Individuals and Institutions
Essays on Social and Political Epistemic Systems

by

Yao Fan

BA, University of Sydney, 2015

BA(HONS), Australian National University, 2016

MA, University of Pittsburgh, 2023

MSc, University of Pittsburgh, 2023

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yao Fan

It was defended on

July 21, 2023

and approved by

Anil Gupta, Alan Ross Anderson Distinguished Professor, Department of Philosophy

Kevin Dorst, Assistant Professor, Department of Linguistics and Philosophy (Massachusetts Institute of Technology)

Dissertation Co-Director: James Shaw, Associate Professor, Department of Philosophy

Dissertation Co-Director: Kevin Zollman, Professor, Department of Philosophy (Carnegie Mellon University)

Individuals and Institutions
Essays on Social and Political Epistemic Systems

Yao Fan, BA(HONS), MA, MSc, PhD

University of Pittsburgh, 2023

This thesis is a collection of three papers in social epistemology. Three epistemic systems of great importance for modern society are studied respectively: the expert system, the democratic system, and the financial system. Throughout the three papers, I try to pay additional attention to the interactions between individuals and institutions, which explains the title of this dissertation.

In the first paper, I discuss ways in which institutional remedies are crucial for individual epistemic performance. In the second paper, I argue that the individual moral requirements of citizens (i.e., participants of democracy) should not be overlooked in presence of the epistemic aggregation effect of democracy at the institutional level. In the third paper, I analyze how individual motives in financial markets and the epistemic functions of financial markets influence and interact with each other.

The proper functioning of each of the three institutions, I believe, is necessary for achieving a good society in our time. In a good society, neither the individual aspects nor the institutional aspects should be overlooked.

Table of Contents

Preface viii

Acknowledgement xiv

1.0 Institutional Remedies for Individual Expert Identification 1

1.1 Introduction 1

1.2 The Models 5

1.2.1 Strategic Testing: a toy model 6

1.2.2 Strategic Testing: choosing difficulty levels 13

1.2.3 Strategic Testing and Strategic Disclosing 17

1.3 Philosophical Reflections: Epistemic Virtues at Institutional Level 23

1.4 Appendix 26

1.4.1 Appendix 1. The First Model 27

1.4.1.1 A1.1 Numerical Examples 29

1.4.1.2 A1.2 Accuracy 32

1.4.2 Appendix 2. The Second Model 34

1.4.3 Appendix 3. The Third Model 38

2.0 Voting Norms in Condorcet Jury Theorem: What We Owe to Other Voters 41

2.1 Introduction 41

2.2 The Classic CJT Framework 43

2.3 Two Varieties of Strategic Voting 47

2.3.1 Second-Best Voting 49

2.3.2 Pivotal Voting 52

2.4 Economists' Case for Pivotal Voting in the CJT 56

2.5 What is Wrong of Pivotal Voting 60

2.5.1 Previous arguments against pivotal voting 60

2.5.2 The relevance of a normative argument 62

2.5.3 A moral argument against pivotal voting 65

2.5.3.1 5.3.1 The Nature of Pivotal Voting 65

2.5.3.2 5.3.2 The Moral Flaw of Pivotal Voting 66

2.6 The Presence of Heterogeneous Values 70

2.7 Final Remark 74

3.0 The Ideal of Information Efficiency of Financial Markets 76

3.1 The Information Efficiency of Financial Markets 78

3.1.1 Information-Arbitrage Efficiency 80

3.1.2 Fundamental Valuation Efficiency 82

3.2 The Ideal of Fundamental Valuation Efficiency 84

3.2.1 The Definition of Fundamental Value 85

3.2.2 Two Examples 87

3.2.2.1 2.2.1 Bond 88

3.2.2.2 2.2.1 Stock 92

3.2.3 Fundamental and Non-Fundamental Information 96

3.3 Anomalies: Challenges in Realizing the Ideal 97

3.3.1 Anomalies to FV-efficiency 97

3.3.2 The Clever-Ticker Effect 99

3.3.3 Market Mechanism for FV-Efficiency 103

3.3.4 Regulation for FV-Efficiency: a note from social epistemology 106

3.4 Is the Ideal Adequate? Challenges from ESG Investing..... 109

3.4.1 What Motivates ESG Investing? 110

3.4.2 New Ideal for New Paradigm.....112

3.4.3 Are Financial Markets suitable for processing ESG information? 114

3.5 What's Next..... 120

Bibliography 122

List of Tables

Table 1: An infallible test. 9

Table 2: A general test. 10

Table 3: A comparison between $x\dagger$ and $x*$ 15

Table 4: A comparison of various tests. 32

Table 5: Examples of Clever Tickers. 100

List of Figures

Figure 1: Left: x and P_x^N; Right: x and the novice's mean squared error. 21

Figure 2: x and novice's decision inaccuracy 35

Figure 3: x and the probability of $Sx = 1$ 35

Figure 4: x and the novice's decision accuracy, with $\sigma = 1$ and $t = 2/3$ 37

Figure 5: x and the novice's decision accuracy, with $\sigma = 1$ and $t = 1/2$ 37

Preface

This thesis is a collection of three papers in social epistemology. Three epistemic systems of great importance for modern society are studied respectively: the expert system, the democratic system, and the financial system. For reasons I will discuss later, all three papers contain technical details that readers may find confusing or overwhelming to some extent. One may lose track of what is the purpose of all these and why would any of them show up in a philosophy dissertation. I hope that this preface can provide some motivation and justification for what I have done in these three papers.

The focus of epistemology has traditionally been on individuals. Descartes' epistemology concerns how certainty can be achieved from first principles through meditation, Locke's epistemology tries to understand how (individual) human understanding operates, and Kant's epistemology was an attempt at the conditions for the possibility of human knowledge from an individual perspective. More recently, philosophers have become interested in how rational individuals should respond to evidence, form rational beliefs, and obtain knowledge. Even the recent literature on social epistemology is often individualistic, focusing on how individuals should react to social evidence, e.g., testimonies, in pursuit of their individual epistemic goals. In all these cases, the subjects under investigation are individuals.

The past studies of individualistic epistemology have proven productive for a better understanding of individual rationality but fail to capture the social aspect of epistemic activities in the modern world. Nowadays, epistemic tasks of importance are rarely carried out by individual epistemic agents. Instead, they are pursued by collective agents, a group, a team, a community, etc., often in a systematic and institutionalized manner.

An early influential study of collective epistemic agents is Edwin Hutchins's *Cognition in the Wild* (1995), which provides a fantastic case study about how ship navigation, a surprisingly complicated and challenging epistemic activity, is done in a systematic and distributed manner by crew members. Another great example is modern science, which is highly collective and institutionalized. Scientific discoveries are not achieved by an individual scientist who is on top of everything, but by a group of scientists each taking their own share of cognitive labor. Scientists are trained for years to follow the protocols for scientific discoveries, from research methodology to how to publish in peer-reviewed journals.

In recent years, considerable philosophical literature has developed around the institution of science, from how to best divide cognitive labors to how to reduce fraud. Despite their enthusiasm for the epistemic institution of science, philosophers have been much less prolific on other epistemic institutions. In this dissertation, I reflect on three epistemic systems that are much less studied by philosophers, namely, the expert system, the democratic system, and the financial system. While their social importance is self-evident, I hope that my readers will also find them of philosophical interest.

Economists have studied social institutions extensively in the past decades. It would be wrong to neglect their contributions. In all three papers, I draw from and reflect on existing economic studies of epistemic institutions. Certain background knowledge in economics, which falls out of philosophers' daily toolbox, is necessary for such reflections. I attempt to provide such backgrounds in these papers. For this reason, an exposition of technical backgrounds precedes philosophical reflections in each of the three papers. Readers from a more technical background may find the exposition redundant, while readers from a more philosophical background may find the reflection inadequate. It's after all difficult to swim in two rivers at the same time. I often

remind myself of the drunk who looks for his lost keys under the lamppost rather than where he actually dropped them. While it's certainly easier to look for things in places where the light is better, it's more important to shine light where it's needed the most.

In the rest of this preface, I give an overview of the three papers.

The first paper is the most technical one among the three. It connects the recent development of Bayesian persuasion models in economics with the philosophical problem of expert identification. Existing philosophical discussion on expert identification is mostly individualistic. The main question there is how an individual can find the right expert to trust on her own. My paper emphasizes the institutional aspect of the expert system. Certain barriers a novice faces in identifying experts can only be overcome at the institutional level. Due to the asymmetry in signaling power between the expert and the novice, even a perfectly rational Bayesian novice cannot avoid being misled by the expert on her own.

The second paper is perhaps the most philosophical one. It addresses an apparent disagreement on voting norms between philosophers and economists under the setup of the Condorcet Jury Theorem (CJT). While economists have moved away from sincere voting since Austen-Smith & Banks (1996), philosophers, represented by Goodin & Spiekermann (2018), are relentlessly defending it. I provide an analysis of this disagreement, along with a defense of sincere voting on moral grounds. It's surprising that such a moral argument has never been explicitly laid out in the existing philosophical literature on voting.

The third paper studies financial markets as an epistemic institution. Despite, or perhaps because of, the existence of a community of experts, namely finance researchers and scholars, who study financial systems for a living, financial systems are rarely put under philosophical scrutiny. While some eminent finance scholars discuss philosophical issues in their less technical writings

(e.g., keynote addresses or policy commentaries), their discussions could often use some refinement from the perspective of a philosopher. In this paper, I try to articulate, analyze, and criticize the ideal of information efficiency of financial markets. This ideal of financial markets qua epistemic institutions is widely appraised by finance scholars and practitioners alike but seldom gets carefully reflected upon.

Throughout the three papers, I try to pay additional attention to the interactions between individuals and institutions, which explains the title of this thesis. In the first paper, I discuss ways in which institutional remedies are crucial for individual epistemic performance. In the second paper, I argue that the individual moral requirements of citizens (i.e., participants of democracy) should not be overlooked in presence of the epistemic aggregation effect of democracy at the institutional level. In the third paper, I analyze how individual motives in financial markets and the epistemic functions of financial markets influence and interact with each other. The proper functioning of any of these three institutions, I believe, is necessary for achieving a good society in our time. In a good society, neither the individual aspects nor the institutional aspects should be overlooked.

Acknowledgement

The submission of this dissertation marks a temporal stop to my journey in professional philosophy. This acknowledgement gives me a good opportunity to look back at the journey and express my gratitude to so many wonderful people I met along the way.

The journey started in early 2016, one and a half years before my arrival at Pitt, when I met Daniel Stoljar for the first time in his office at Australian National University. Before ANU, I primarily studied mathematics and logic. When I arrived at ANU to start my Honours year in philosophy (which is a research-oriented fourth year of undergraduate study, unique of Australian higher education system), my background in philosophy was quite limited.

Despite my limited background, the philosophy department at ANU successfully transformed me into a researcher of philosophy. I wrote an Honours thesis and three papers during my Honours year under the supervision of Daniel Stoljar, Alan Hajek, Koji Tanaka and Phil Dowe. While I learnt much from each one of my supervisors there as well as other members and visitors of the ANU philosophy community, I'm particularly grateful to Daniel, from whom I learnt how to do analytic philosophy. My ability to write philosophy benefited greatly from Al Hajek, with whom I iterated many versions of a paper, which ended up getting me admitted to Pitt.

Getting admitted to Pitt was indeed a surprise. No one at ANU at that time knew much about what's going on at Pitt. Later, I learnt that my writing sample, written with much help from Al, somehow impressed Dmitri Gallow, a junior faculty at Pitt Philosophy who later moved to Australia. I made my decision to go to Pitt over other schools largely because of the logic and formal epistemology community in the Pittsburgh area, formed by Pitt and CMU philosophers.

While I did learn some logic and formal epistemology during my time at Pitt (and even wrote my dissertation on it), I learnt a lot more what I didn't expect back then.

I took a lot of classes (145 credit units, or 48.33 courses accumulated over six years) and learnt from a lot of people during my time in Pittsburgh. It's impossible to give a comprehensive list of people who I have learnt from. I can only name a few occasions that happen to be most influential for me.

In the Spring of my first year, I took Teddy Seidenfeld's seminar on Foundations of Statistics at CMU. That seminar was truly enlightening for me. There I saw what good mathematical/technical-informed philosophy looks like for the first time. I enthusiastically put everything I can grasp from that seminar, which is far from satisfying, due to my limited background in Statistics back then, into a set of lecture notes. Teddy read through that set of notes at the end of the semester. He told me that it can count toward the term paper required for the seminar. So, I guess he is not too disappointed. Although very little of what I learnt from that seminar and subsequently from Teddy gets reflected in my dissertation, they genuinely informed me of how to do mathematically informed philosophy.

Two big events happened in my second year. First, I was lucky to get the opportunity to work as a Research Assistant for Bob Brandom for two years. I mostly worked on the Expressivist Logics project during those two years. Although I decided not to work on that project after the RA-ship, largely because I prefer to identify myself as a philosopher instead of a logician, I'm glad that I encountered the writings of the French logician Jean Yve Girard during that time. I learnt much from my weekly conversations with Bob and wish I have come to those conversations more prepared than I was. Bob is always very generous with his time, and always enthusiastic and encouraging. I'm particularly grateful for him passing down the most important lesson he learnt

from his distinguished career as a scout, that the best way to improve your mud reading skill is to spend more time on mud. Although less known to the broader philosophy community, Bob was in fact an eagle scout.

The second remarkable thing in that year is that I finally learnt Kant's theoretical philosophy from Steve Engstrom. To date, I think that it is the most important thing I learnt in my Ph.D. years. Reading Kant deeply changed how I think about theoretical philosophy, epistemology and otherwise. This could not have happened without Steve, who is not only the best Kant scholar that I know but also a great philosopher. Steve is very willing to engage in questions that Kant has never touched on in writing and provide me with answers based on his authentic understanding of Kant. Talking to Steve is like talking to Kant, perhaps in a deeper and more profound way. Anwar ul Haq, another student of Steve's, once told me that he finds Steve's book so deep that it's more difficult to understand that Kant.

In the third year, apart from continuing working as Bob's RA, I took several classes in Microeconomics with PhD students in Pitt Economics. A lot of what I learnt then went into this dissertation. In addition to that, I did a supervised reading with Japa Pallikkathayil on Political Philosophy. The bulk of the readings was John Rawls' *A Theory of Justice* and subsequent criticisms. I was deeply moved by the book. I recall saying to Teddy on some occasion that after reading Rawls, I find all those technical stuffs no longer important. It is reading Rawls and political philosophers after Rawls (many of whom are Rawls' students) that moved me to write my dissertation of *social and political* epistemology. Meanwhile, I cannot speak higher of Japa, as a teacher and as a philosopher. In the last semester of my Ph.D., I was lucky to have the opportunity to seat in Japa's Ethics Core seminar, where we read Korsgaard's *Source of Normativity*, which eventually helped me understand Kant's practical philosophy.

The fourth year of my Ph.D. happens to overlap with the COVID-19 pandemic. I spent most of my time during that period away from my Ph.D., producing online videos about philosophy in Chinese with a few friends. I'm particularly lucky to meet Shimin Kuang, who worked with me for many months and taught me much about Heidegger and Nietzsche. What impresses me the most of Shimin, is not him as a friend or as a collaborator, but him as a person and his indefatigable pursuit of the greatness of soul (μεγαλοψυχία).

I returned to Pittsburgh at the beginning of the fifth year. Out of various considerations, I decided that I want to do a M.Sc. in Intelligent Systems simultaneously with my Ph.D. in Philosophy. I was lucky to be supported by James and Japa of the philosophy department, and Michael Lewis of the Intelligent Systems program. With their helps, I finished the program by Spring 2023. I'm of course grateful to Huao Li, with whom I co-authored my first peer reviewed publication. Who would have thought that my first peer reviewed publication will be on Artificial Intelligence.

Finally, I'm of course very grateful to all my committee members, without whom I wouldn't have finished the Ph.D. I'm particularly grateful to James. While all three papers I wrote fell out of James' specialized areas, James managed to provide me with tons of helpful feedbacks. Recently, some younger students in the department as well as some newly admitted students asked me what James is like as a supervisor. I said that "I cannot speaker higher of James either as a philosopher, as a supervisor, or as a person". Other members of course also helped a lot. For example, the second paper in this dissertation largely came out of my reflection on John Roemer's book, *How We Cooperate*, which I read with Kevin Zollman a couple years ago.

