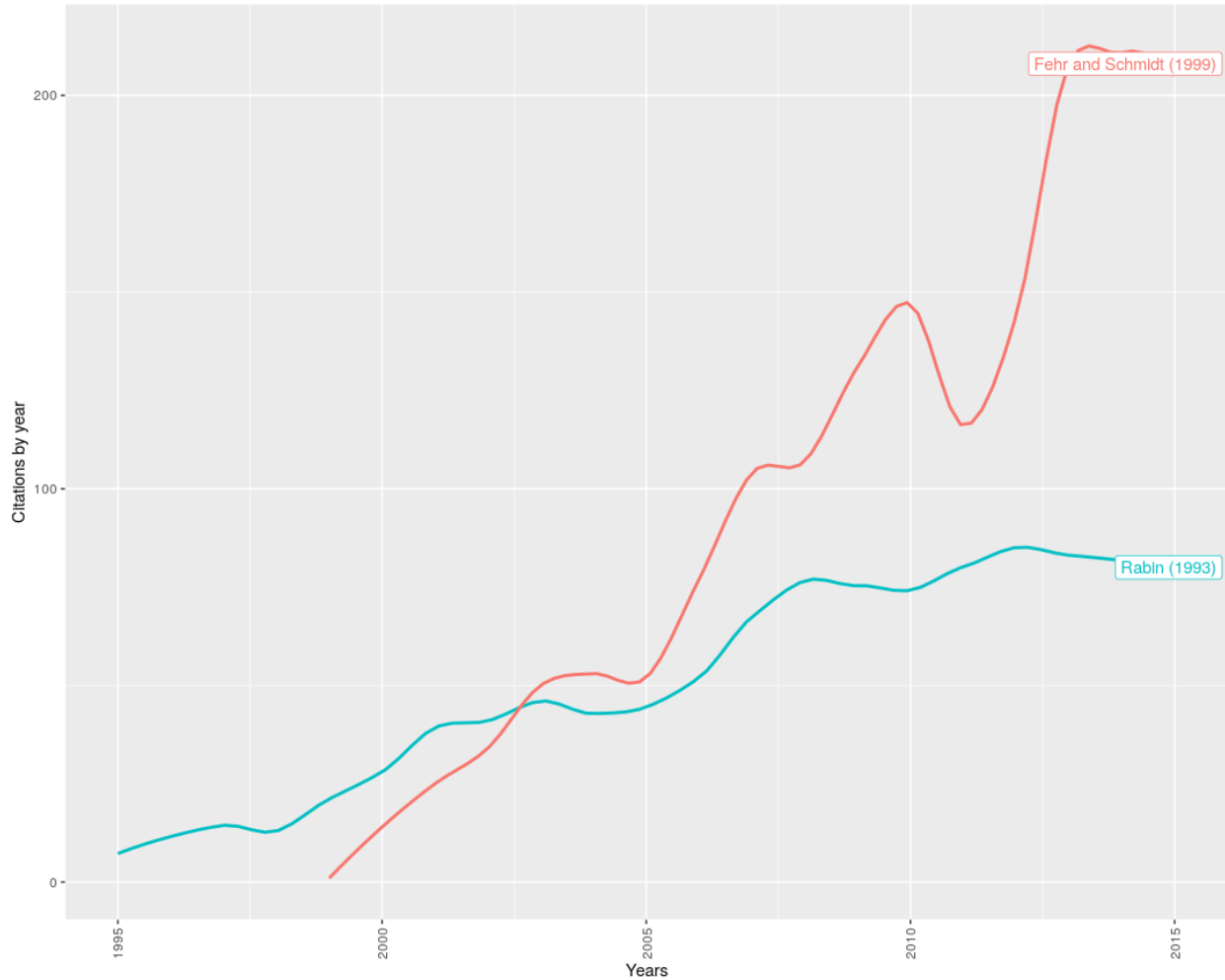


Figure 1: Citations by year of Fehr and Schmidt (1999) and Rabin (1993) (Source: Web of Science, curves are smoothed using local polynomial regression)

(and his numerous co-authors) who studied market interactions in the lab (Heukelom, 2011; Jullien and Vallois, 2014; Svorenčík, 2016). Behavioral economics developed rapidly after the 1980s when it became part of the Sloan Foundation and the Russel Sage Foundations as a research program between 1984 and 1992. Smith's experimental approach was part of the BE program within the Sloan-Sage Foundations in 1984. Between 1987 and 1988, working groups were organized to integrate the multiple lines of work developed in BE. The group working on experimental economics that was supposed to be led by Smith proved more difficult to integrate. While Smith wanted to work on market experiments and how markets

could steer individual behavior towards the equilibrium, the founding fathers of behavioral economics Kahneman and Thaler were more interested in how individuals deviated from a theoretically defined equilibrium (Heukelom, 2014, 161). The working group on experimental economics ended up being directed by Camerer instead of Smith. There were some tensions underlying this decision at the time. The separation between behavioral and experimental market economics resulted from “a persistent low-intensity conflict at the methodological and theoretical level” (Guala, 2008, 156) which had its roots in a deeper methodological division in the history of psychology (Svorenčik, 2016).