As I said at the beginning, the dissertation really marks a (perhaps temporary) stop to my journal in professional philosophy. I joined the business of non-professional philosophy much

earlier on, when I was about six years old, reflecting on death and personal identity. (This is of course a retrospective recollection. I only learnt what I was doing back then is philosophy from my philosophy classes in undergraduate studies.) I plan to stay in the business of non-professional philosophy, as a self-sponsored philosopher. Either way, I find the journey extremely rewarding and worthwhile. I am grateful to all the wonderful people I met along the way and feel extremely lucky.

I'd like to thank my parents at the very end of this acknowledgement. Thanks for your unconditional support and love. I'd like to thank my mom for showing and teaching me how to be a good person and thank my father for encouraging me to an idealist, who pursues spiritual achievement, instead of a materialist, who pursues mundane success.

1.0 Institutional Remedies for Individual Expert Identification

Abstract: Goldman (2001) and subsequent literature treat the problem of expert identification as a problem of the study of individual doxastic agents with social evidence. In this paper, I show that system-oriented social epistemology, a different category under Goldman's (2010, 2011) taxonomy of social epistemology, has much to say about the problem of expert identification. Making use of models from the information design literature in economics, I show how purported experts can influence the novice's decision of trust by sending verifiable signals strategically, in pursuit of higher chances of being trusted. The expert's strategic behaviors lead to epistemic failures that can only be avoided with certain institutional regulations, namely regulations on the minimal testing standard. An individual doxastic agent, no matter how hard she tries, cannot avoid such failures on her own.

1.1 Introduction

Human beings have produced more knowledge than any single person can grasp in her lifetime. It is striking to think about the scope of what an ordinary person of our time knows. Like most of my readers, she knows that we live on a round planet that circles the sun in an enormous almost empty space called the universe; she knows that the planet has been warming up rapidly in the past century and CO2, among other greenhouse gases, is causing it; she knows that her body is made up of trillions of cells and almost every cell has the same DNA; etc.

She, of course, did not discover all these by herself. In fact, like most of my readers, she barely has any direct evidence for any of the listed claims. She learned them from others. She learnt them from textbooks, schoolteachers, documentary films, newspapers, and other sources she judges to be credible. The credibility of these sources ultimately comes from experts who possess direct evidence for these claims. Even though she, due to a lack of professional training, may not be able to comprehend the evidence experts present in support of these claims, she nevertheless decides to trust the experts' conclusions.

This description applies to every person in our modern society. No matter whether you are a renowned scholar or an average steel worker, much of what you know comes from others. Hardwig (1985, 1991) argues that epistemic deference to experts is necessary in any sufficiently developed society. A self-reflective person should suspend her own judgment and trust the expert's opinion when she judges herself to be epistemically inferior to the expert. On the extreme end, even if she has some arguments that she finds plausible against the expert's opinion, after failing to comprehend the expert's reply in dismissal of her arguments, the novice can, or perhaps should, rationally suspend her judgment and take the expert's word. Hardwig's proposal sounds particularly compelling when we reflect on how we should react to physicians' opinions on our health. Even if I may not be able to comprehend some of the physician's evidence and arguments, I, as a reflected novice in medicine, should nevertheless take the physician's word. Trust is inevitably blind to some extent.

Granting that it is permissible, or perhaps required by rationality, to take the expert's opinion blindly once one decides to trust an expert, the decision to trust someone should not be made blindly. It should not surprise anyone that it is a risky move to take someone at her word. Trusting an incompetent self-professed expert can lead to disastrous consequences. One could end

up injecting disinfectant in reaction to a potential virus threat as a consequence of trusting a fake expert. Even worse consequences could obtain if epistemic agents of bigger scales — e.g., the state — trusted the wrong expert.

The problem of expert identification was first discussed in Goldman (2001). The difficulty of expert identification comes from the fact that a novice cannot make sense of most evidence possessed (or even most messages sent) by competing experts but must make her decision of trust based on social evidence, e.g., what credentials an expert has, or how popular certain positions are among experts.

After assessing several different types of evidence, Goldman (2001) ended up recommending the expert's past track records of cognitive successes as a promising source of evidence. The novice should wait until the expert's statements become verifiable (or "exoteric") to the novice and see how often the expert achieves cognitive successes. This is possible because some statements, e.g., predictions, can be easily verified retrospectively. The fact that the expert frequently makes true predictions or achieves other cognitive successes in the past gives the novice good reasons to trust the expert's future statements and predictions.

In this paper, I present three mathematical models to show that verifiable signals can be misleading as well. Assuming that there are always track records of cognitive successes that can be verified by the novice, the problem of expert identification does not disappear. The idea, in a nutshell, is that the expert can strategically build up her track records by tackling easier cognitive challenges. Although she cannot decide whether a certain prediction she makes will end up being a success or not, she can nevertheless decide which predictions to make in the first place.

I call this phenomenon "strategic testing", as we may consider the process of making predictions and building track records a process of testing. Unlike medical licensing tests, these

tests are not conducted by meta-experts. Rather, these tests are conducted by Nature, like predicting the return of Halley's Comet. The phenomenon of strategic testing results in an epistemic failure that the novice cannot avoid by her individual efforts but can be mitigated through institutional arrangements. In contrast to Goldman (2011), which takes the problem of expert identification as a paradigmatic example of the study of individual doxastic agents with social evidence under Goldman's (2010, 2011) tripartite taxonomy of social epistemology, this paper shows that "system-oriented social epistemology", another category in Goldman's taxonomy, has much to say about the problem of expert identification.

One may underestimate the importance of models discussed in this paper if one fails to see the significance of the problem of expert identification. Identifying the true expert is not merely about acquiring one true belief (that this expert is competent) or making one right decision (namely, to trust this expert). By trusting a competent expert, a novice can acquire many true beliefs in the expert's domain of expertise that are not accessible to him without the expert. These acquired beliefs will then allow the novice to make good decisions that he could not have made previously. Trusting an incompetent expert, by contrast, not only deprives the novice of the opportunity to learn from a competent expert, but also brings him many false beliefs, which further misguide the novice in decision making. One false belief about the expert's competency and one bad decision about which expert to trust will lead to a sequence of false beliefs and bad decisions. If the novice can improve, albeit slightly, in identifying competent experts, the novice would expect a massive improvement in his epistemic performance. More broadly, if something can be done to help the novices better identify competent experts in general, the community's overall epistemic performance would improve extensively.

Three models are presented in Section 2 to illustrate two similar but distinct phenomena. Section 2.1 presents the simplest possible model of strategic testing to help the reader understand the nature of the phenomenon. In Section 2.2, with the help of a more realistic characterization of tests, an institutional remedy to the epistemic failure is introduced, namely, the regulation of minimal testing standard. In Section 2.3, the model is further developed to include another strategic phenomenon, called "strategic disclosing". There, the expert is allowed to decide when to disclose her test results. With the power of strategic disclosing, the same epistemic failure appears under weaker assumptions. In both cases, the epistemic failure can be avoided with external regulations on minimal testing standards but cannot be avoided by the novice's individual effort.

1.2 The Models

We assume that there are two Bayesian agents in our models: an information sender "the expert" (she) and an information receiver "the novice" (he). The novice must decide whether to trust the expert or not, based on his judgment of the expert's competency, whereas the expert wants to be trusted by the novice regardless of her competency. The expert can influence the novice by sending verifiable signals as results of tests (i.e., establishing track records), which reveal her competency to the novice. The novice gets to observe the results of the test (whether it's a cognitive success or not) as well as the correlation between different test results and the expert's competency, which we may call the *distribution* of the test. The distribution of a test reflects how difficult it is to achieve a cognitive success in this test. The expert gets to choose which test to take but must honestly report the testing result. That is, the expert can control the distribution of the test, but not its result. Such information revealed to the novice is said to be *verifiable*. The models thus differ

from signaling games that have been studied by philosophers before. Signals in signaling games are typically not verifiable but come at a different cost for different types of agents. In our models, known as models of information design, all signals are verifiable.

1.2.1 Strategic Testing: a toy model

We start with a toy model adapted from Kamenica and Gentzkow (2011). Although this model is unrealistic in various respects, its simplicity helps us understand the nature of the strategic phenomenon of interest.

Denote the true state of the expert's competency by $\omega \in \{0,1\}$. $\omega = 1$ means that the expert is competent and thus trustworthy, while $\omega = 0$ means that the expert is incompetent and thus untrustworthy. Suppose that 30% of purported experts are indeed competent and both the expert and the novice know about this fact. For simplicity, without loss of generality, I assume that the expert does not know better about the true status of her own competency than the novice. The novice and the expert thus start with a common prior $P_0(\omega = 1) = 0.3$ vindicated by statistical evidence.[1]

[1] There are at least two reasons to think that the expert and the novice would start with different priors in more realistic settings. (1) One may think that the expert would know better, although still not perfectly, about her competency than the novice does. In economics, we say that the expert in this case has private information. See Hedlund (2017) for a study of this case. (2) The expert may have the same amount of evidence but is simply more confident about herself than the novice. Imagine an incompetent expert who is confident about her competency on false grounds. Alonso and Camara (2016) provides a thorough analysis of this case. What is relevant for this paper is that the strategic phenomenon I discuss persists in both cases. Relaxing the assumption will only introduce complications that are unnecessary for the purpose of this paper.

The novice wants to trust all and only competent experts. In practice, the novice will choose to trust the expert if and only if he is confident enough about the expert's competency. To this end, he chooses a threshold $t \in [\frac{1}{2}, 1]$ for trust. The novice will trust the expert if and only if his credence $P_l(\omega = 1)$ is greater than or equal to t. Different choices of t correspond to different strategies of trust. For example, a novice with $t = \frac{1}{2}$ can perhaps be called a credulist novice, who prefers to trust the expert unless he thinks the expert is strictly more likely to be incompetent than competent. A novice with higher values of t would only choose to trust the expert when he is more confident in the expert's competency. Perhaps, he thinks it is more important to avoid trusting an incompetent expert (i.e., committing an error of false positive) than to avoid not trusting a competent one (i.e., committing an error of false negative). In the extreme case, a novice can have $t = 1$ and only trust experts that he believes to be competent for sure. Such a novice will not trust any incompetent experts but will also fail to trust most if not all competent experts.

We can measure the (in)accuracy of the novice's decision by how often the novice makes a mistake, either trusting an incompetent expert or not trusting a competent one.[2] Formally, we may denote the novice's action by a, which is a function taking the novice's credence in the expert's competency as input and returning 1 (standing for "to trust") or 0 (standing for "not to trust").

$$a = \begin{cases} 1, & P_l(\omega = 1) \geq t \\ 0, & P_l(\omega = 1) < t \end{cases}$$

[2] The result does not depend on this choice of scoring rule. In the appendix, I illustrate the same phenomenon with the Brier score, which I use to measure belief accuracy. Since what is at stake in this model is the novice's decision, it's intuitive to focus on how likely the novice will make the right decision, which justifies the scoring rule I use. The discussion of belief accuracy is left to the appendix to avoid distractions.

The inaccuracy of the novice's decision, measured by how often the novice's decision mismatches the expert's true state of competency, is $E\left[|\omega - a\left(P_l(\omega = 1)\right)|\right]$, where $P_l(\omega = 1)$ stands for the novice's belief in $\omega = 1$ at the time he makes the decision. Since this number measures how often the novice makes a mistake, the lower the number gets, the more accurate the novice's decision is. We may call its complement, $1 - E\left[|\omega - a\left(P_l(\omega = 1)\right)|\right]$, "the novice's decision accuracy". To improve the novice's decision accuracy is equivalent to reducing his decision inaccuracy. In the following, I switch between talks of accuracy and talks of inaccuracy every now and then when I see fit.

Suppose that the novice is somewhat cautious and uses $t = \frac{2}{3}$ as his threshold. Without further information, the novice would choose not to trust the expert since 0.3 is below the threshold. Thus, the novice makes a mistake if and only if the expert is competent. With no testing at all, the novice's decision inaccuracy is 0.3. As to the expert, without any testing, the expert's chance of being trusted is 0 since the novice will always choose not to trust.

Is there anything the expert can do to improve the situation? Yes, the expert can take a test that reveals her competency to the novice as well as to herself. For example, the expert can take an infallible test that all and only competent experts can pass. The test will return $Comp$ if and only if the expert is competent (i.e., $\omega = 1$) and it will return $Incomp$ if and only the expert is incompetent (i.e., $\omega = 0$). Such a test can be represented by the following table.

Table 1: An infallible test.

If ω=0, the test will return Incomp with probability 1. If ω=0, the test will return Comp with probability 1.

	$\Pr(Comp\|\omega)$	$\Pr(Incomp\|\omega)$
$\omega = 0$	0	1
$\omega = 1$	1	0

This test helps both the expert and the novice. Without the test, the expert's chance of being trusted is 0. The novice, without further information, would simply refuse to trust the expert. With the test taken and its result revealed, the expert's chance of being trusted improves from 0 to 0.3. With 0.3 probability, the expert turns out to be competent. Observing the result of the test, the novice updates his belief to $P_l(\omega = 1) = 1$ via a regular Bayesian update and decides to trust the expert. With 0.7 chance, the test would return $Incomp$ and the novice thus learns that the expert is incompetent and decides not to trust the expert. The novice, thanks to the test, always makes the right decision and thus is perfectly accurate in trusting competent experts. Thus, both the expert and the novice would prefer the expert to take the infallible test than not. This example may strike one as trivial. What it shows, however, is that the expert can influence the novice's beliefs and decisions by revealing true information to the novice. With the infallible test, the expert reveals maximally accurate information to the novice and promotes her interests and the novice's interests at the same time.

Will the expert take the infallible test if given the power to take tests? No, not when other tests that can further promote her interests are available.

A typical test measures the state of interest, in our case, ω, with some errors. The infallible test is the special case with no error. Generalizing the infallible test to include the possibility of errors, we get that a general test can be characterized by two parameters ϵ_1 and ϵ_2, as in the

following table. ϵ_1 and ϵ_2 stand for the test's rates of false positive and false negative, which are also known as type 1 error and type 2 error, respectively.

Table 2: A general test.

If ω=0, the test will return Comp with probability ϵ_1 and Incomp with probability 1-ϵ_1.

If ω=1, the test will return Comp with probability 1-ϵ_2 and Incomp with probability ϵ_2.

	$Pr(Comp\|\omega)$	$Pr(Incomp\|\omega)$
$\omega = 0$	ϵ_1	$1 - \epsilon_1$
$\omega = 1$	$1 - \epsilon_2$	ϵ_2

As shown in the table, if $\omega = 0$, with probability $1 - \epsilon_1$, the test measures correctly and returns $Incomp$, and with probability ϵ_1, it runs into an error of false positive and returns $Comp$ instead. Similarly, if $\omega = 1$, with probability $1 - \epsilon_2$, the test returns $Comp$ correctly and with probability ϵ_2, it commits an error of false negative and returns $Incomp$. Since a test is uniquely determined by its error rates, we may denote a test by a tuple (ϵ_1, ϵ_2) for convenience. For example, the infallible test can be denoted as $(\epsilon_1 = 0, \epsilon_2 = 0)$.

Suppose that testing options are abundant in the sense that for any pair of $\epsilon_1, \epsilon_2 \in [0,1]$, there is a test (ϵ_1, ϵ_2) available to the expert.[3] This allows the expert to optimize her chance of

[3] This assumption can be loosened so that there are only countably or perhaps finitely many tests available. The continuous case is an idealization of the more realistic countable/finite cases. If only finitely many tests are available, the expert will optimize for her interest over all finitely many available tests in the same way as in the model.

being trusted across all possible tests by specifying ϵ_1 and ϵ_2. For the specific model we have so far specified, the expert's chance of being trusted is maximized by test $(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$.

In the case of infallible tests, the signal $Comp$ would increase the novice's belief in the expert's competency to 1, which is unnecessarily high from the expert's point of view. The novice would trust the expert as long as his credence reaches $\frac{2}{3}$. Once the novice's credence gets to $\frac{2}{3}$, the expert has no interest in further improving it. By carefully calibrating the false-positive rate (ϵ_1) of the test, the expert makes it the case the novice's credence would precisely go to $\frac{2}{3}$ after observing $Comp$. At the same time, the new test $(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$ is more likely to return $Comp$ than the infallible tests, because it will return $Comp$ for sure when the expert is competent and will also return $Comp$ with some probability ($\frac{3}{14}$) when the expert is not.