From their dominated position, behavioral and experimental market economists made efforts to become accepted into the mainstream (Heukelom, 2014, 120) to avoid being classified as outsiders (Earl and Peng, 2012) or becoming “ghettoized” (Maas and Svorenčik, 2016, 58). Furthermore, both approaches had a common interest regarding lab experiments, which they both used, and open controversies would not have helped convince mainstream economists of the value of this method. Some have qualified the approach of behavioral and experimental market economists as a Trojan horse as many strategies were put in place to be accepted as part of mainstream economics (Earl and Peng, 2012). For instance, the first publications from Kahneman and Tversky in economics were not easy to achieve, and required tinkering with regards to the rhetoric used to present their experimental results and modelling techniques (Heukelom, 2014, p.119). In the case of experiments in economics, we also find multiple strategies to make them more acceptable in the discipline (Maas and Svorenčik, 2016, p.39). While there were clear differences between both approaches, this dynamic contributed to the emergence of a semantic ambiguity about the term “experimental economics”, which sometimes includes BE and sometimes does not. This semantic ambiguity is still present today in the way Starmer describes the situation we have just presented:

“[there are] disagreements between folks who view themselves as experimental economists and others who identify themselves as behavioral economists. I see apparent lines of fracture between different camps using experimental methods.”(Maas and Svorenčik, 2016, p.177)

One episode in the mid-1990s regarding this opposition is worth mentioning briefly, as it bears directly on the topic of inter-personal games. It involves Hoffman, McCabe, and Smith (Hoffmann et al., 1995) commenting on the presentation of their work on the dictator game by Camerer and Thaler (1995b), who then responded to this comment (Camerer and Thaler, 1995a, p.239). Interestingly, while both camps in this exchange do not necessarily agree on whether the behaviors observed in this type of game pose a deep problem for standard models, they do seem to agree that their explanation is better through social norms than through social preferences (or altruism).¹⁹

The dominated position of BE within economics started to change in the late 1990s as the field grew exponentially (Geiger, 2017). This changing position coincides with the emergence of controversies about behavioral economics that became much more open and virulent. For instance, in 1999 the *Economic Journal* dedicated a special issue to discuss the controversial aspects of “experimental economics” (which included behavioral economics). It is worth noting that, in this special issue, behavioral economist Loewenstein (1999) replied to the virulent criticism made by Binmore (1999) in that same issue about the way behavioral economists run experiments. Most of Binmore’s 1999 point explicitly echoes the criticisms of behavioral economics that experimental market economists were making less openly at that time. While Loewenstein’s 1999 reply clarified the distinction between BE and experimental market economics, the semantic ambiguity of the term “experimental economics” still exists as of today.

This controversy appeared the same year as Fehr and Schmidt’s 1999 theory of inequity aversion. In a sense, one intended contribution of Fehr and Schmidt is to reconcile the opposition between behavioral economics and experimental market economics. In the very first pages of the paper, the authors explicitly state that:

“There are also many well-controlled bilateral bargaining experiments which indicate that a nonnegligible fraction of the subjects do not care *solely* about material payoffs [... Camerer and Thaler 1995].... However, there is also evidence

¹⁹The dictator game is an ultimatum game in which the receiver has no choice but to accept the offer of the proposer.

that seems to suggest that fairness considerations are rather unimportant. For example, in competitive experimental markets ... *almost all* subjects behave as if they are only interested in their material payoff ... [Smith and Williams 1990...]... In this paper we ask whether this conflicting evidence can be explained by a single simple model. Our answer to this question is affirmative” (Fehr and Schmidt, 1999, 817-8)

In the early 2000s, the work of Fehr and Schmidt (1999) participated to motivate a set of contributions on pro-social behavior involving many behavioral economists, notably Camerer, Gintis, Fehr himself, who collaborated with Henrich, an anthropologist who was the first author of the main project conducted by this group (Henrich et al., 2004). These contributions rapidly generated some controversies. In particular, Houser, McCabe, and Smith (2004, 88) questioned the very existence of non-selfish behavior and suggested that pro-social behavior could be interpreted through the notion of reciprocity and that it is in the self-interest of individuals to reciprocate. On the other hand, Binmore (2005) accused Henrich and his co-authors to attack a strawman because the “selfishness axiom” of standard economics they postulate does not exist. It is in this historical context that our controversy begins, with Shaked explicitly expressing sympathy for the work of Smith who “established and robust experimental observations of Competitive Markets” (Shaked, 2005, 2).