With test $(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$ taken, the expert's chance of being trusted increases to 0.45, which is a further improvement from 0.3 as in the case of the infallible test. On the other hand, it brings the novice's decision inaccuracy to 0.15, which is better (i.e., lower) than 0.3 in the case of no test but worse than 0 in the case of the infallible test. I call this phenomenon "strategic testing", where the expert strategically chooses the test that maximizes her chance of being trusted at expense of the novice's decision accuracy. It is misleading for the novice for two reasons: (1) the novice fails to achieve the ideal epistemic state brought by the infallible tests and (2) the novice would end up trusting 45% of all experts despite that he knows that only 30% of purported experts are competent!

Knowing the test is strategically chosen by the expert to influence his choice, why would the novice decide to update upon the result of this test? Why wouldn't the novice just ignore the expert's test result to avoid misleading information? Notice that the novice's accuracy of decision

is, in fact, improved, in presence of test ($\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0$). More generally, I show in Appendix A1 that in this model updating on test results will never have negative impacts on the novice's epistemic performance. For this reason, it is always rational for the novice to update his belief with the information revealed by the expert.

This toy model shows how the existence of verifiable signals helps the novice to improve epistemic performance. However, it also shows how the expert can strategically design the verifiable signal to influence the novice. The novice, by himself, cannot avoid such influence. In this model, the expert and the novice started with the same amount of information about the expert's competency level ω. What the expert has in her favor is the power to choose which test to take. Utilizing this power, the expert manipulated the novice's beliefs and decisions for her own interests.

An obvious limitation of this toy model is how tests are characterized. In reality, we do not know much about the error rates of tests and cannot choose which test to take by specifying its rates of errors. This feature of the model prevents us from having a fruitful discussion about potential remedies. Ideally, we would like every expert to take the infallible test to reveal her true state of competency so that the novice's epistemic performance can be optimized. But an infallible test rarely exists if at all. This limitation can be overcome with a more realistic characterization of tests, which also allows us to discuss institutional remedies for the identified suboptimality.

1.2.2 Strategic Testing: choosing difficulty levels

Instead of specifying the rates of errors, the expert now must choose tests according to their difficulty levels. A more challenging test is harder to pass but will have more influence on the novice's beliefs if passed. An easier test, on the other hand, is more likely to have positive, albeit smaller, impacts over the novice. This tradeoff is very common in our everyday life. For example, a graduate student must choose which journal to submit her paper to. A top journal is harder to get in but boosts her chance of getting a job if her paper does get accepted. A lesser journal is less impressive to search committees but makes it easier for her to secure a publication. The strategic phenomenon identified in this model is essentially the same as in the previous model. A more realistic characterization of tests, however, facilitates philosophical discussions of potential remedies, as well as prepares us for the further sophistication discussed in the third model.

A test, generally speaking, measures the underlying state of interest (ω) with some error (ϵ). We may think that the error ϵ in detecting ω is exogenously determined by currently available technology. For simplicity, let us assume that the error ϵ is normally distributed, with mean $\mu = 0$ and standard deviation $\sigma = 0.5$, i.e., $\epsilon \sim N(0, 0.5)$. The random variable $S = \omega + \epsilon$ is a test that measures ω with error ϵ. One can update his belief about ω upon observing the value of S in a familiar Bayesian manner.

Most tests in the real world, however, would not return a real number as its result. A typical test has its results divided into finitely many different categories, either in terms of letter grades (ABCDF), pass/fail, or finitely many different grade scales (e.g., 100 points). Tests are so constructed for practical reasons. A test with infinitely many different potential outcomes is infeasible while fewer categories make test results more suitable for communication and comprehension. For our purpose, we focus on binary tests with two potential outcomes: success

or failure. The phenomena we identify however hold for all tests with finitely many different potential outcomes.

We can convert the real-valued signal S into a binary test by discretizing it with a testing standard x.

$$S_x = \begin{cases} 1, & P(\omega = 1|S) > x \\ 0, & P(\omega = 1|S) \leq x \end{cases}$$

When the posterior in $\omega = 1$ upon observing the value of S exceeds the testing standard x, the binary test S_x would return 1. And it will return 0 otherwise.[4]

The expert chooses which test to take by specifying a testing standard $x \in [0,1]$, which can be plausibly interpreted as the difficulty level of the test. The novice, knowing the testing standard x, observes the value of S_x, which is either 0 or 1, and updates his belief accordingly. It is important to emphasize again that the novice knows exactly how the test result is generated, what the testing standard is, and updates correctly using the Bayes rule. The technical treatments of this model can be found in Appendix A2.

As in the first model, the choice of test that is optimal for the expert is suboptimal for the novice. We may denote the expert's optimal test, i.e., the test that maximizes the expert's chance of being trusted, by $x^{\dagger}$ and denote the novice's optimal test, i.e., the test that minimizes the

[4] This way of discretization is equivalent to setting a threshold s_t on the value of S, so that the binary test returns 1 if and only if $S > s_t$. We can translate back and forth between these two ways of discretization via the fact that $P(\omega = 1|S) > x$ if and only if $s > \sigma^2 * \log\frac{(1-p)*x}{p*(1-x)} + \frac{1}{2}$. It's easier for us to think about the testing standard as a threshold x on the posterior belief instead of s_t on the value of S, as x must fall between 0 and 1 while s_t and S could be any real number.

novice's decision inaccuracy, by x^*. The information about $x^\dagger$ and x^* of our particular example are summarized in the following table.

Table 3: A comparison between $x^\dagger$ and x^*

	Testing Standard	Expert's chance	Novice's decision inaccuracy
x^*(optimal for the novice)	0.5	0.269	0.139
$x^\dagger$(optimal for the expert)	0.258	0.389	0.17

The expert, given the power to choose the testing standard, would choose $x^\dagger$ to maximize her chance of being trusted, while the novice's epistemic performance can be optimized by x^* which is distinct from $x^\dagger$. Thus, we have identified the same phenomenon as in the first toy model.

Unlike the first model where the expert gets to choose the error rates of the test, the error rate of tests in this model is fixed by the exogenous error term ϵ. The expert has no control over the distribution of ϵ. Instead, the expert influences the novice's decision by choosing a testing standard x, for the difficulty level of the test.

In this model, the novice's epistemic performance can be improved if we force the expert to take a test with testing standard x^* instead of $x^\dagger$. This remedy is feasible since both x^* and $x^\dagger$ are based on the same technology reflected by ϵ. Instead of directly asking the expert to take a test with testing standard x^*, this remedy can be achieved by setting a minimal testing standard at x^*. The minimal testing standard at x^* inhibits the expert from taking tests with standards below x^*. When such regulation is in place, the expert, motivated by her self-interest, will choose x^* over other available testing standards. The novice's epistemic performance is thus optimized. This

regulation of minimal testing standard removes the suboptimality in general for this model. Detailed treatments can be found in Appendix A2. An intuitive explanation is as follows: once passing the optimal level $x^{\dagger}$, the expert's chance of being trusted decreases as the testing standard goes up. Since, as shown in the appendix, x^{*} is always greater than or equal to $x^{\dagger}$, when all values below x^{*} are not available, x^{*} maximizes the expert's chance of being trusted. Once all testing levels below x^{*} are prohibited, it's to the best interest of the expert to choose x^{*}.

This model vindicates the widespread practice of setting a minimal testing standard in real world expertise licensing against potential libertarian deregulation positions on certification. For example, in the United States, like in most other countries, many private agencies can issue medical licenses, but they must provide tests that are above the state's minimal standard. In principle, passing a medical test that is below the state's minimal standard also shows something about the expert's competency. Why not allow people to do that? This model shows that purported medical workers will tend to take easier tests, should those tests be available, to maximize their chance of being trusted by the novice. What is more important for the problem of expert identification is that this model applies to track record building as well. Not only can an expert choose to make easier licensing tests, whose reliability is backed up by meta-experts, an expert can choose to take on easier cognitive challenges, making easier-to-make predications, in pursuit of a higher chance of being trusted.

This model, like all models, has its own limitations. So far, we have been assuming that the novice makes his decision based on a threshold t, which is known to the expert. The strategic moves of the expert are made possible by this assumption. One may wonder what if the novice doesn't make his decision by a fixed threshold, but instead makes his decision in a continuous manner. Say, the novice randomizes his decision in accordance with his credence in the expert's

competence. Even if the novice does use a threshold to make his decision, what if the expert is ignorant about it?

In the next section, I introduce a further sophistication, where the expert is ignorant about how the novice makes a decision except that her chance of being trusted increases as the novice's credence in her competency does. The phenomenon of strategic testing persists in that model, once the expert is allowed to disclose her test results strategically.

1.2.3 Strategic Testing and Strategic Disclosing

How would the expert decide which test to take if she does not know how the novice makes his decision? In this case, it is no longer possible for the expert to maximize her chance of being trusted. The expert must instead aim to maximize the expectation of the novice's posterior in her competency, assuming that the novice is more likely to trust her as his credence increases.

The Martingale Property of Bayesian updating, commonly known as the Reflection Principle[5] in the philosophical literature, requires that for any prior probability distribution $P(\cdot)$ and any test S, the expectation of the posterior $E[P(\cdot\,|S)]$ must be equal to the expectation of the prior $E[P(\cdot)]$.For this reason, any test seems to be equally good (or bad) at increasing the expectation of the novice's posterior as other tests or no test at all. Thus, it may seem that the expert must be indifferent about all tests or even no test. This is true if the expert must always disclose the result of testing. However, the indifference vanishes when the expert can decide when to disclose her testing results strategically.

[5] For a recent philosophical discussion of this principle, see Huttegger and Nielsen (2020).

The practice of strategic disclosure is almost ubiquitous in our daily life. For example, a regular academic job candidate will not disclose rejections from academic journals in her job application. She will only report the good news: acceptance from the Journal of Philosophy, "Revise and Resubmit" from Philosophical Review, etc. For the ones that did not get accepted or "Revise and Resubmit", the candidate would simply not mention them. The search committee, who reads the candidate's application, thus would not even know if the candidate has submitted those papers to a journal or not, and *a fortiori* which journals she submitted to. Similarly, a lawyer would tell you when and where he was admitted to the bar but would not mention that he has failed seven times before he passed the bar exam, even if he did fail that many times.

The general idea of strategic disclosure is as follows. After taking a test, I would only disclose results that I find in my favor. I would avoid mentioning negative testing results whenever possible. The information receiver, in that case, would not even know if I have taken a test at all, and *a fortiori* what test I took or what result I got.

The phenomenon of strategic disclosure is especially important in the expert-novice scenario under our investigation. It is arguably not as important in the academic job market example. The members of a search committee are experts who can directly assess the candidate's competency via, say, her writing samples, without relying on her credentials or track records. Such direct assessments are not available to a novice in an expert-novice scenario, due to the novice's ignorance about the expert's domain of expertise. The novice must rely on the test results disclosed by the expert. In the following, I use results from DeMarzo, Kremer and Skrzypacz (2019) to discuss the impact strategic disclosure has on the expert-novice scenario.

To include the phenomenon of strategic disclosure in our model, we must first add the genuine possibility of not taking a test into the model, otherwise the novice would simply know

that the result is negative whenever the expert refuses to disclose. A *null result* represents the possibility that the expert did not take a test in the first place and thus has no result to be disclosed. A null result cannot be verified. When the expert reports a null result, the novice cannot distinguish between the state where the expert has no result to disclose from the state where the expert has some result but prefers not to disclose it. When the expert reports the null result, the novice also does not learn which test the expert has taken, as he does not even know if the expert has taken a test. Meanwhile, all other test results, if disclosed, are verifiable. Like before, the expert cannot fabricate test results and must report honestly if she chooses to disclose. This clearly maps onto real-world scenarios. While you can ask someone for a certificate/transcript to verify her claimed testing results, it is difficult to challenge someone's claim that she did not take a test, as there is no certificate issued when no test is taken.

We illustrate the general phenomenon with a concrete example and put further details in Appendix A3 for interested readers. As in the previous model, there is an underlying state of interest $\omega \in \{0,1\}$ that represents the expert's competency. Let $P(\omega = 1) = 0.3$ like before. Random variable $S = \omega + \epsilon$ measures ω with error $\epsilon \sim N(0,1)$. We, in addition, assume that with some probably, say 0.02, the expert did not take a test. In that case, $S = \emptyset$. We then discretize S into S_x with a threshold hold x, as follows:

$$S_x = \begin{cases} 1, & P(\omega = 1|S) > x \\ 0, & P(\omega = 1|S) \leq x \\ \emptyset, & S = \emptyset \end{cases}$$

The expert in this model would have to make two decisions, namely which test to take and when to disclose. Denote the expert's disclosing strategy by θ, which decides for each potential test result $S_x \in \{0, 1, \emptyset\}$ whether to disclose it or not. Formally, it is convenient to think θ as a map from $\{0, 1, \emptyset\}$ to $\{D, N\}$, where N in the codomain stands for "not to disclose" and D stands for "to

disclose". When $S_x = \emptyset$, the expert has nothing to disclose. Thus, $\theta(\emptyset) = N$, regardless of the value of x.

We analyze this scenario as a game using the concept of Nash equilibrium. A Nash equilibrium is a profile of players' strategies so that no player would benefit from deviating to a different strategy with the other player's strategy held fixed. The expert chooses a pair (x, θ) to maximize the expectation of the novice's posterior belief in her competency, where x is the testing standard as in the last model and θ is the expert's disclosing strategy. The novice will have to choose a strategy to form his posterior beliefs. In the case of disclosure, the novice would update his belief in a familiar Bayesian manner, after observing x and S_x. In the case of nondisclosure, the novice cannot update his belief in the familiar Bayesian manner since there is no verifiable signal he can update on. Instead, he must choose a posterior P^N in reaction to the null signal according to the expert's strategy. This fact that the novice must choose a posterior P^N when no information is disclosed and the assumption that players have common knowledge about each one's strategy profile, which is essential for the concept of Nash equilibrium, play crucial roles in solving this game. It is difficult to give an intuitive description of how the game is solved without going into some of the technical details. In the following, I present the main result from DeMarzo, Kremer and Skrzypacz (2019). Interested readers should consult the appendix for details.

Proposition: At Nash equilibrium, the expert's equilibrium strategy (x_E, θ_E) minimizes the expectation of the novice's non-disclosure posterior P^N.[6]

[6] DeMarzo, Kremer and Skrzypacz (2019, p. 2181, Proposition 1). See my Appendix A3 for a more intuitive and accessible proof.

This proposition turns the problem of solving the game into an optimization problem. It further reduces to a one-dimensional optimization problem once we realize that, for our example, the disclosing strategy θ^* that maps 0 and $\emptyset$ to N, 1 to D, minimizes P^N regardless of the choice of x. (See Appendix A3 for more details.) Thus, we can hold θ^* fixed and search for the x that minimizes P^N.

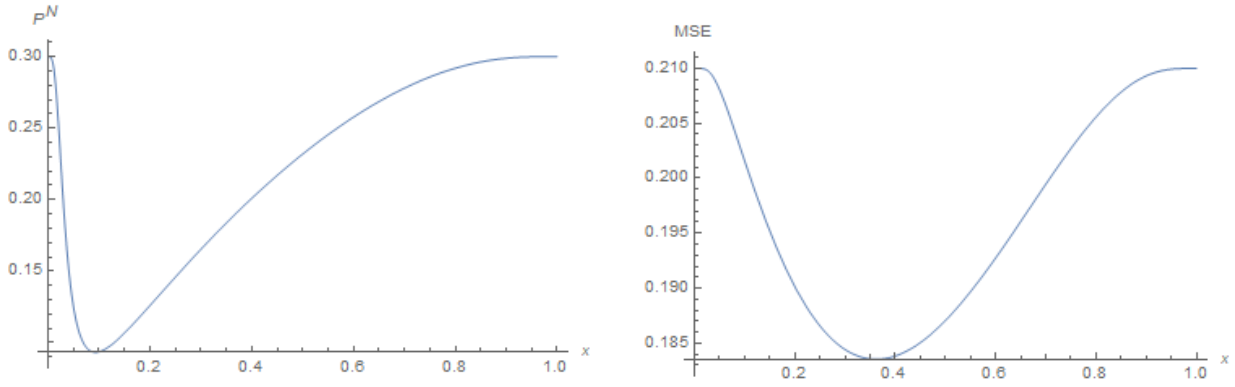

Figure 1: Left: x and P_x^N; Right: x and the novice's mean squared error.

The figure on the left of Figure 1 shows the relation between x and P_x^N. P_x^N is minimized at $x = 0.0933$. This means that the expert would choose testing standard $x = 0.0933$ at equilibrium. The question of our interest is how the novice performs epistemically. Since the novice's decision to trust is no longer part of the model, we must adopt an alternative measurement to assess the novice's epistemic performance. The only sensible measurement in this case is the novice's belief accuracy. While there are many ways to define it, the most standard way is to take the mean squared error (MSE) of his belief, i.e., $E\left[\left(\omega - P(\omega|S_x)\right)^2\right]$, a measurement of how inaccurate his belief is. His belief accuracy is, thus, $1 - E\left[\left(\omega - P(\omega|S_x)\right)^2\right]$. The right half of Figure 1 shows the relation between x and the novice's MSE. The novice's epistemic performance is maximized (i.e., his MSE is minimized) at $x = 0.3622$, where the novice's MSE is 0.1835. At the Nash equilibrium, the novice's MSE, in contrast, is $0.205 > 0.1835$. Thus, the expert, if left

ungoverned, will choose a testing standard that is suboptimal for the novice's epistemic performance, as in the last model, even though the expert is ignorant of the novice's decision strategy.

Like in the second model, if we set a minimal testing standard at the level that is optimal for the novice, the expert will choose that minimal level and thus take a test that is the best for the novice. This point can be seen from Figure 4.1. Recall that the expert in equilibrium chooses the testing standard x to minimize P_x^N. If we set the minimal testing standard at $x = 0.3622$, the expert would choose 0.3622 since it is the minimizer of P_x^N among all testing standards that are available to the expert. DeMarzo, Kremer and Skrzypacz (2019, p. 2189, Proposition 4) show that this phenomenon as well as the remedy of minimal testing standard hold under very general conditions.

This model exhibits similar epistemic failure of the novice without the assumption that the expert knows the novice's strategy of decision making. The expert is ignorant about how the novice makes his decision except that her chance of being trusted increases as the novice's credence in her competency increases. We may think that this is the case where the expert takes one test and decides whether to disclose it for all novices, in contrast to the two previous models, where the expert tailors her test for a single novice with a known strategy of decision. The analysis of the model makes use of the concept of Nash equilibrium, which contends that the expert will deviate whenever there are more profitable strategies available. At the unique Nash equilibrium, the expert will choose a test and a disclosing strategy that is suboptimal for the novice's epistemic performance. The identified suboptimality in the novice's epistemic performance, like before, cannot be avoided by the novice's own effort. Rather, it must be eliminated by some external regulations, namely, to put the expert under the regulation of a minimal testing standard.

1.3 Philosophical Reflections: Epistemic Virtues at Institutional Level

The proposed regulation, by eliminating the identified suboptimality, provides a Pareto improvement[7] of the community's epistemic performance measured by true beliefs. It has been shown that when the regulation of minimal testing standard is imposed, both the novice and the expert learn better about the expert's competency. The outcome under regulation thus Pareto dominates the outcome without regulation, for every player has more accurate beliefs when the regulation is in place. The only factor that prevents this epistemically preferable situation from obtaining is the expert's interest in being trusted, which is epistemically irrelevant both at an individual level and at a community level.

There is no doubt that a veritist, like Goldman (1999), who takes true beliefs as the fundamental epistemic value, will be in favor of the proposed regulation. However, one does not have to be a veritist to find the proposed regulation epistemically virtuous. There is no good reason to think that an improvement in true beliefs, on its own, will lead to losses in justified beliefs or knowledge. So long as one thinks that true beliefs are epistemically valuable, even if not the only epistemic value, one would find the proposed regulation leading to a Pareto improvement. Even for those who prefer to measure one's epistemic performance solely by justified beliefs or by knowledge, the regulation will still more likely than not lead to a Pareto improvement in the community's epistemic performance, given that these desiderata either presuppose true beliefs (as

[7] A Pareto improvement is a change in allocation that harms no one and helps at least one person. The improvement under consideration is not Pareto if measured in terms of the players' utilities. When the regulation is in place, the expert's chance of being trusted decreases. However, the improvement is Pareto if we focus on the agents' epistemic performance only.

in the case of knowledge) or are strongly correlated with true beliefs (as in the case of justified beliefs). Since there is no epistemic trade-off involved, even epistemic anti-consequentialists, like Berker (2013), should be satisfied with the regulation.

An alternative way to eliminate the identified phenomenon in the third model is to deprive the expert of the power to disclose strategically. This solution is, however, less appealing for the following three reasons. (1) In the third model, if the expert must disclose honestly, the expert will be indifferent to all tests, as well as no test, because of the reflection principle of Bayesian updating. This leaves the expert with no incentive to take tests and thus deprives the novice of the chance to learn about the expert's competency from the expert's testing results.[8] (2) This solution does not help with the first two models since strategic disclosure is not assumed there. It, thus, does not stop the expert from testing strategically, so long as the expert has some beliefs about how people make decisions. (3) It is much more morally questionable to hold people obligated in disclosing their testing results. One may argue that the mandated disclosure of testing results would compromise, for example, individuals' right of privacy.

Since the publication of Goldman's seminal paper on the problem of expert identification, this issue has been discussed from the perspective of an individual doxastic agent with social evidence. The novice takes the external circumstances as given and contemplates what he should do to achieve best epistemic results in presence of social evidence. In this paper, I adopt mathematical models from economics to show that this is also a problem of what Goldman calls

[8] Like in this first model, nontrivial tests (i.e., $x \notin \{0,1\}$) always improve the novice's belief accuracy. This point can be seen from the figure on the right of fig. 4.1 and is proven generally in Kamenica and Gentzkow (2011, prop. 1).

"system-oriented social epistemology", the study of epistemic systems and epistemic institutions in pursuit of better epistemic outcomes. I argue that the identified failure in epistemic outcomes can only be solved at an institutional level, with the regulation of the minimal testing standard.

One may think that perhaps the identified suboptimality can be solved at an individual level, not by the novice, but the expert. One may say that the expert, who misleads the novice with her power of testing, is to be blamed for the failure in the epistemic outcome, just like in cases where epistemic failures obtain as a result of someone lying to another. I would like to end the paper with a brief reflection on this point.

When the proposed regulation is not in place, has the expert committed a wrong in making those strategic moves? What the expert did in the model is distinctly not lying or deceiving. The expert did not convey any false information to the novice. A verifiable test result never hurts the novice's epistemic performance and sometimes improves it. What is potentially culpable of the expert is that in presence of the option x^* that is epistemically best for the novice, the expert chooses the option $x^\dagger$ that is best for her own interest but suboptimal for the novice.

One may think that the expert conducted a wrong in this case. For example, one may justify this claim on the ground that in the Kingdom of Ends, as one may imagine, an expert would always choose the test that best promotes the novice's epistemic status, given how important it is for the novice to find the competent experts. One may add that the wrong conducted by the expert is of a particularly epistemic kind, as it undermines the novice's capacity as an epistemic agent. Although this view has a ring to it, it is simply counterintuitive that the expert should be held responsible for failing to optimize for the novice's epistemic performance.

I think this is a case where institutional virtues and individual virtues diverge.[9] An example of such a divergence can be found in John Rawls' *A Theory of Justice*. Rawls proposes to design the basic social institutions of a just society in accordance with his two principles of justice but only requires the individuals in a just society to follow and promote just social institutions. Individuals in a just society are not obligated to regulate their actions in line with the two principles. Similarly, in our case, it is a virtue of our epistemic institutions to regulate tests with a minimal testing standard to promote epistemic values at the community level. It is, however, at best supererogatory for an expert to choose tests that are optimal for the novice. For this reason, if we only look for moral or epistemic virtues at an individual or a transactional level, we will miss structural phenomena and institutional virtues that are crucial. This is what motivates me to focus my discussion on epistemic institutions, instead of individual epistemic agents, in this paper.

1.4 Appendix

In this appendix, I present detailed technical treatments of models presented in the paper. The appendix is divided into three parts, one for each model. Most calculations discussed in the appendix are undertaken with Mathematica. The code I used can be found in the supplemented Mathematica file.

[9] For an example of such cases in epistemology, see Mayo-Wilson, Zollman and Danks (2011).

1.4.1 Appendix 1. The First Model

In this model, the underlying state of interest is the expert's competency, noted by $\omega \in \{0,1\}$, where 0 stands for incompetent and 1 for competent. ω follows a Bernoulli distribution $B(p)$ such that for an arbitrary expert chosen from all self-regarded experts, $\Pr(\omega = 1) = p$. Two agents both know this statistical fact and thus start with a common prior $P_0(\omega = 1) = p$. For simplicity, we assume that the expert has no private information about the true value of ω.

The novice decides whether to trust the expert based on his belief in $\omega = 1$. To this end, the novice sets a threshold t such that he decides to trust if and only if his posterior in ω is greater than or equal to t. The expert chooses a test to inform the novice of the true value of ω. The test has two potential outcomes: $Comp$ and $Incomp$. The expert chooses a test by specifying two error rates $\epsilon_1 = \Pr(Comp|\omega = 0)$ and $\epsilon_2 = \Pr(Incomp|\omega = 1)$, for the false positive rate and false negative rate of the test, respectively. ϵ_1 and ϵ_2 uniquely determines a test and thus, we can denote a test by (ϵ_1, ϵ_2).

Upon observing the test result, the novice would update his belief based on the signal he saw and the error rates. The expert wants to maximize the probability that the novice chooses to trust her. The expert pursues this end by choose the values of ϵ_1 and ϵ_2. Once ϵ_1 and ϵ_2 are fixed, the novice observes a verifiable signal generated according to ϵ_1 and ϵ_2. The expert cannot control the signal apart from choosing ϵ_1, ϵ_2.

We first consider if there is a test (ϵ_1, ϵ_2) the expert can choose so that the novice always chooses to trust the expert regardless of the testing result. If so, this test would maximize the expert's chance of being trusted to 1, for the novice will always choose to trust the expert. This would require the following two conditions to hold.

$$P_0(\omega = 1|Comp) = \frac{P_0(Comp|\omega = 1)P_0(\omega = 1)}{P_0(Comp|\omega = 1)P_0(\omega = 1) + P_0(Comp|\omega = 0)P_0(\omega = 0)}$$
$$= \frac{(1 - \epsilon_2) * p_0}{(1 - \epsilon_2) * p_0 + (1 - \epsilon_1) * (1 - p_0)} \geq t \quad (1)$$

$$P_0(\omega = 1|Incomp)$$
$$= \frac{P_0(Incomp|\omega = 1)P_0(\omega = 1)}{P_0(Incomp|\omega = 1)P_0(\omega = 1) + P_0(Incomp|\omega = 0)P_0(\omega = 0)} \quad (2)$$
$$= \frac{\epsilon_2 * p_0}{\epsilon_2 * p_0 + (1 - \epsilon_1) * (1 - p_0)} \geq t$$

Some calculation would reveal that these two inequalities cannot be satisfied at the same time unless $p_0 \geq t$ at the first place. That is, unless the novice would have trusted the expert without any testing at the first place, there is no way for the expert to take a test so that the novice trusts her regardless of the result. Thus, the best the expert can expect is that the novice would choose to trust her upon seeing $Comp$, and not to trust her upon seeing $Incomp$.[10] To this end, we reverse the direction of inequality in the second condition above and get the following two conditions, which characterizes this case, known as a separating equilibrium in economics, for the novice takes different actions toward different signals:

$$P_0(\omega = 1|Comp) = \frac{P_0(Comp|\omega = 1)P_0(\omega = 1)}{P_0(Comp|\omega = 1)P_0(\omega = 1) + P_0(Comp|\omega = 0)P_0(\omega = 0)}$$
$$= \frac{(1 - \epsilon_2) * p_0}{(1 - \epsilon_2) * p_0 + (1 - \epsilon_1) * (1 - p_0)} \geq t \quad (3)$$

[10] $Comp$ and $Incomp$ are mere labels. It does not make a difference if we switch them. That is, the expert can equivalently choose tests so that the novice trusts her if and only if the result is $Incomp$.

$$P_0(\omega = 1|Incomp)$$

$$= \frac{P_0(Incomp|\omega = 1)P_0(\omega = 1)}{P_0(Incomp|\omega = 1)P_0(\omega = 1) + P_0(Incomp|\omega = 0)P_0(\omega = 0)} \quad (4)$$

$$= \frac{\epsilon_2 * p_0}{\epsilon_2 * p_0 + (1 - \epsilon_1) * (1 - p_0)} < t$$

To maximize the chance of being trusted, the expert only needs to maximize the probability the novice sees $Comp$. The novice sees $Comp$ with probability $\Pr(\omega = 0) * \Pr(Comp|\omega = 0) + \Pr(\omega = 1) * \Pr(Comp|\omega = 1) = (1 - p_0) * \epsilon_1 + p_0 * (1 - \epsilon_2))$. Thus, the expert solves the following constraint optimization problem to find the optimal test (q_0, q_1) that maximizes her chance of being trusted.

$$\arg\max_{0 \le q_0, q_1 \le 1} (1 - p_0) * \epsilon_1 + p_0 * (1 - \epsilon_2), subject\ to\ (3), (4)$$

This concludes the formal treatment of this simple model of information design. Next, I use the numerical examples I discussed in the paper to give some mathematical intuitions about this phenomenon.

1.4.1.1 A1.1 Numerical Examples

Let $p = 0.3$ and $t = \frac{2}{3}$. Following the general solution procedure, we find that $(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$ maximizes the expert's chance of being trusted. The Mathematica code for solving the optimization problem is supplemented.

The novice's posteriors after observing different results of test $(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$ can be calculated using the Bayes Rule as follows:

$$P_l(\omega = 1|Comp) = \frac{Pr(Comp|\omega = 1)P_0(\omega = 1)}{Pr(Comp|\omega = 1)P_0(\omega = 1) + Pr(Comp|\omega = 0)P_0(\omega = 0)}$$

$$= \frac{(1 - \epsilon_2) * p_0}{(1 - \epsilon_2) * p_0 + \epsilon_1 * (1 - p_0)} = \frac{2}{3}$$

$$P_l(\omega = 1|Incomp) = \frac{Pr(Incomp|\omega = 1)P_0(\omega = 1)}{Pr(Incomp|\omega = 1)P_0(\omega = 1) + Pr(Incomp|\omega = 0)P_0(\omega = 0)}$$

$$= \frac{\epsilon_2 * p_0}{\epsilon_2 * p_0 + (1 - \epsilon_1) * (1 - p_0)} = 0$$

A regular Bayesian update would lead the novice's posterior to $\frac{2}{3}$ after observing $Comp$ from test ($\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0$), and to 0 after observing $Incomp$. Thus, the novice would choose to trust the expert if and only if he observes $Comp$. The result $Comp$ could obtain in two ways: either the test correctly reports a competent expert as competent, or it incorrectly reports an incompetent expert as competent. The probability of observing result $Comp$ is thus $P_0(\omega = 1) * \Pr(Comp|\omega = 1) + P_0(\omega = 0) * \Pr(Comp|\omega = 0) = 0.3 * (1 - \epsilon_2) + 0.7 * \epsilon_1 = 0.45$. By choosing to take this test instead of the infallible one, the expert raises her chance of being trusted from 0.3 to 0.45.

How did this happen? The martingale property of Bayesian updating only requires the expectation of a Bayesian agent's posterior equal to his prior. The martingale property, however, does not constrain the distribution of the agent's posterior. By optimizing the signal to be sent, the expert makes the novice's posterior as concentrated on $\Pr(\omega = 1) = \frac{2}{3}$ as possible. In this example, the novice's posterior is either 0 or $\frac{2}{3}$. The expert's chance of being trusted is thus maximized.

Notice that $\epsilon_2 = 0$, which means this test has zero false negative rate. Thus, after observing the test result, the novice would make the right decision (i.e., to trust the expert) whenever the

expert is competent. In contrast to the case without testing, where the novice always chooses not to trust the expert, the novice in this case would trust all competent experts and some incompetent ones. Intuitively, this strategically chosen test improves the expert's chance of being trusted by moving the novice's errors from false negative to false positive. It is worth noting that the novice also benefits from this test compared to no test taken. The novice's decision inaccuracy drops from 0.3 to 0.15, in presence of the test. The novice will trust $\frac{3}{14}$ of incompetent experts, which constitute 70% of all self-regarded experts and thus makes a mistake with probability 0.15.

The phenomenon exists under fairly general conditions. More general treatments can be found in Kamenica and Gentzkow (2011). I shall illustrate the point with some relevant examples here.