All these episodes allow us to appreciate how the controversy that we studied is part of a larger cyclical dynamic that reveals a broader structure of opposition between two communities within economics. This broader structure may partly be responsible for the failure of the participants to reach a consensus. Our analysis suggests not only that the controversy is related to the history of the behavioral *versus* experimental market economics divide, but also that Binmore and Shaked can be situated as siding with the market experimentalists and Fehr and Schmidt as active members of the behavioral economics side in this opposition, despite their initial reconciliatory ambition. As far as we know, these points are absent from the literature on the history of behavioral economics, which does not focus so much on developments from European scholars. It should be noted that, if these points are not

necessarily self-evident when one reads the papers constituting the controversy, it is partly because of the semantic ambiguity of the term “experimental economics”, which the authors employ interchangeably with “behavioral economics”.²⁰

4.2.2 Interdisciplinarity and the relation between theoretical and empirical work

If some sub-disciplinary boundaries play a role in our controversy, as we have seen in the previous section, the same can be argued for other inter-disciplinary boundaries (section 2).

As it is well-known, the main specificity of BE is its claim to borrow insights from psychology to model individual behavior and conduct lab experiments (Camerer and Loewenstein, 2004). Several philosophers and sociologists of science have pointed out that interdisciplinary research is, by itself, a source of controversy (Andersen, 2013; Turner et al., 2015; Allwood et al., 2020). This is particularly true in economics where most researchers explicitly favor mono-disciplinary approaches (Fourcade et al., 2015). Through training and practice, researchers acquire rules of reasoning and judgment as well as implicit norms that are specific to their discipline. Interdisciplinary research can be a source of controversies as it involves the confrontation with other such rules and norms that cannot readily be made explicit (Hesse, 1966; Dunbar, 1995, 1999; Bird, 2008).

The implied clash of epistemic values between the different scientific communities concerned by some form of interdisciplinarity often crystallizes as an opposition between data-driven *versus* theory-driven approaches (MacLeod, 2018). For example, in biophysics the interaction between physics and biology can be difficult as the two communities do not share the same interests: “physicists are theory-friendly, whereas biologists tend not to be” (Bird, 2008, 33). In the same fashion, the interaction between ecologists and economists is challenging because, for the former, the theoretical framework and the normative principles underlying economic models are hard to rationalize (MacLeod, 2018, 13). The special issue of the JEBO in which the last exchanges of our controversy were published closes a similar

²⁰The set of open and virulent controversies about behavioral economics after the mid-2000 is broader than the one we have presented here and includes notable Camerer (2008) *versus* Gul and Pesendorfer (2008).

point on how theories are evaluated differently in economics, where parsimony prevails at the cost of abstraction, and in psychology, where the accuracy of predictions for many specific situations prevails at the cost of a high degree of freedom, while behavioral economics stands in-between (Croson and Gächter, 2010).²¹

This tension underlying behavioral economics is indeed clearly present in the controversy that we studied, as Fehr and Schmidt describe their methodological position as follows:

[...] there is no guarantee that the outcome of this endeavor will provide us with an empirically informed model that is tractable enough for applied economics. Thus, we may be faced with a situation where we have to sacrifice precision in the modeling of individual preferences in order to keep the models tractable for applied work. (Fehr and Schmidt, 2010, 107)

For Fehr and Schmidt, inequity aversion is a model to stimulate empirical investigations that can summarize much of the knowledge they derived about social preferences from experiments. They recognize that group welfare, norms, reciprocity, and reputations are as important as preferences for altruism and envy in understanding pro-social behavior. However, their theoretical results are only relevant to them to interpret the results of experiments as the interplay between selfish behavior and other-regarding behavior, without being specific about the precise types of social preferences. This commitment to favor data-driven research did not lead Fehr and Schmidt to develop their 1999 model in further theoretical work. This strategy proved to be useful to engage in empirically-oriented interdisciplinary collaboration on social preferences with anthropologists, psychologists, and neuroscientists (Glimcher et al., 2009; Kosfeld et al., 2005; Gintis et al., 2008). After 2010, more than fifty percent of Fehr's work is published outside of economics journals, while still being about social preferences (Truc, 2022a).

By contributing to the behavioral sciences more generally, behavioral economists are blurring the frontiers between disciplines and criteria of evaluation. This approach is also

²¹See also (Croson, 2006) for a similar point of view.

advocated by another participant in our controversy, Gintis, who explicitly criticizes Binmore on this point:

[W]e are more generally behavioral scientists, not just economists, and we have published in first ranked professional journals in biology, anthropology, psychology, sociology, and political science. Our results do not “belong” to economics, and indeed can be evaluated profitably by psychologists, sociologists, and indeed any professional acquainted with the scientific method. The papers published in *Science* and *Nature* are extensions of research published widely in all the behavioral sciences. (Gintis, 2011, 101)

This state of affairs, is itself a point of controversy, as Binmore seems to regard this strategy as a means to avoid criticisms from peers in economics:

“[A] vocal minority of behavioralists have succeeded in outflanking the skepticism of mainstream economists by appealing directly to a wide range of social and life scientists with the aid of publications in journals like *Science* and *Nature* that do not normally publish the work of economists. They have thereby been successful in discrediting psycho-sociological explanations of the data (in terms of cultural norms) in favor of traditional economic explanations (in terms of the optimization of utility functions) while simultaneously decrying the very economic methodology that they seek to persuade others to adopt” (Binmore, 2010, 148).

The new sources of legitimacy from the natural sciences allow behavioral economists to dismiss the concerns of standard economists for the historically dominant theoretical questions, or for any question related to the boundaries of economics (Gul and Pesendorfer, 2008; Gintis, 2011). The importance of theoretical work in economics is challenged by the interaction with other disciplines that do not operate with such a strict distinction between empirical and theoretical work, and that have historically valued empirical contributions

more highly than theoretical ones. By switching the role of theoretical models as a unificationist medium to a tool at the service of empirical investigation, the approach of Fehr and Schmidt does indeed present major differences with the more traditional theoretically oriented economists. Such differences were indeed apparent in the controversy, when (Binmore and Shaked, 2010a, 97) accused, Fehr and Schmidt of considering “that the subjects behave ‘as if’ motivated by inequity aversion and that their theory ‘helps to organize and interpret the data’”. In other words, using the model mainly to *interpret* experimental is problematic for a theoretician such as Binmore.²²

The ambiguity of the status of theoretical models in BE is indeed non-trivially related to the issue of the realism of hypotheses. On the one hand, behavioral economists identify the as-if methodology as one of the historical moments symbolizing when economics went wrong about how to model individual behavior (Thaler, 1980; Camerer, 2006; Thaler, 2015). On the other hand, more recently, some behavioral economists like Chetty (2015, 29) explicitly defend that theories should be evaluated on their applied value and adopt “an updated reading of the ‘as if’ approach”.²³

To sum up, the study of our controversy reveals that some of its underlying points of contentions are due to the interdisciplinary ambitions of behavioral economics. In a sense, the oppositions of the participants can be taken as manifestations of the one between subtle versions of rationalism (the primacy of theoretical work in the constitution of knowledge) from the theoreticians Binmore and Shaked and empiricism (the primacy of experience in the constitution of knowledge) from behavioral economists, which is not uncommon in scientific controversies as suggested in Section 2. Our point is not that Binmore and Shaked are only interested in theoretical work (they are experimentalists as well; see Binmore et al., 1985). Our point is rather that both sides of the controversy exhibit a different articulation

²²Similar arguments about the as-if approach in behavioral economics can be found in Berg and Gigerenzer (2010); Moscati (2021)

²³This tension is exacerbated by at least two other historical and methodological factors. Firstly, BE is part of the more general shift of the economics profession towards more empirical and applied work (Hamermesh, 2013; Backhouse and Cherrier, 2017a; Angrist et al., 2020). Secondly, there is an unacknowledged heterogeneity of interpretations of rational choice theory and game theory from both proponents and opponents of behavioral economics engaged in controversies (Guala, 2006; Herfeld, 2020).

and hierarchy between theory and data, which lead to significantly different approaches to experiments in economics (see Backhouse and Cherrier, 2017a; Backhouse and Cherrier, 2017b for a more general perspective on the articulation between theory and data within economics in the last few decades).

5 Conclusion

The goal of this paper was to analyze the controversy on inequity aversion that opposed Shaked and Binmore against Fehr, Schmidt, Eckel, and Gintis in the light of the tools we can find in sub-disciplines that have specialized in the study of controversies for several decades. Our analyzes reveal several insights that contribute either to the historical and methodological literature on behavioral economics or to the literature on the study of controversies. We showed that the controversy was about more than what the participants said explicitly, as it was part of a larger cycle of related controversies largely unacknowledged. The controversy happened to be an object that crystallizes a set of tensions regarding the interdisciplinary boundaries of economics, even though this issue was not mentioned by the participants. Finally, the controversy also proved to be a relevant object to reveal or to characterize some subtle mechanisms that participate in the way controversies achieve closure or remain open. In our case, the strong rhetorical tone, the dynamical expansion of the topic, and the implicit epistemological positions of the participants all contributed to leave the controversy unresolved. As we saw in our analysis of the reception of the controversy (section 4.1.3), while its existence is sometimes acknowledged, most behavioral economists do not seem to take sides explicitly (Harrison (2011) can be read as implicitly siding with Shaked and Binmore). The openness of the controversy suggests that its consequences will be internalized (i.e., a change (or not) in empirical modelling practices in behavioral economics that will not be acknowledged explicitly) or that it will re-open in the future as a new cycle.