(1) One may think that any realistic test necessarily suffers from two kinds of statistical errors, which makes the condition $\epsilon_2 = 0$ particularly questionable. The phenomenon persists if we mandate some positive rate of false negative, by putting some constraints on ϵ_2. Say, let $\epsilon_2 \geq 0.1$. The test thus must have a false negative rate of at least 0.1. Solving the model with this additional constraint, we find that the expert can improve her chance of being trusted to 0.405 under with $\epsilon_1 = \frac{27}{140}$ and $\epsilon_2 = 0.1$.

(2) One may find the asymmetry between ϵ_1 and ϵ_2 in the example somewhat concerning. Maybe, this peculiar asymmetry is the sole source of the expert's influence over the novice. It is not so. We can restrict the expert's choice to tests that are symmetric in the sense that $\epsilon_1 = \epsilon_2$ only. A symmetric test is equally sensitive in detecting competent experts and incompetent experts. The same calculation procedure, which can be found in the appendix or the supplemented Mathematica file, shows that the expert maximizes her chance of being trusted with $\epsilon_1 = \epsilon_2 \approx 0.176471$, where the expert's chance of being trusted is about $0.370588 > 0.3$.

1.4.1.2 A1.2 Accuracy

I discussed the novice's (in)accuracy of decision in the paper. We can alternatively measure how well the novice is doing epistemically by his belief (in)accuracy. There are different ways to define belief accuracy. A standard measure is by the mean squared error of his belief compared to the true state, which is used in the last model of the paper. The novice's belief accuracy is thus $E\left[\left(\omega - P_0(\omega|S)\right)^2\right]$, where S denotes the testing result the novice observes and $P_0(\omega|S)$ is the posterior belief of the novice after observation. I list the agent's belief and decision inaccuracies for examples discussed in the paper here for comparison.

Table 4: A comparison of various tests.

Test	Belief Inaccuracy	Decision Inaccuracy	Chance of Trust
No Test	0.21	0.3	0
$(\epsilon_1 = 0, \epsilon_2 = 0)$	0	0	0.3
$(\epsilon_1 = \epsilon_2 = 0.176471)$	0.130841	0.3	0.370588
$(\epsilon_1 = \frac{3}{14}, \epsilon_2 = 0)$	0.1	0.15	0.45
$(\epsilon_1 = \frac{27}{140}, \epsilon_2 = 0.1)$	0.118487	0.135	0.405

It should not be surprising that there is a positive correlation between the novice's belief inaccuracy and his decision inaccuracy. After all, the novice's decision of trust is made according to his belief and the threshold t he chooses.

It is worth noting that the symmetric test $(\epsilon_1 = \epsilon_2 = 0.176471)$ does not improve the novice's decision inaccuracy, from the case of No Test, but still improves the novice's belief accuracy and the expert's chance of being trusted. With No Test, the novice will only commit false

negative error (i.e., failing to trust real expert) in his decision. The symmetric test, so to speak, converts certain false negative error to false positive error and thereby increases the expert's chance of being trusted. Although the novice ends up making the same number of errors (but of different types) with the symmetric test, the test nevertheless improves the novice's belief accuracy.

In general, any non-trivial tests, that is any test that has a non-zero correlation with the expert's competency, would improve the novice's belief inaccuracy. No test will diminish the novice's epistemic performance, either in terms of belief accuracy or of decision accuracy. This point is proven under general conditions in Kamenica and Gentzkow (2011) and can easily be established for my particular model as follows.

Let $0 < p_0 < 1$.

Without testing, the novice's expected belief inaccuracy is $E[(\omega - 0)^2] = p_0 * (1 - p_0)$.

With testing, the novice's expected belief inaccuracy is $E\left[\left(\omega - P(\omega = 1|S)\right)^2\right]$.

It is rational for the novice to update on the revealed information if

$$E\left[\left(\omega - P(\omega = 1|S)\right)^2\right] \geq E[(\omega - 0)^2].$$

A straightforward calculation would reveal that this inequality always holds, and the equal sign obtains if and only if $\epsilon_1 = 1 - \epsilon_2$. The formulation and verification of this inequality can be found in the supplementary Mathematica file.

When $\epsilon_1 = 1 - \epsilon_2$, the test does not distinguish between $\omega = 0$ and $\omega = 1$. No matter what ω is, the test would respond in the same way. Namely, the test will return $Comp$ with probability ϵ_1, which is equal to $1 - \epsilon_2$, and return $Incomp$ with probability $1 - \epsilon_1$, which is equal to ϵ_2, regardless of what ω is. Such a test contains no information about the value of ω and thus the novice's belief would not change after observing the result of such a test. For any test,

where $\epsilon_1 \neq 1 - \epsilon_2$, the novice's belief accuracy would strictly improve after observing the test result, as it contains some information about ω.

As to the novice's decision accuracy, whenever there is a separating equilibrium, observing the test result would improve the novice's decision accuracy. Sometimes, although the test improves the novice's belief accuracy by delivering new information, the new information may not be enough to move the novice's decision into a separating equilibrium. However, the novice generally benefits from learning new true information in the measurement of belief or decision accuracy.

1.4.2 Appendix 2. The Second Model

Solving the second model is, in fact, more intuitive than the first one. Like before, the expert knows the novice's threshold of trust and chooses the testing standard x to maximize her chance of being trusted. No matter what value x takes, failing the test S_x will not raise the novice's credence in $\omega = 1$. Unless the novice's prior already exceeds the threshold at the first place, the novice will always decide not to trust the expert upon observing $S_x = 0$, regardless of the value of x. For this reason, there is no x such that the novice would always choose to trust the expert regardless of the testing result. The best the expert can hope for, like in the first model, is that the novice will decide to trust her whenever she passes the test. When the value of x is relatively low, observing that $S_x = 1$ would not raise the novice's credence beyond the threshold t. We may say that a testing standard x differentiates the novice's reaction if x is large enough so that the observation of $S_x = 1$ would raise the novice's credence beyond the threshold t. If x is already large enough to differentiate the novice's reaction, further increasing x will not change the

novice's decision upon observing $S_x = 1$, despite that it does further increase the novice's posterior belief. Meanwhile, the chance of passing the test decreases as the testing standard x increases. Therefore, the expert wants to find the minimal x that will differentiate the novice's reaction to maximize her chance of being trusted.

Figure 2 represents the relation between the testing standard x and the novice's decision accuracy, given parameters specified above. When x is small, the novice would choose not to trust the expert regardless the testing result. Thus, his decision inaccuracy is 0.3 in the corresponding area on the left of the diagram, for he fails to trust all and only competent experts. On the right end of the diagram, as the testing standard increases to 1, the probability of $S_x = 1$ decreases to zero. The novice's decision inaccuracy thus increases back to 0.3, as eventually no one will pass the test and the novice will not trust anyone. In between, lies the optimal choice $x^* = 0.5$ that minimizes the novice's decision inaccuracy, where the novice makes mistakes with probability ~0.138749.

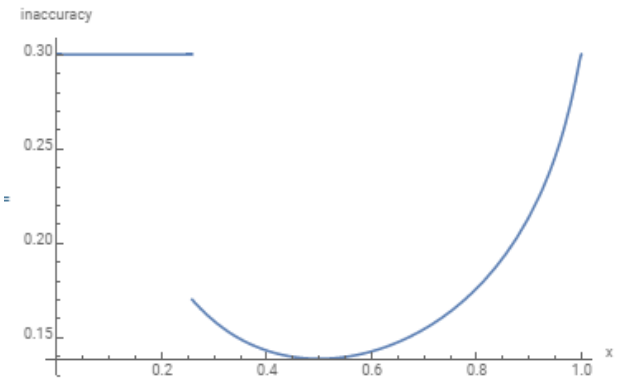

Figure 2: x and novice's decision inaccuracy

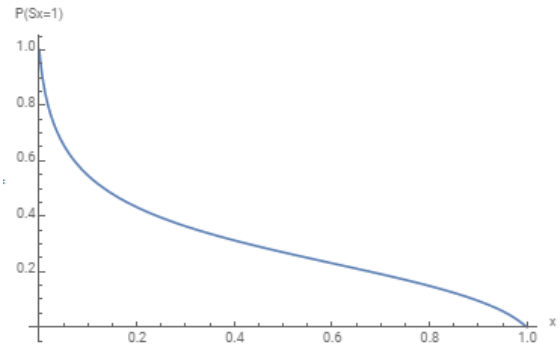

Figure 3: x and the probability of $S_x = 1$

Figure 3 shows the relation between the testing standard x and the expert's chance of passing the test. Not surprisingly, the expert's chance of passing monotonically decreases from 1 to 0 as x increase from 0 to 1. When x is large enough to differentiate the novice's reaction, the expert's chance of being trusted equals to the

probability of passing the test. Thus, the best testing standard x from the expert's point of view is the minimal x that would differentiate the novice's reaction, which is ~0.257734 ($x^{\dagger}$) in our example. Notice that $x^{\dagger}$, which is optimal from the expert's point of view is lower than x^{*}, which minimizes the novice's decision inaccuracy. The novice's decision inaccuracy under $x^{\dagger}$ is 0.170179, worse than 0.138749 under x^{*}.

When we inhibit the expert from choosing $x < x^{*}$ with a regulation on the minimal testing standard, the expert would choose x^{*} to maximize her own interest. This point can be seen from Figure 2, where the chance of seeing the good signal ($S_x = 1$) decreases as x increases. This holds in general, so long as $x^{\dagger} \leq x^{*}$.

If the minimal testing standard is set at x^{*}, the expert will always choose x^{*}, because $x^{*} \geq x^{\dagger}$ except in degenerate cases. By degenerate cases, I mean cases where no value of x would differentiate the novice's reaction and thus x^{*} and $x^{\dagger}$ are not well-defined. Such cases occur when there is too much noise in the test. In any non-degenerate cases, where x^{*} and $x^{\dagger}$ are well defined, x^{*}, which by definition minimizes the novice's decision inaccuracy, must differentiate the novice's reaction. On the other hand, $x^{\dagger}$, by definition, is the minimal value of x that differentiates the novice's reaction. Therefore, $x^{*} \geq x^{\dagger}$ in all non-degenerate cases. For this reason, when we set the minimal testing standard at x^{*} so that the expert can only choose all and only $x \geq x^{*}$, the expert, motivated by her self-interest, would choose x^{*} to maximize her chance of being trusted. The novice's epistemic performance, measured by his decision accuracy, thus gets maximized. This justifies why the regulation with minimal testing standard optimizes the novice's epistemic performance.

With some different choice of parameters, the novice's interest and the expert's interest can accidentally coincide so that $x^{\dagger} = x^{*}$. Here I give an example to help better understanding the model.

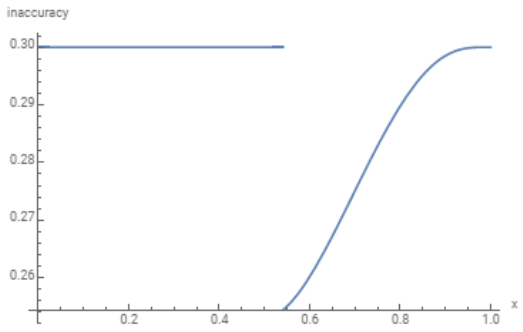

Figure 4: x and the novice's decision accuracy, with $\sigma = 1$ and $t = 2/3$

Let us change the distribution of ϵ to the standard normal distribution $N(0,1)$ and leave all other parameters intact as in the example in the main text. Figure A2.1 on the left shows the relation between the testing standard x and the novice's decision inaccuracy. In this case, $x^{\dagger} = x^{*} = 0.541341$. At that point, the novice's decision inaccuracy obtains its minimal value of 0.254406 and the expert's chance of being trusted obtains its maximal value of 0.136782.

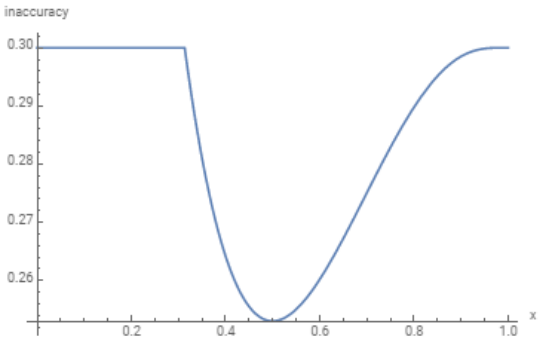

Figure 5: x and the novice's decision accuracy, with $\sigma = 1$ and $t = 1/2$

This coincidence results from the novice's relatively high threshold of trust $t = 2/3$. If we let $t = 1/2$, the relation between the novice's decision accuracy and the testing standard x will be as shown in the graph on the right. The novice's decision accuracy is maximized by $x^{*} = 0.5$, while the expert's chance of being trusted is maximized by $x^{\dagger} = 0.313426$.

When the novice adopts a higher threshold of trust $t = \frac{2}{3}$, the testing standard $x = 0.5$, which would minimize his deicision inaccuracy should he choose $t = 1/2$, is not large enough to differentiate his reaction. The minimal differentiating standard $x^{\dagger}$ thus lands on the right of $x =$

0.5. As a consequence, any $x \geq x^{\dagger}$ would further decrease the novice's decision accuracy. Setting the novice's threshold of trust aside, it shouldn't be suprisingly that in both this example, where $\epsilon \sim N(0,1)$, and the example in the main text, where $\epsilon \sim N(0, 0.5)$, the optimal testing standard from the novice's perspective is 0.5. This is an artifact of our parameter choices that $\omega \in \{0,1\}$ and that ϵ is symmetrically distributed. There is no reason to think that $x = 0.5$ has any special significance, since there is no good reason to think the error must be normally or symmetrically distributed or that ω must be either 0 or 1.

1.4.3 Appendix 3. The Third Model

Proposition. At Nash equilibrium, the expert chooses (x, θ) that minimizes the novice's non-disclosure posterior P^N.

Proof. Suppose the expert chooses (x_E, θ_E) at equilibrium. When the expert discloses her testing result, i.e. $\theta_E(S_{x_E}) = 1$, the novice would updates his belief, denoted by $P^D(x_E, S_{x_E})$, to $P(\omega = 1|S_{x_E})$ after he learns both the testing standard x and the test result S_x as in the previous model. When the expert does not disclose the result, the novice must choose his posterior belief P^N in react to the null signal. Given the expert's choice of (x_E, θ_E) known to the novice, the novice would choose $P_E^N = P(\omega = 1|\theta_E(S_{x_E}) = 0)$. In this case, the expectation of the novice's posterior is $E\left[\theta_E(S_{x_E})P^D(x_E, S_{x_E}) + \left(1 - \theta_E(S_{x_E})\right)P_E^N\right]$, which is equal to $E\left[\theta_E(S_{x_E})P^D(x_E, S_{x_E}) + \left(1 - \theta_E(S_{x_E})\right)P(\omega = 1|\theta_E(S_{x_E}) = 0)\right]$.

Now consider an arbitrary deviation (x', θ'). I claim that the deviation is not profitable if and only if $P(\omega = 1|\theta'(S_{x'}) = 0) \geq P(\omega = 1|\theta_E(S_{x_E}) = 0)$. This means that the expert's equilibrium strategy (x_E, θ_E) must minimize the novice's non-disclosure posterior $P_E^N = P(\omega =$

$1|\theta_E(S_{x_E}) = 0)$. The crucial step here is that the concept of Nash equilibrium requires the novice's reaction to the null signal, which is part of his strategy, to be held fixed, when we assess if the expert benefits from deviating to (x', θ'), For this reason, when the expert deviates to (x', θ'), in the case of non-disclosure, the novice's posterior belief would be $P_E^N = P(\omega = 1|\theta_E(S_{x_E}) = 0)$, but not $P(\omega = 1|\theta'(S_{x'}) = 0)$. Thus, the deviation is only profitable if $P_E^N = P(\omega = 1|\theta_E(S_{x_E}) = 0) > P(\omega = 1|\theta'(S_{x'}) = 0)$.