6 Appendix

6.1 Questions asked to the scholars involved in the controversy

Our analysis of the controversy is partly based on our personal correspondence with the participants and Barkley Rosser. We contacted these scholars in January 2020 and asked them the following questions:

To Avner Shaked:

- When did you decided to make and distribute the first paper about the rhetoric of IA? What was the intention behind this first paper?
- How did the collaboration between you and Binmore regarding this controversy started?
- How did the controversy moved from relatively informal papers to the Journal of Economic Behavior & Organization, and how was it decided/organized on your side?
- Was the tone of the papers discussed with your co-author? Or with other economists?
- How do you feel about how the controversy ended? What do you feel are the consequences of this controversy up until now?

To Kenneth Binmore:

- How did the collaboration between you and Shaked started and why did you decided to join the controversy after the first paper by Shaked?
- How did you feel about the first exchange between Shaked and FS? What did you want to change when writing the second paper within the JEBO?
- Was the tone of the papers discussed with your co-author? Or with other economists?
- How do you feel about how the controversy ended? What do you feel are the consequences of this controversy up until now?

To Klaus Schmidt (responded jointly with Ernst Fehr):

- How important do you think is this controversy in your career as an academic and for the work you produced after?
- When did you decide to answer to the first article of Shaked? And why?
- How did the controversy moved from relatively informal papers to Journal of Economic Behavior & Organization, and how was it decided/organized?
- Following that question: was the tone of the response discussed with your co-author? Or with other economists?
- How do you feel about how the controversy ended? What do you feel are the consequences of this controversy up until now?

To Rosser:

- When did you first heard about the controversy and how was it decided on your side to involve the JEBO?
- How did you feel about the controversy before the JEBO became involved? And after?
- How do you feel about how the controversy ended?

References

- Allwood, J., Pombo, O., Renna, C., and Scarafile, G., editors (2020). *Controversies and Interdisciplinarity: Beyond disciplinary fragmentation for a new knowledge model*, volume 16 of *Controversies*. John Benjamins Publishing Company, Amsterdam.
- Andersen, H. (2013). The Second Essential Tension: on Tradition and Innovation in Interdisciplinary Research. *Topoi*, 32(1):3–8.
- Angner, E. (2019). We’re All Behavioral Economists Now. *Journal of Economic Methodology*, 26(3):195–207.
- Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2020). Inside Job or Deep Impact? Extramural Citations and the Influence of Economic Scholarship. *Journal of Economic Literature*, 58(1):3–52.

- Backhouse, R. E. and Cherrier, B. (2017a). The Age of the Applied Economist. *History of Political Economy*, 49(Supplement):1–33.
- Backhouse, R. E. and Cherrier, B. (2017b). “It’s Computers, Stupid!” The Spread of Computers and the Changing Roles of Theoretical and Applied Economics. *History of Political Economy*, 49(Supplement):103–126.
- Baltas, A. (2000). Classifying Scientific Controversies. In Machamer, P. K., Pera, M., and Baltas, A., editors, *Scientific Controversies: Philosophical and Historical Perspectives*, pages 40–49. Oxford University Press, Oxford.
- Bardsley, N., editor (2010). *Experimental economics: rethinking the rules*. Princeton University Press, Princeton.
- Berg, N. and Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, 18(1):133–165.
- Binmore, K. (1999). Why Experiment in Economics? *The Economic Journal*, 109(453):F16–F24.
- Binmore, K. (2005). Economic man – or straw man? *Behavioral and Brain Sciences*, 28(06).
- Binmore, K. (2010). Social norms or social preferences? *Mind & Society*, 9(2):139–157.
- Binmore, K. (2011). Review. *Journal of Economic Methodology*, 18(2):177–181.
- Binmore, K. and Shaked, A. (2010a). Experimental economics: Where next? *Journal of Economic Behavior & Organization*, 73(1):87–100.
- Binmore, K. and Shaked, A. (2010b). Experimental Economics: Where Next? Rejoinder. *Journal of Economic Behavior & Organization*, 73(1):120–121.
- Binmore, K., Shaked, A., and Sutton, J. (1985). Testing noncooperative bargaining theory: A preliminary study. *The American Economic Review*, 75(5):1178–1180. tex.publisher: American Economic Association.
- Bird, A. (2008). Incommensurability Naturalized. In Soler, L., Sankey, H., and Hoyningen-Huene, P., editors, *Rethinking Scientific Change and Theory Comparison*, pages 21–39. Springer Netherlands, Dordrecht.
- Braesemann, F. (2019). How behavioural economics relates to psychology – some bibliographic evidence. *Journal of Economic Methodology*, 26(2):133–146.
- Bruhin, A., Fehr-duda, H., and Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4):1375–1412.
- Camerer, C. (2006). Behavioral Economics. In Blundell, R., Newey, W., and Persson, T., editors, *World Congress of the Econometric Society*, pages 181–214. Cambridge University Press, London.
- Camerer, C. and Loewenstein, G. (2004). Behavioral economics: Past, present, future. In Camerer, C. F., Loewenstein, G., and Rabin, M., editors, *The Foundations of Positive and Normative Economics: A handbook*, pages 3–51. Princeton University Press, Princeton.
- Camerer, C. and Thaler, R. (1995a). Response from Colin Camerer and Richard H. Thaler. *The Journal of Economic Perspectives*, 9(4):239–240. Publisher: American Economic Association.
- Camerer, C. and Thaler, R. H. (1995b). Anomalies: Ultimatums, Dictators and Manners. *Journal of Economic Perspectives*, 9(2):209–219.

- Camerer, C. F. (2008). The Case for Mindful Economics. In Caplin, A. and Schotter, A., editors, *The Foundations of Positive and Normative Economics: A handbook*, pages 43–69. Oxford University Press, Oxford.
- Chang, H. (2014). *Is water H₂O? evidence, realism and pluralism*. Number 293 in Boston studies in the philosophy of science. Springer, Dordrecht Heidelberg New York London.
- Chang, H. (2017). Is pluralism compatible with scientific realism? In Saatsi, J., editor, *The routledge handbook of scientific realism*, pages 176–186. New York: Routledge.
- Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review*, 105(5):1–33.
- Collins, H. (1981a). Stages in the Empirical Programme of Relativism. *Social Studies of Science*, 11(1):3–10.
- Collins, H. and Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science*, 32(2):235–296.
- Collins, H. M. (1981b). The Place of the ‘Core-Set’ in Modern Science: Social Contingency with Methodological Propriety in Science. *History of Science*, 19(1):6–19.
- Croson, R. (2006). Contrasting methods and comparative findings in psychology and economics. In De Cremer, D., Zeelenberg, M., and Murnighan, J. K., editors, *Social psychology and economics*, pages 301–317. Psychology Press, New York. OCLC: 841171684.
- Croson, R. and Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73(1):122–131.
- Dascal, M. (1995). Observations sur la dynamique des controverses. *Cahiers de linguistique française*, 17:99–121.
- Dascal, M. (1998). The Study of Controversies and the Theory and History of Science. *Science in Context*, 11(2):147–154.
- Davis, J. B. (2010). *Individuals and Identity in Economics*. Cambridge University Press, Cambridge.
- Debaz, J. (2017). Entre science et société, les controverses comme enquêtes collectives. *Zilsel*, 2(2):149.
- Dhermy-Mairal, M. and Fages, V. (2017). Une étude de controverses à nouveaux frais ? Des normes pour comprendre le champ scientifique. *Zilsel*, 2(2):167.
- Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4):559–579. Publisher: [The University of Chicago Press, Philosophy of Science Association].
- Dreber, A., Fudenberg, D., and Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, 98:41–55.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In *The nature of insight.*, pages 365–395. The MIT Press, Cambridge, MA, US.
- Dunbar, K. (1999). How Scientists Build Models In Vivo Science as a Window on the Scientific Mind. In Magnani, L., Nersessian, N. J., and Thagard, P., editors, *Model-Based Reasoning in Scientific Discovery*, pages 85–99. Springer US, Boston, MA.
- Earl, P. E. and Peng, T.-C. (2012). Brands of Economics and the Trojan Horse of Pluralism. *Review of Political Economy*, 24(3):451–467.