A proper proof can be achieved by comparing the following four expectations.

$$(a) \qquad E\left[\theta_E(S_{x_E})P^D(x_E, S_{x_E}) + \left(1 - \theta_E(S_{x_E})\right)P_E^N\right]$$

$$(b) \qquad E\left[\theta_E(S_{x_E})P^D(x_E, S_{x_E}) + \left(1 - \theta_E(S_{x_E})\right)P(\omega = 1|\theta_E(S_{x_E}) = 0)\right]$$

$$(c) \qquad E\left[\theta'(S_{x'})P^D(x', S_{x'}) + \left(1 - \theta'(S_{x'})\right)P_E^N\right]$$

$$(d) \qquad E\left[\theta'(S_{x'})P^D(x', S_{x'}) + \left(1 - \theta'(S_{x'})\right)P(\omega = 1|\theta'(S_{x'}) = 0)\right]$$

(a) is, by definition, the expectation of the novice's posterior. $(b) = (a)$, since at equilibrium, the novice knows the expert's strategy (x_E, θ_E) and reacts correctly to the expert's null signal with $P_E^N = P(\omega = 1|\theta_E(S_{x_E}) = 0)$. (c) is, by the definition of Nash equilibrium, the expectation of the novice's posterior if the expert deviates to (x', θ').

(d) is the expectation of the novice's posterior, should he know that the expert chooses (x', θ'). We use (d) as a middle term to compare (a) and (c) so that we can decide when a deviation is profitable. Recall that the martingale property requires that the expectation of posterior must equal to the prior, regardless of the choice of test. Thus, $(b) = (d)$. Given $(b) = (a)$, we have $(a) = (d)$. $(c) > (d)$ if and only if $P_E^N > P(\omega = 1|\theta'(S_{x'}) = 0)$. Thus, $(c) > (a)$ if and only if $P_E^N = P(\omega = 1|\theta_E(S_{x_E}) = 0) > P(\omega = 1|\theta'(S_{x'}) = 0)$. That is, the deviation is

profitable if and only if $P(\omega = 1|\theta_E(S_{x_E}) = 0) > P(\omega = 1|\theta'(S_{x'}) = 0)$. For (x_E, θ_E) to be the expert's equilibrium strategy, there cannot be any profitable deviation available to the expert. This means that for all (x', θ'), $P(\omega = 1|\theta'(S_{x'}) = 0) \leq P(\omega = 1|\theta_E(S_{x_E}) = 0)$. That is, at equilibrium, the expert chooses (x, θ) to minimize the novice's non-disclosure posterior $P^N = P(\omega = 1|\theta(S_x) = 0)$. Q.E.D.

This proposition allows us to find the Nash equilibrium of the scenario by solving the optimization problem $\min_{(x,\theta)} P(\omega = 1|\theta(S_x) = 0)$, which can be further simplified by separating the search of x and θ. For a given testing standard x, we first find the θ that solves $\min_{\theta} P(\omega = 1|\theta(S_x) = 0)$. Let us denote this value by P_x^N. Then we search across all values of x to find the x that solves $\min_{x} P_x^N$. This iterated procedure leads us to the Nash equilibrium (x, θ), which is unique for being a solution of an optimization problem.[11]

The procedure further simplifies for our example. Recall that θ is a map from $\{\emptyset, 0, 1\}$ to $\{0,1\}$ and that $\theta(\emptyset) = 0$. There are, thus, only four different disclosing strategies available to the expert. It is easy to check for any $x \in (0,1)$, the θ that maps 0 and $\emptyset$ to 0, 1 to 1 minimizes P_x^N. Thus, we can hold θ fixed and search across different values of x to find the Nash equilibrium.

[11] See DeMarzo, Kremer and Skrzypacz (2019, p. 2182, corollary 2) for details about the uniqueness result.

2.0 Voting Norms in Condorcet Jury Theorem: What We Owe to Other Voters

2.1 Introduction

Over the past twenty years, philosophers have become increasingly interested in the epistemic properties of social institutions, in particular, of democracy.[12] One may view this movement of epistemic democracy in philosophy as downstream from the 20th century information revolution that happened in economics, when information economics models were widely adopted to study social institutions. One most celebrated result from the information revolution is the rediscovery and the generalization of the Condorcet Jury Theorem (the CJT).

While this development is celebrated in philosophy and in economics alike, philosophers and economists turn out to have distinct takes on some of the basic assumptions of the CJT, namely, the Sincerity Assumption. Recently, Goodin & Spiekermann (2018) gave an extensive defense for the classic version of the CJT, including the Sincerity Assumption, whereas economists have moved away from it in favor of pivotal voting (viz. a type of strategic voting to be defined in Section 3.2) since Austen-Smith & Banks (1996). Why would philosophers, who are well aware of and often get inspired by the relevant development in economics, decide to fly in the face of it?

Several responses have been given in the past, all focusing on the empirical adequacy of pivotal voting.[13] I argue such responses are wrong-headed. Instead, I propose a moral argument

---

[12] For example, see Anderson 2006, Estlund 2008, Zagzebski 2012, Goodin & Spiekermann 2018.

[13] Previous arguments against pivotal voting to be discussed in this paper include Brennan & Lomasky (1997, Chapter 4), Dunleavy (1997) and Goodin & Spiekermann (2018, Chapter 4.3).

against pivotal voting that justifies philosophers' disagreement with economists. In a nutshell, a pivotal voter fails what is morally required of participants of democracy (i.e., citizens) by treating others as mere obstacles, i.e., parts of the environment that they must navigate through to optimize their own benefit. This failure comes from the direct projection of behavioral norms from the economic sphere to the political sphere.

My argument against pivotal voting is distinct from previous ones for being moral and hence normative in nature. While a normative argument can hardly move an economist who is in search of an empirically adequate descriptive/positive theory of voting behaviors, I argue that it should be of central interest to philosophers. It follows that the Sincerity Assumption should be understood as a normative requirement, instead of a realistic description of voters. This makes sincere voting a normative implication of the CJT, along with other better known normative implications, e.g., universal suffrage.

Some background must be covered before my main point can be addressed. In Section 2, I provide a brief review of the classic CJT, followed by a discussion of its normative implications. In Section 3, I define and discuss two types of strategic voting: second-best voting and pivotal voting. In Section 4, I present economists' arguments for why they prefer pivotal voting over sincere voting in the context of the CJT. While these technical backgrounds may strike both economists and philosophers as superfluous for different reasons, it is necessary to put the two in conversation.

In Section 5.1, I review previous arguments against pivotal voting and say why I think that they all fail to justify philosophers' disagreement with economists. I then argue for the relevance of a moral and thus normative argument and explain what it implies for interpreting CJT-type

results, in Section 5.2. Following that, I present my moral argument against pivotal voting in Section 5.3.

In Section 6, I allow voters in the CJT setup to have heterogeneous preference, representing their fundamental differences in values. While this case is important in its own right, it helps deliver a similar, perhaps more devastating, moral objection against sincere voting. I give a response in defense of sincere voting regarding the moral objection, which reveals the limits of the extensionalist formulations of voting used in CJT-type results.

I end the paper with some general big-picture remarks in Section 7.

2.2 The Classic CJT Framework

In the classic CJT framework, a group of n voters vote to make a choice between two alternatives, $\{L, R\}$, using the majority rule. For simplicity, let's assume that n is odd and thus there will not be any tie.

Between the two alternatives, L, R, one alternative is objectively better than the other, for every voter in the election. We may call this objectively better alternative the "correct alternative". The correct alternative varies across different possible states of the world. Again, for simplicity, we may assume that there are only two possible states of world, $\{s_L, s_R\}$. For each possible state of the world, one of the two alternatives is objectively better. Assume that it's common knowledge among voters that alternative L is objectively better in state s_L and alternative R is objectively better in state s_R. The voters are, however, uncertain about what the actual state of the world is. Each voter is expected to cast a vote for one of the two alternatives, and whichever alternative gets the most votes will get elected.

Two features of this setup are worth highlighting. First, all voters in this setup have identical values. In any state of the world, whatever is the best for anyone is the best for everyone. The potential disagreement among voters is purely epistemic, regarding what the correct alternative is, not what 'correct' means. Second, the uncertainty concerning the correct alternative comes from the uncertainty about what the actual state of the world is alone. Once voters agree on which state is actual, there will be no disagreement left over which candidate is correct. That is, for any given state, there is a unique correct alternative that is commonly known among voters.[14]

The classic Condorcet Jury Theorem follows from the following three assumptions (Goodin & Spiekermann, 2018, pp. 17-20):

> **Competence**: Each voter's belief about the correct alternative is true with probability $p_c > 0.5$ (and this holds for both states and is the same for all voters).

For simplicity, we assume that everyone is competent ($p_c > 0.5$) and equally competent (at the same level p_c). This is apparently unrealistic and can be relaxed in many ways. But this simple version is adequate for our purpose.

> **Sincerity**: All voters vote for the alternative they believe to be the correct alternative.

---

[14] The second feature is in place for the convenience of our discussion of heterogeneous preferences in Section 6. One may want to challenge this formulation, on the ground that voters are often uncertain about which candidate is correct for a given state, in addition to being uncertain about which state is actual. From there, one may want to make the distinction between "state uncertainty" and "candidate uncertainty". While this point is intuitive in certain ways, it can be accommodated by fine-graining states so that voters always know which candidate is correct in each state. In Section 6, I will allow voters to disagree which candidate is correct in a particular state. But such disagreement comes from the heterogeneity of values among voters and thus is not epistemic.

The Sincerity Assumption will be the focus of this paper.

> **Independence**: The beliefs of all voters are statistically independent, given the true state of the world regarding the correct alternative.

Like the other two assumptions, the Independence Assumption has been extensively criticized. However, it will not be discussed in this paper.

From these three assumptions, two results, both known as the Condorcet Jury Theorem, are derived.

> **Non-asymptotic Result**: For any given level of competency $p_c \in (0.5,1)$ and any given electorate of size n, if we add two (or more) voters to the electorate, the resulting larger electorate will be more likely to choose the correct alternative than the previous smaller electorate.

This result suggests that a larger electorate will do better at making decisions and is often cited as a justification for universal suffrage. The asymptotic result alone, however, is not sufficient to provide an epistemic justification for democracy. It only shows that larger electorates make better decisions than smaller ones. It doesn't follow that elections make good decisions or decisions that are better than alternative decision mechanisms.

> **Asymptotic Result**: When the number of voters goes to infinity, the correct alternative always gets elected.

The asymptotic result is supposed to show the wisdom of the crowd and serve as an epistemic justification for democracy. However, reading literally, it doesn't have much normative import at all, since we will never have infinite voters. What provides the normative import is what I call "the approximate result".

Approximate Result: When the number of voters is large enough, the correct alternative gets elected with very high probability.

The approximate result, as its name suggests, is approximate and doesn't bear the kind of mathematical rigor the first two results have. This is why this result is typically not included as part of the Condorcet Jury Theorem in the existing literature. However, I think that this is what people have in mind when they say that the Condorcet Jury Theorem provides an epistemic justification for democracy.

A more precise statement of the approximate result is as follows:

For any level of accuracy $P \in (0,1)$, given a competence level $p_c \in (0.5, 1)$, there is a number $n > 0$ such that a group of size n (or larger) will elect the correct alternative with probability higher than P.

The argument for the epistemic justification of democracy, based on the approximate result, goes as follows.

(1) The setup applies, and the assumptions hold.

(2) There is a level of accuracy $P \in (0,1)$ such that electing the correct alternative with probability P is sufficient to show that democracy has an epistemic advantage.

(3 the Approximate Result) For any level of accuracy $P \in (0,1)$, given a competence level $p_c \in (0.5, 1)$, there is a number $n > 0$ such that a group of size n (or larger) will elect the correct alternative with probability higher than P.

(4) Given the competence level p_c that holds among voters, the number of voters in our world is greater than or equal to (or at least can be greater than or equal to) the number of voters required to achieve the desired level of accuracy P.

(Conclusion) Democracy has an epistemic advantage.

All three assumptions, as well as the setup, have been criticized extensively in the past. The main disagreement between the philosophers and the economists, however, concerns the Sincerity Assumption. Goodin & Spiekermann (2018) discussed various ways to weaken the Competence Assumption (Chapter 3.1) and the Independence Assumption (Chapter 5). But, when it comes to the Sincerity Assumption, which economists abandoned without hesitation, Goodin and Spiekermann decide to stick with its original form. To better understand this disagreement, we must introduce alternative ways to vote in addition to sincere voting, namely, strategic voting.

2.3 Two Varieties of Strategic Voting

While CJT assumes that people vote sincerely by faithfully reporting their true preferences, people often fail to do so for various reasons. When people misrepresent their preferences in voting, we may say that they voted *insincerely* or *strategically*. The word 'strategic voting' has been used in several different ways in the existing literature and in colloquial conversations. Instead of following their colloquial uses, I define 'sincere', 'insincere' and 'strategic' as technical terms as follows. According to my definition, 'strategic' is synonymous with 'insincere', referring to misrepresenting the voters' true preference in general.[15]

It's worth first distinguishing between voting *behaviors* and voting *strategies*. A voting behavior refers to the actual or projected action of a voter in casting a ballet, while a voting strategy

[15] `insincere' and `strategic' are not synonymous in modern colloquial English. `insincere' has a normative charge that is not always present in `strategic'. In this article, I use `sincere', `insincere' and `strategic' as technical terms, based on the definition provided in the text.

is a plan of actions for different scenarios. Formally, a voting strategy is a function that maps voting scenarios to voting behaviors.

A voting behavior is *sincere* if it reports one's true preference faithfully. That is, the ballot casted coincides with one's true preference.[16] A voting behavior is *strategic* (or equivalently, *insincere*) if it is not sincere. This definition of sincere voting behavior is extensional in the sense that it only turns on whether the ballot and the preference coincide. The potentially associated motivations, intentions, reasons, etc., is left out of the definition on purpose.

A voting strategy is *sincere* if it maps all voting scenarios to a sincere voting behavior. That is to say, a voter who follows a sincere voting strategy reports her true preference in all voting scenarios. Clearly, there is only one voting strategy that is sincere. We shall call it "sincere voting". A voter who adopts sincere voting as her voting strategy is a sincere voter.

Any voting strategy that is not sincere is *strategic*. Equivalently, a voting strategy is strategic if it leads to insincere voting behavior in *some* scenarios. There are potentially infinitely many different voting strategies that are strategic. Adopting any one of them makes one a "strategic voter". Strategic voting strategies, as defined, are intensional.

To illustrate my definitions, let's consider a somewhat absurd example. Suppose that there are two candidates in an election, Candidate 1 and Candidate 2. One can adopt the following voting strategy: if one saw odd numbers of red cars on her way to the voting booth, she will vote for Candidate 1, and she will vote for Candidate 2 otherwise. Assume that the voter does have a

[16] This definition is extensive on purpose. Whether a voting behavior is sincere or not only depends on if the ballot coincides with the voter's true preference. The motivations, intentions, reasons, etc., behind a behavior is left out of the definition on purpose.

preference over the two candidates independent of the number of red cars seen on the way to the voting booth.[17] Let's say it's Candidate 1. Then the above-mentioned strategy is strategic, even though it may not necessarily lead to an insincere (i.e., strategic) voting behavior. For example, the voter could happen to see 5 red cars on her way and ends up voting for her true preference, Candidate 1. But the mere fact that the resulting voting behavior is sincere does not make the adopted voting strategy sincere. This voter is, by my definition, a strategic voter. This example also shows that the technical definition of 'strategic' I use is against its colloquial meaning. A strategic voter, per my definition, doesn't have to cast her vote in any deliberate or manipulative way.

Voting strategies, not voting behaviors, take the central stage of this paper. Throughout the paper, I use phrases like "strategic voting" as a shorthand for "strategic voting strategy", since the latter phrase is cumbersome. In the following, I introduce two types of strategic voting (i.e., two voting strategies that are strategic): second-best voting and pivotal voting.

2.3.1 **Second-Best Voting**

What I call "second-best voting" is the best known and most discussed kind of strategic voting. It is traditionally discussed under a setup distinct from the classic CJT. It assumes that (1) voters have heterogenous preferences, and (2) there are at least three candidates. Although second-best voting is *not* the focus of this paper, it's worthwhile to have it in view, for a better general

[17] Throughout the paper, we assume that the voter's preferences are invariant across scenarios.

understanding of strategic voting, as well as a more realistic illustration of the definitions I gave above.

Suppose that there are three candidates in an election: A, B and C. Every voter casts a single vote for one candidate and whoever gets the most votes gets elected (i.e., the plurality rule). Suppose that my true preference over the three candidates is A > B > C. If I were to vote sincerely, I would vote for A.