- Eckel, C. and Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior & Organization*, 73(1):109–119.
- Fabiani, J.-L. (1997). Controverses scientifiques, controverses philosophiques: Figures, positions, trajets. *Enquête*, (5):11–34.
- Fehr, E. and Schmidt, K. (2005). The Rhetoric of Inequity Aversion- A Reply. Working.
- Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fehr, E. and Schmidt, K. M. (2010). On inequity aversion: A reply to Binmore and Shaked. *Journal of Economic Behavior & Organization*, 73(1):101–108.
- Fischbacher, U., Fong, C. M., and Fehr, E. (2003). Fairness, Errors and the Power of Competition.
- Fourcade, M., Ollion, E., and Algan, Y. (2015). The Superiority of Economists. *Journal of Economic Perspectives*, 29(1):89–114.
- Freudenthal, G. (1998). “Controversy”. *Science in Context*, 11(2):155–160.
- Fritz, G. (2019). Scientific controversies. In Leßmöllmann, A., Dascal, M., and Glöning, T., editors, *Science Communication*, pages 311–334. De Gruyter.
- Geiger, N. (2017). The Rise of Behavioral Economics: A Quantitative Assessment. *Social Science History*, 41(03):555–583.
- Gintis, H. (2011). The future of behavioral game theory. *Mind & Society*, 10(2):97–102.
- Gintis, H., Henrich, J., Bowles, S., Boyd, R., and Fehr, E. (2008). Strong Reciprocity and the Roots of Human Morality. *Social Justice Research*, 21(2):241–253.
- Glimcher, P. W., Camerer, C., Fehr, E., and Poldrack, R. A., editors (2009). *Neuroeconomics: decision making and the brain*. Elsevier, Amsterdam Boston Heidelberg London New York Oxford Paris San Diego San Francisco Singapore Sydney Tokyo, first edition edition. OCLC: 845386625.
- Graham, S. S. and Walsh, L. (2019). There’s No Such Thing as a Scientific Controversy. *Technical Communication Quarterly*, 28(3):192–206.
- Gravoglu, K. (2000). Controversies and the Becoming of Physical Chemistry. In Machamer, P. K., Pera, M., and Baltas, A., editors, *Scientific Controversies: Philosophical and Historical Perspectives*, pages 177–198. Oxford University Press, Oxford.
- Grüne-Yanoff, T. (2016). Interdisciplinary success without integration. *European Journal for Philosophy of Science*, 6(3):343–360.
- Guala, F. (2006). Has Game Theory Been Refuted? *The Journal of Philosophy*, 103(5):239–263.
- Guala, F. (2008). Experimental Economics, History of. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics: Volume 1 – 8*, pages 1958–1962. Palgrave Macmillan UK, London.
- Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1):1–15.
- Guala, F. and Mittone, L. (2010). Paradigmatic experiments: The Dictator Game. *The Journal of Socio-Economics*, 39(5):578–584.
- Gul, F. and Pesendorfer, W. (2008). The case for mindless economics. In Caplin, A. and

- Schotter, A., editors, *The Foundations of Positive and Normative Economics: A handbook*, pages 3–42. Oxford University Press, Oxford.
- Hacking, I. (2000). Multiple Personalities, Internal Controversies, and Invisible Marvels. In Machamer, P. K., Pera, M., and Baltas, A., editors, *Scientific Controversies: Philosophical and Historical Perspectives*, pages 213–229. Oxford University Press, Oxford.
- Hamermesh, D. S. (2013). Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature*, 51(1):162–172.
- Hands, D. W. (2010). Economics, psychology and the history of consumer choice theory. *Cambridge Journal of Economics*, 34(4):633–648.
- Harrison, G. W. (2011). The methodological promise of experimental economics. *Journal of Economic Methodology*, 18(2):183–187.
- Heidl, S. (2016). *Philosophical problems of behavioural economics*. Routledge INEM advances in economic methodology. Routledge, Taylor & Francis Group, London ; New York.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (2004). *Foundations of Human Sociality*. Oxford University Press.
- Herfeld, C. (2020). The Diversity of Rational Choice Theory: A Review Note. *Topoi*, 39(2):329–347.
- Hesse, M. B. (1966). *Models and analogies in science*. University of Notre Dame Press, Notre Dame, Ind. OCLC: 984798137.
- Heukelom, F. (2011). What to Conclude from Psychological Experiments: The Contrasting Cases of Experimental and Behavioral Economics. *History of Political Economy*, 43(4):649–681.
- Heukelom, F. (2014). *Behavioral Economics: A History*. Cambridge University Press, New York.
- Hoffmann, W., McCabe, K., and Smith, V. (1995). Ultimatum and Dictator Games. *The Journal of Economic Perspectives*, 9(4):236–239. Publisher: American Economic Association.
- Houser, D., McCabe, K., and Smith, V. (2004). Cultural group selection, co evolutionary processes and large-scale cooperation (by Joseph Henrich). *Journal of Economic Behavior & Organization*, 53(1):85–88.
- Jullien, D. and Vallois, N. (2014). A probabilistic ghost in the experimental machine. *Journal of Economic Methodology*, 21(3):232–250.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263.
- Kitcher, P. (2000). Patterns of Scientific Controversies. In Machamer, P. K., Pera, M., and Baltas, A., editors, *Scientific Controversies: Philosophical and Historical Perspectives*, pages 21–39. Oxford University Press, Oxford.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435(7042):673–676.
- Kovaka, K. (2021). Fighting about frequency. *Synthese*, 199(3-4):7777–7797.
- Lakatos, I. (1978). *The methodology of scientific research programmes*. Number 1 in Philosophical papers. Cambridge University Press, Cambridge.

- Lemieux, C. (2007). À quoi sert l'analyse des controverses ? *Mil neuf cent*, n° 25(1):191.
- Lisciandra, C. (2018). The role of psychology in behavioral economics: The case of social preferences. *Studies in History and Philosophy of Science Part A*.
- Loewenstein, G. (1999). Experimental Economics From the Vantage-point of Behavioural Economics. *The Economic Journal*, 109(453):25–34.
- Longino, H. E. (2013). *Studying human behavior: how scientists investigate aggression and sexuality*. The University of Chicago press, Chicago.
- Maas, H. and Svorenčík, A., editors (2016). *The Making of Experimental Economics: Witness Seminar on the Emergence of a Field*. Springer International Publishing : Imprint: Springer, Cham, 1st ed. 2016 edition.
- Machamer, P. K., Pera, M., and Baltas, A. (2000). *Scientific Controversies: Philosophical and Historical Perspectives*. Oxford University Press, Oxford.
- MacLeod, M. (2018). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese*, 195(2):697–720.
- Małecka, M. (2021). Values in economics: a recent revival with a twist. *Journal of Economic Methodology*, 28(1):88–97.
- Mongin, P. (2006). A concept of progress for normative economics. *Economics and Philosophy*, 22(1):19–54.
- Moscatti, I. (2021). Process Models Are As-If Models: An Antirealist Account of Economic Theories of Decision-Making. Working Paper, Available at SSRN: <https://ssrn.com/abstract=3724961>.
- Nagatsu, M. (2021). *A modern guide to philosophy of economics*. Elgar modern guides. Edward Elgar Publishing, Cheltenham.
- Paternotte, C. and Grose, J. (2013). Social Norms and Game Theory: Harmony or Discord? *The British Journal for the Philosophy of Science*, 64(3):551–587.
- Pestre, D. (2007). L'analyse de controverses dans l'étude des sciences depuis trente ans: Entre outil méthodologique, garantie de neutralité axiologique et politique. *Mil neuf cent*, n° 25(1):29.
- Pinch, T. (2015). Scientific controversies. In Wright, J. D., editor, *International Encyclopedia of the Social & Behavioral Sciences*, pages 281–6. Elsevier, Oxford, 2nd edition.
- Popper, K. R. (1934). *The Logic of scientific discovery*. Routledge classics. Routledge, London, repr. 2008 (twice) edition. OCLC: 315976473.
- Raynaud, D. (2015). *Scientific Controversies: a Socio-Historical Perspective on the Advancement of Science*. Routledge. OCLC: 1001887961.
- Sent, E.-M. (2004). Behavioral Economics: How Psychology Made Its (Limited) Way Back Into Economics. *History of Political Economy*, 36(4):735–760.
- Shaked, A. (2005). The Rhetoric of Inequity Aversion. Technical report, University of Bonn.
- Svorenčík, A. (2016). The Sidney Siegel Tradition: The Divergence of Behavioral and Experimental Economics at the End of the 1980s. *History of Political Economy*, 48(suppl 1):270–294.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1):39–60.

- Thaler, R. H. (2015). *Misbehaving: the making of behavioral economics*. W.W. Norton & Company, New York, first edition edition.
- Truc, A. (2022a). The Disciplinary Mobility of Core Behavioral Economists. *Working Paper*.
- Truc, A. (2022b). Interdisciplinary influences in behavioral economics: a bibliometric analysis of cross-disciplinary citations. *Journal of Economic Methodology*, 29(3):217–251.
- Turner, V. K., Benessaiah, K., Warren, S., and Iwaniec, D. (2015). Essential tensions in interdisciplinary scholarship: navigating challenges in affect, epistemologies, and structure in environment–society research centers. *Higher Education*, 70(4):649–665.
- Vromen, J. (2022). As If Social Preference Models. In Heilmann, C. and Reiss, J., editors, *The Routledge handbook of the philosophy of economics*, Routledge handbooks in philosophy, pages 125–137. Routledge, Taylor & Francis Group, New York London, first published edition.
- Yang, Y., Onderstal, S., and Schram, A. (2016). Inequity aversion revisited. *Journal of Economic Psychology*, 54:1–16.