Suppose that there are some shared beliefs among voters about the distribution of all voters' preferences. It's commonly believed that 40% of voters have preference C > B > A, 35% B > A > C, and 25% A > B > C. If everyone were to vote sincerely, C, my least preferred candidate, would get elected. In this case, 60% of all voters, including me, will become deeply disappointed, since C is their least preferred candidate. Facing this reality, those with preference B > A > C ("B-voters") and those with preference A > B > C ("A-voters") may come together and vote for B, to prevent C from being elected. B-voters may tell me (an A-voter) the following: "There is no way that A is going to win the election. You should vote for B instead to prevent C from winning."

The belief that sincere voting will lead to worst outcome gives me *prima facie* reasons[18] to misrepresent my preference and vote for my second-best option B instead. Clearly, there is nothing magical about having three candidates or voting for the *second*-best option. If there are more than

[18] When one decides what to do, there are often various reason for competing actions in presence. While the belief that sincere voting will lead to worst outcome for me gives me a reason to vote for B, I may have other reasons not to do so. For example, I may consider voting as an opportunity to express my true preference to other voters, and thus care more about expressing than the outcome of the election. This gives me a reason to vote for A. These two conflicting reasons are both *prima facie* in the sense that neither of them decides my final action by itself. The final decision is made with all *prima facie* reasons evaluated and compared together.

three candidates, one may similarly have reasons to vote for the nth-best candidate to prevent even less preferred candidate from being elected. For convenience, we refer to all cases by "*second-best voting*" where one misrepresents her true preference to get certain less preferred yet more promising (a.k.a., electable) candidate elected, when she finds her top preferences unlikely to win.

This example showcases why I emphasize the distinction between voting strategies and voting behaviors as defined earlier. It's difficult to characterize second-best voting by voting behaviors alone, unless we include the voter's intention as part of her voting behavior. However, we can easily capture the essence of second-best voting as a voting strategy, where the voter always votes for the less preferred yet more promising candidate, when she finds her top preference unlikely to win. The intensional structure of voting strategy (at least partly) captures the otherwise elusive intentional content of voting.

Like in our previous red-car example, a voter following the second-best voting strategy may end up conducting a sincere voting behavior if she finds it likely for her most preferred candidate to win. But this doesn't make her a sincere voter, for a sincere voter always votes sincerely.

Second-best voting is, strictly speaking, irrelevant to the classic CJT since the two assumptions it relies on do not hold. But it can appear in certain extensions of the classic CJT. The first assumption can be met by recasting the story into an epistemic version. Instead of having heterogenous preference over candidates, all voters have the same preference as in the CJT setup in the sense that they all want to elect the correct candidate. The difference comes in beliefs. 40% of voters believe that C is more likely to be the correct candidate than B, who is more likely than A (C > B > A), etc. Being in the minority, I think that I know better than others and choose to vote strategically, not to promote my own selfish interests, but to promote the best interests of everyone.

The second assumption does fail in the classic CJT setup but will hold in certain multi-candidate extensions. [19]

If second-best voting is the only kind of strategic voting, there will be no alternative to sincere voting in the classic CJT setup and, thus, the Sincerity assumption will be justified on the ground of a lack of alternatives. Of course, this is not true. People misrepresent their true preferences for all kinds of reasons. Among these, we shall focus on what is known as "sophisticated voting" in economics. To avoid the positive connotation in the word 'sophisticated', I shall call it "pivotal voting" instead. Unlike second-best voting, pivotal voting does fit squarely with the classic CJT setup.

2.3.2 **Pivotal Voting**

Let's go back to the classic CJT setup. Suppose that Alice, Bob, and Carol are to vote between two options $\{L, R\}$. There are two possible states of the world $\{s_l, s_r\}$. L is the correct choice in state s_l, and R is correct in s_r. Voters are uncertain about which state is actual. All voters have the same preference in the sense that they all want the correct option of the actual state to be elected. Should they know which state is actual, they will unanimously elect the correct candidate. Unfortunately, they do not know which state is actual. Instead, they each receive (an independent realization of) an informative private signal $\sigma \in \{\sigma_l, \sigma_r\}$ about the actual state of world. This signal allows each of them to form a belief about which state is actual.

---

[19] See List & Goodin (2002) for a multiple-candidate (≥ 3) version of the CJT. Some of our later discussion over pivotal voting can be applied to second-best voting in the context of multiple-candidate CJT. I shall not attempt that in this paper.

The signal σ is informative in the sense that its distribution is correlated with the actual state of the world. If s_l is the actual state, they will more likely than not to receive σ_l. Similarly, if s_r is the actual state, they will more likely receive σ_r. Suppose that it's common knowledge that their signals are drawn from the same distribution. This makes them epistemic peers for having the same epistemic capacity. Still, they may end up forming different beliefs about the actual state, since they each receive an independent realization of the signal. We expect that each of them will form a correct belief with probability > 0.5, although they may receive a false signal and form a false belief by chance. Thus, the Competence Assumption is met. Let's assume that the other two assumptions also obtain, i.e., they all vote independently and sincerely based on the signal they received.

Let's take Bob's individual perspective to see why he may have reasons to misrepresent his true preference. Suppose that Bob received the signal σ_l. The first thing to notice is that Bob's vote will only make a difference to the outcome when there is a tie between Alice and Carol. If Alice and Carol agree, Bob's vote simply wouldn't make a difference. Thus, Bob only needs to think about what he should do when Alice and Carol disagree. Since Alice and Carol voted sincerely, a disagreement will only occur when one of them received σ_l, and another received σ_r. With these reflections, it's as if another piece of information becomes available to Bob, namely, one of Alice and Carol received σ_l, and another σ_r. This gives Bob reasons to cast his vote as if he has received three independent signals, $\{\sigma_l, \sigma_l, \sigma_r\}$, instead of one. This inferred piece of information could make a difference to Bob's choice.

Let's make an additional assumption to illustrate the point. Suppose that the signal σ is highly accurate in s_l but is not as accurate in s_r.[20] It's highly unlikely for one to receive a false signal (σ_r) if s_l is actual. In this case, the signal pattern $\{\sigma_l, \sigma_l, \sigma_r\}$ is much more likely to obtain in s_r than in s_l. The fact that someone observed σ_r will convince Bob that s_r is the actual state of the world, even though the other two voters, including himself, received σ_l. From Bob's perspective, it will be to everyone's benefit, if he votes for *R* instead. It's worth noting that this case doesn't violate the Competency Assumption, as the signal is (equally) informative for all voters with more than 50% accuracy.

Let's use some concrete numbers to make the point clearer. Suppose that Alice, Bob, and Carol all have the same prior belief that s_l and s_r are equally likely, i.e., $P(s_l) = P(s_r) = 0.5$, before they observe their own private signal. The signal σ is independent and identically distributed for all voters. The accuracy (or quality) of σ is much higher in s_l than in s_r. Let's say that $P(\sigma_l|s_l) = 0.9$ and $P(\sigma_r|s_r) = 0.6$. Now let's compute Bob's posterior over s_l assuming that Bob received σ_l and that there is a tie between Alice and Carol, viz. that two σ_l's and one σ_r are received among the three of them.

$$P(\sigma_l) = P(\sigma_l|s_l)P(s_l) + P(\sigma_l|s_r)P(s_r) = 0.75$$

$$P(\sigma_r) = 1 - P(\sigma_l) = 0.25$$

[20] Experiments and resulting signals often face trade-offs between two types of errors: false positive and false negative. This trade-off corresponds to the difference in accuracy across different states. Take a rapid Covid test for example. A rapid test is highly accurate when the actual state is negative. If the patient is negative, it's highly unlikely that a rapid test will report positive. However, it's less accurate when the actual state is positive. A positive patient can get a negative result from a rapid test from time to time. This is why a positive result from a rapid test is very informative while a negative result doesn't mean too much.

$$P(s_l|\sigma_l,\sigma_l,\sigma_r) = \frac{P(\sigma_l,\sigma_l,\sigma_r|s_l)P(s_l)}{P(\sigma_l,\sigma_l,\sigma_r)} = \frac{\binom{2}{1}P(\sigma_l|s_l)P(\sigma_l|s_l)P(\sigma_r|s_l)P(s_l)}{\binom{2}{1}P(\sigma_l)P(\sigma_l)P(\sigma_r)}$$

$$= \frac{2 * 0.9 * 0.9 * 0.1 * 0.5}{2 * 0.75 * 0.75 * 0.25} = 0.288$$

Thus, even though Bob observed σ_l himself, he finds it highly unlikely that s_l is actual if there is a tie between Alice and Carol. When there is a tie, he would like to break it in favor of s_r, regardless of the signal he received.

The Competency Assumption is satisfied in this case. Since Bob originally finds s_l and s_r equally likely, sincere voting will make lead Bob to choose L (R) if and only if he receives σ_l (σ_r). In state s_l, Bob will be correct 90% of time, whereas he will be correct 60% of time in state s_r. On average, Bob expects to be correct 75% of time, which makes his vote better than a random guess.

It's worth noting that Bob is indeed misrepresenting his true preference (or his true belief) in this case. If you ask Bob who he thinks is the correct candidate when he casts his vote, Bob will say *L* without hesitation. After all, Bob only observed one signal, not three, and formed his belief based on that signal. Then you may ask Bob why he chooses to vote for *R* instead of *L*. Bob may say the following: "My vote will only make a difference when there is a tie. I do not expect a tie to happen. I know that both Alice and Carol are voting sincerely. If a tie happens in this case, I want to break the tie in the correct way by voting for *R* instead of *L*."

The way Bob casts his vote is known as "sophisticated voting" in the economics literature. The key to sophisticated voting is to vote as if you are breaking a tie, or in other words, that your vote is pivotal. For this reason, sophisticated voting may better be called "pivotal voting", to avoid the positive connotation in the word 'sophisticated'. The motivation for pivotal voting is simple. Your vote only makes a difference when it's pivotal, that is, when there is a tie among all other voters.

Like second-best voting, a pivotal voter may end up voting sincerely if it happens to break the tie in the right way. In our case, Bob will vote for R if his private signal is σ_r, not because he is a sincere voter, but because it's the right way to break the tie between Alice and Carol.

Unlike second-best voting, pivotal voting, as illustrated in our example, doesn't require that voters have heterogeneous preferences or that there are more than two candidates. One may think that the motivation behind second-best voting is the heterogeneity in preference among voters, while the motivation behind pivotal voting is the heterogeneity in beliefs. This makes pivotal voting fit squarely with the classic CJT setup, where voters only disagree on beliefs, not on values. Indeed, economists find it much more preferable than sincere voting, and proposed to replace the sincerity assumption with it in the CJT.

2.4 Economists' Case for Pivotal Voting in the CJT

The 20th century economic literature on Condorcet Jury Theorem started assuming that people vote sincerely. Early influential papers include Ladha (1992), Miller (1986), and Young (1988). However, economists soon reached a unanimous agreement to abandon sincere voting in favor of pivoting voting, following Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1997).

The main criticism against sincere voting in Austen-Smith and Banks (1996) is that sincere voting is not rational in the sense that it does not constitute a Nash equilibrium. I resist using the word "rational" for the property of constituting Nash equilibrium. While Nash equilibrium is the received solution concept in modern game theory, it doesn't make it the only conception of

rationality for human social interactions. Instead, I call a voting strategy *Nash* if a Nash equilibrium is obtained when *all* voters adopt that voting strategy.

Austen-Smith and Banks (1996) showed that sincere voting is not Nash under general conditions. Feddersen and Pesendorfer (1997) showed that pivotal voting is, in general, Nash. Going through either of the two results will take us far afield. But the general idea of both is already present in our example above.

A Nash equilibrium is a profile of players' strategies so that no player would benefit from deviating to a different strategy with the other player's strategy held fixed. In the example above, we saw that Bob is motivated to deviate from sincere voting when the other two voters are voting sincerely. Let's take another brief look at it.

Suppose Alice, Bob and Carol all vote sincerely in the example discussed above. This constitutes a strategy profile of three players, where each of them adopts the strategy of sincere voting, i.e., to vote according to their posterior beliefs about the actual state s after observing their own private signal σ. We have seen above that Bob (or anyone) has the incentive to adopt a different strategy, namely, pivotal voting, to improve the probability of electing the correct candidate in the case where there is a tie between Alice and Carol.

If there is no tie between Alice and Carol, whatever strategy Bob adopts has no influence on the outcome of the election. Suppose that there is a tie between Alice and Carol. Let's see how sincere voting and pivotal voting perform for Bob.

First, we calculate the distribution of the state assuming there is a tie.

$$P(s_l|\sigma_l,\sigma_r)=\frac{\binom{2}{1}P(\sigma_l,\sigma_r|s_l)P(s_l)}{\binom{2}{1}P(\sigma_l,\sigma_r)}=\frac{P(\sigma_l|s_l)P(\sigma_r|s_l)P(s_l)}{P(\sigma_l)P(\sigma_r)}=\frac{0.9*0.1*0.5}{0.75*0.25}=0.24$$

$$P(s_r|\sigma_l,\sigma_r)=1-P(s_l|\sigma_l,\sigma_r)=0.76$$

Suppose that Bob votes sincerely. Following the previous setup, $P(\sigma_l|s_l) = 0.9$ and $P(\sigma_r|s_r) = 0.6$. Thus, Bob will be right with probability 0.9 if s_l is actual and 0.6 if s_r is actual. This leads the overall expected accuracy to be $0.24 * 0.9 + 0.76 * 0.6 = 0.672$. That is, given that Alice and Carol vote sincerely and there is a tie between them, if Bob votes sincerely, the election will elect the correct candidate 67.2% of the time.

We have seen before that if there is a tie between Alice and Carol, Bob will find s_r much more likely regardless of his private signal. Thus, Bob will always vote for R if Bob adopts pivotal voting. Thus, the election will elect the correct candidate if and only if s_r is actual. This makes it 76% accurate, which is a significant improvement from sincere voting (67.2%). Unilaterally deviating from sincere voting to pivotal voting increases the accuracy of the election significantly in case that there is a tie between Alice and Carol. If there is no tie, Bob's strategy has no influence on the election. Therefore, unilateral deviation leads to a sure gain in some cases and no loss in any case. This gives Bob an incentive (which, as discussed before, is purely epistemic) to vote pivotally instead of sincerely and thus shows that sincere voting doesn't constitute a Nash equilibrium in this case.

While my example assumes a very specific probability distribution of s and σ across different states, Austen-Smith and Banks (1996, Theorem 2&3) showed that this assumption can be vastly generalized, and such phenomena obtain under fairly general conditions. On the other hand, it shouldn't be surprising that pivotal voting constitutes a Nash equilibrium (under certain reasonably weak background assumptions). The thinking pattern Bob goes through in our example is characteristic of Nash equilibrium. Bob holds others' strategies as fixed and elaborate over different scenarios to decide what should be done in each of them.

Feddersen & Pesendorfer (1997) extended CJT-type results to pivotal voting, showing that a large electorate of pivotal voters will always elect the correct candidate (p. 1042, Theorem 3). Pivotal voting, therefore, aggregates information as sincere voting does, without facing the charge of not constituting a Nash equilibrium. In fact, they add, pivotal voting outperforms sincere voting in information aggregation, as a large group of pivotal voters can elect the correct candidate even when a large group of sincere voters fail to do so (1997, p. 1044, Example 5.1). This leads to the unanimous abandonment of sincere voting among economists.

A characteristic case in favor of pivot voting is when the signals voters receive are *biased* toward one state, say s_l. In this case, no matter what the actual state is, a voter is more likely to receive σ_l than σ_r. That is, $P(\sigma_l|s_l) > P(\sigma_l|s_r) > 0.5$.The classic CJT fails in this case, as the Competence Assumption is systemically violated. A large electorate of sincere votes will elect *L* for sure regardless of the actual state. A large electorate of pivotal voters, however, will elect the correct candidate. In state s_r, a good portion of voters who received σ_l will end up voting for *R* after reflection, as Bob did in the example above.

The results of Feddersen and Pesendorfer (1997) are widely considered a modern extension of the classic CJT, and, indeed, a remarkable achievement in the information-aggregation approach to political institutions. The Sincerity Assumption is replaced by more realistic ones (at least by economist's standards); the Competence Assumption is relaxed; and a stronger result is delivered. Why wouldn't philosophers follow the lead?

2.5 What is Wrong of Pivotal Voting

In this section, I discuss what is wrong of pivotal voting from a philosopher's point of view. I first present three objections from the existing literature, including Brennan & Lomasky (1997, Chapter 4), Dunleavy (1997) and Goodin & Spiekermann (2018, Chapter 4.3). I argue that they are all wrong-headed in focusing on the empirical adequacy of pivotal voting and thus fail to justify the disagreement between philosophers and economists. In Section 5.2, I argue for the relevance of a moral and thus normative argument in this context. Finally, I present my moral argument against pivotal voting in Section 5.3.

2.5.1 Previous arguments against pivotal voting

In Chapter 4.3, Goodin & Spiekermann (2018) presented two arguments against strategic voting in general terms. [21] The first argument concerns second-best voting alone and thus will not be discussed here. The second argument, however, is explicitly addressed to Austen-Smith & Banks (1996). They contend that pivotal voting is exceedingly demanding on voters and thus fails to be realistic.

"Since strategic thinking provides no clear, unique alternative guide to behaviour, even for voters who try their best to engage in it, the default rule rules by default … It is probably wiser, at

[21] Goodin & Spiekermann (2018) used the term "strategic voting" in a way similar to this paper, referring to all insincere voting strategies. However, they didn't clearly distinguish different types of strategic voting as I did in this paper.

this point, to step back from complicated game-theoretic reasoning altogether, if we are to construct a credible model of voter behaviour in the real world." (p. 49)

It is true that pivotal voting is demanding on voters. It wasn't straightforward for us to figure out how Bob ought to vote in a pivotal manner in our three-voter example. It will only become more challenging, as the size of the electoral increases.[22]

Similar objections that challenge the empirical adequacy of pivotal voting have been raised in the past. Brennan & Lomasky (1997, Chapter 4) shows that the chance of being pivotal in a large electorate for any individual voter is both negligible and extremely sensitive to initial conditions. It's highly unrealistic that voters in the real world would make their decisions conditional on such tiny and elusive probabilities. Dunleavy (1997, pp 74-7) argues that in realistic voting scenarios, "being pivotal" can be interpreted in many different ways. Even in small electorates, different interpretations will lead to different models and different optimal voting strategies. Pivotal voting, thus, fails to provide clear guidance for voters in any reasonably realistic elections, in cases where, for example, people can choose to abstain, and the election can end up with ties.[23] For this reason, pivotal voting cannot be an accurate model of voter behaviors.

---

[22] Some recent literature on the computational complexity shows that second-best voting is very demanding computation-wise. For a recent review, see Veselova (2016). Little has been said about pivotal voting in the literature. However, Bayesian agents, which pivotal voting assumes, are notoriously demanding on computation.

[23] Dunleavy's point, as I read it, concerns the barriers individual voters face when practicing pivotal voting. There are well-defined coherent pivotal voting models where the election can end up in ties and people can choose to abstain. Dunleavy's point is that there are different ways to interpret what "pivotal" means for any reasonably realistic voting scenario. These different interpretations lead to different models for the same scenario and different models recommend different voting strategies. Once we have a model, like the one discussed earlier in this paper, what

What voters in the real world actually do can only be found out from empirical evidence. Conducting empirical studies on this matter turns out to be particularly challenging, because the true preference of voters cannot be observed. Limited existing studies all focus on second-best voting, and no consensus has been reached.[24] The lack of empirical evidence is partly why above-mentioned authors rely on theoretical arguments to dispute the empirical adequacy of pivotal voting.

I argue that all these three arguments, whether seemingly plausible or not, fail to justify philosophers' disagreement with economists. This failure does not come from a mere lack of empirical support. Rather, the real problem is that they focus on the empirical adequacy of pivotal voting alone and implicitly takes empirical evidence as the final tribunal for the dispute. If empirical adequacy is what matters after all, what makes philosophers think that they know better than economists about how people actually behave?

2.5.2 The relevance of a normative argument

The real reason why philosophers hesitate to let go of sincere voting, I think, is that they find sincere voting morally superior to pivotal voting. Even if people are perfectly capable of pivotal voting, many philosophers would argue against doing so, in spite of the economic reasons

---

"pivotal" means is precise and well-defined. However, the step from a realistic voting scenario to a model is not always transparent.

[24] Most recently, Spenkuch's work on second-best voting (2018, p.74) suggests that neither sincere voting nor second-best voting is empirically adequate. The study however does not concern pivotal voting at all.

in its favor discussed above (namely, more robust information aggregation results and being a Nash equilibrium).

Pursuing this line of thought requires us to read the Sincerity assumption (and its alternatives) as a normative claim instead of a descriptive one. Goodin & Spiekermann (2018) clearly read the Sincerity assumption as a descriptive claim and thereby takes CJT-type results as *non-ideal* theories. A non-ideal theory draws conclusions from certain *non-ideal* yet *realistic* assumptions. The focal non-ideal yet realistic assumption in this case is that people vote sincerely (i.e., the Sincerity assumption). However, I do not see this assumption as either non-ideal or realistic. In fact, sincere voting strikes many as ideal but non-realistic at a pre-theoretical level, quite opposite to what Goodin and Spiekermann claim.

A normative reading of the Sincerity assumption pushes CJT-type results more toward an ideal theory.[25] Under this reading, the classic CJT requires voters to vote sincerely for the information aggregation effect. Similarly, Feddersen and Pesendorfer's result requires voters to vote pivotally so that a more robust information aggregation effect can take place under a Nash equilibrium. This normative reading certainly leaves most economists unimpressed. After all, what they are after is an empirically adequate descriptive/positive theory of voting behaviors, but not a normative and ideal one.

However, this normative reading, I argue, is exactly what philosophers should care about. Recall my discussion of the normative implications of the classic CJT in Section 2. Take the best-

[25] CJT-type results rely on several assumptions in addition to the Sincerity assumption. I do not think that there is a clear cut between ideal theories and non-ideal theories. How ideal a theory is apparently depends on how realistic its assumptions are and comes in degree.

known implication of universal suffrage for example. The argument says that we should adopt universal suffrage so that we can enjoy the epistemic benefits shown by the non-asymptotic result. The argument takes the following form.

(1) If we adopt universal suffrage, assuming that other assumptions of the classic CJT hold, such-and-such epistemic benefits (as shown by the non-asymptotic result) will obtain.

(2) We want such-and-such epistemic benefits to obtain.

(Conclusion) We should adopt universal suffrage.

The exact same argument form applies to the Sincerity assumption and its alternatives.

(1) If everyone votes sincerely, assuming other assumptions of the classic CJT hold, such-and-such epistemic benefits (as shown by the classic CJT) will obtain.

(2) We want such-and-such epistemic benefits to obtain.

(Conclusion) Everyone should vote sincerely.

This, of course, is not an argument for sincere voting over pivotal voting, because the epistemic benefits from everyone voting pivotally are even greater, as shown by Feddersen & Pesendorfer (1997). Instead, this argument is meant to show how a normative reading of the Sincerity assumption should be understood.

Although sincere voting is shown to be inferior to pivotal voting along the epistemic dimension, I argue that it is superior in the moral dimension. Although pivotal voting can potentially provide a stronger epistemic justification for democracy based on the results of Feddersen & Pesendorfer (1997), it fails on moral grounds. Such disapproval is not uncommon. One can easily come up with examples where certain epistemically beneficial means are prohibited on moral grounds.

Anyone who cares about democracy as a whole, not just the epistemic dimension of it, should endorse the sincere voting based CJT-type results, for moral considerations. This does not stop the classic CJT from providing an epistemic justification for democracy.

In the following, I present a moral argument against pivotal voting in CJT-type results. In Section 6, I discuss if sincere voting is indeed morally superior by raising a similar moral objection against it in the presence of heterogenous preferences among voters.

2.5.3 A moral argument against pivotal voting

A better understanding of pivotal voting is needed before the normative argument can be delivered. In the following, I first give a more detailed analysis on the nature of pivotal voting (Section 5.3.1) and then deliver the moral argument (Section 5.3.2).

2.5.3.1 5.3.1 The Nature of Pivotal Voting

The greatest advantage economists see in pivotal voting is that it always leads to a Nash equilibrium (under reasonably weak background assumptions). This fact is barely surprising since the reasoning pivotal voters go through is an instance of Nash optimization, which is the cornerstone of the concept of Nash equilibrium. Clarifying the relations among pivotal voting, Nash optimization and Nash equilibrium will help clarify my moral criticism.

A Nash equilibrium is a profile of players' strategies such that no player will benefit from unilaterally deviating from her current strategy. A profile of strategies includes every player's strategy for action. For example, the profile we start our example with is for everyone to vote sincerely. This profile is not a Nash equilibrium, because, as discussed in section 4, Bob would

benefit from changing his strategy to pivotal voting unilaterally, assuming that Alice and Carol stick with sincere voting.

While Nash equilibrium concerns profiles of strategies, it imposes assumptions about how individuals behave. Each player is assumed to be a Nash agent who optimizes in a particularly Nash way. Namely, a Nash agent calibrates her own strategy to maximize her own utility holding others' strategies as given and fixed. We can call this kind of optimization "Nash optimization". Pivotal voting is a mere application of Nash optimization in the case of voting.

The concept of Nash equilibrium and Nash optimization was originally proposed as a solution concept for *non-cooperative* games. This point is emphasized at the very beginning of John Nash's original publication, which is titled "Non-Cooperative Games" (1951, p.286), as follows: "Our theory, in contradistinction, is based on the *absence* of coalitions in that it is assumed that each participant acts independently, without collaboration or communication with any of the others" (italics in the original).

Pivotal voting, taken as a voting norm, is projecting what is required of economic rationality in human non-cooperative interactions to voting, which is a characteristic type of interaction in the political sphere of human life. The political sphere, if anything, is cooperative.

2.5.3.2 5.3.2 The Moral Flaw of Pivotal Voting

Different spheres of human life are governed by different norms. In the economic sphere, people relate to each other as merchants and customers and are justified to treat others as part of the environment against which they optimize their own benefit, as described by Nash optimization. On the other hand, people ought to relate to each other as fellow citizens in the political sphere. Instead of taking others as part of the environment, they must recognize and respect others' equal citizenship in the democratic procedure to form a general will.

The main problem of pivotal voting understood as a voting norm, in a nutshell, is that it projects the behavioral norms of non-cooperative interactions that are characteristic of the economic sphere directly onto the political sphere where people are expected to interact in a cooperative manner, and thus fails to respect the distinct nature of the political sphere of human social life.

How to treat others as full-fledged fellow citizens is a central topic (or perhaps, the central topic) of political philosophy. A comprehensive account on how to relate to other fellow citizens is far beyond the ambition of this paper. However, it suffices for the purpose of this paper to show that treating others as part of the environment, as pivotal voters do, violates what is required of fellow citizens.

My normative argument against pivotal voting goes as follows.

(1) Pivotal voters treat other voters in a parametric way that is indifferent from how they treat the environment.

(2) By treating others as part of the environment against which they optimize for their own profit, pivotal voters fail to satisfy what is required of participants of democracy (i.e., citizens in a democracy).

(C) Pivotal voting, viewed as a voting norm, thus fails on normative grounds.

Premise (1) can be most clearly seen from the fact that the environmental factors are commonly modelled as an additional player, often called “Nature” in game theory. The formal treatment of other players from the perspective of a pivotal voter is indifferent from the formal treatment of any environmental factor.

This point is clearly made in Roemer (2019, pp. vii-viii), as follows.

"In Nash equilibrium, each player treats all other players' actions *parametrically*, that is, as part of his or her environment. How do we define Nash optimization? *Taking the actions of the others as fixed*, what is the best action for me? I think a model of cooperation should show explicitly how each individual in the group contemplates how others will coordinate with him or her—others should be viewed, not as part of the environment, but as part of the action. This means that optimization must be done in a non-Nash way."

Being so motivated, Roemer (2019) proposes "Kantian optimization" as an alternative to Nash optimization for modelling cooperative behaviors among agents where other players are treated differently from environmental factors. This paper, instead, focuses on the normative aspects of voting norms in the CJT setup. We shall, thus, avoid going into the details of Roemer's fascinating work.

Premise (2) would take book-length efforts to get fully established. I shall only give a sketch here.

The purpose of democracy is to form a general will that aims at the common good of citizens. Toward this end, participants of democracy (i.e., citizens) must recognize other participants as fellow citizens and thus have others in view *in the correct way*. A citizen must not only recognize the existence of other citizens, but also recognize the fact that *she is but one citizen among many*. This relationship of fellow citizens requires her not to treat other citizens as a mere environmental factor against which she optimizes her own interests.[26] While pivotal voters do recognize the existence of other voters, they fail to see others as fellow citizens. Instead, others are

[26] This recognition is, in important ways, analogous to the recognition of the existence of other persons, as discussed in *The Possibility of Altruism*, by Thomas Nagel.

viewed as parts of the environment that pivotal voters must navigate through. In Kantian terms, one may say that pivotal voters treat other voters as mere obstacles, which is a counterpart of mere means. Doing so is not even treating other voters as fellow citizens, not to mention as ends.

The conclusion follows from the two premises.

It's worth noting that the criticism here is not that pivotal voters are selfish and fail to take others' interests into consideration. So far, we assume that all agents have identical preferences and thus there is no real difference between being self-interested and being altruistic. One optimizes for everyone's benefit by optimizing for herself. Rather, the proposed criticism concerns the way in which pivotal voters interact with other voters, namely, to treat other voters merely as an additional environmental factor. However, the objection does seem to carry a stronger rhetoric force in presence of heterogeneous preferences among voters, which I will briefly discuss in the next section.

My criticism against pivotal voting is an instance of the broader criticism against the unreflective application of economic methods toward non-economic matters. Some criticisms of this kind have been made in the past under the title of 'economic imperialism'.[27] What's more important is that my criticism provides a new perspective for assessing voting norms. Namely, we should assess how a voting norm prescribes the manner in which a voter has other voters in view.

So far, I have been assuming that voters have identical preferences. In the next section, I will present the case of heterogeneous preferences. While this case is important in its own right, it also helps clarify a similar moral objection concerning sincere voting.

---

[27] See Udehn (1996) for an extended discussion of economic imperialism.

2.6 The Presence of Heterogeneous Values

In this section, I carry the previous discussion further by allowing voters to have heterogeneous preferences. The motivation is to accommodate the presence of the heterogeneity of values in a democratic society. The heterogeneity of value, as noted by John Rawls in *Political Liberalism* (p. 36), is "not a mere historical condition that may soon pass away" but "a permanent feature of the public culture of democracy". Following Rawls, I take the case of heterogeneous values to be the central case of democracy, and thus take the case of heterogeneous preference to be the central case of the CJT.

In the presence of heterogeneous preferences, the disagreement among voters is no longer purely epistemic. Recall that I formulated voters' epistemic disagreements as concerning the actual state alone. In the case of homogeneous preferences, once the voters agree on which state is actual, there is no disagreement left over candidates. Here, we allow voters to have different values, represented by their state-dependent preferences over candidates. Even after the voters resolve their epistemic disagreements concerning states, they can still disagree on which candidate should be elected. This remaining disagreement comes from their fundamental disagreement of values.

The first challenge CJT-type results face in this setup is that there is no longer a clear notion of the correct candidate. In the classic CJT, in each state, there is an indisputable correct candidate that everyone prefers over the other candidate. This is no longer true when voters have different state-dependent preferences over candidates.

Goodin & Spiekermann's response to this challenge is as follows.

"If the correct answer to the same question is different for different people, what makes a group decision the correct one? One plausible proposal is to endorse the answer supported by most voters. We call that outcome 'democratically-epistemically correct', for short …… Even if there

is no one option that is correct from the point of view of literally everyone, it's democratically better (assuming some substantive safeguards against majority tyranny are in place) to settle upon the option that is correct from the point of view of the largest number of people". (p. 196)

They went on to show that if each voter votes sincerely, the probability the 'democratically-epistemically correct' outcome gets elected goes to 1 as the size of electoral increases, just like in the classic CJT.

Feddersen & Pesendorfer (1997, p. 1031) takes the same stance on this matter with less reported philosophical reflection. They define the correct candidate as the candidate that will get elected if there is no uncertainty over states among voters. This is equivalent to saying that the correct candidate is the candidate that is preferred by most voters.

We shall follow their lead as this seems to be the most intuitive (and, at this moment, the only) way to accommodate heterogeneous values in CJT-type results. Once we identify the candidate preferred by most voters as the correct candidate, regardless of how we call it, the mathematics regarding both sincere voting and pivotal voting stays the same. On the other hand, the presence of heterogeneous values seems to make quite some difference to the moral assessment of the two voting strategies, especially of sincere voting.

In this case, my argument against pivotal voting becomes more intuitive, as a pivotal voter who treats others as background obstacles no longer shares the same preference with many other voters. The same argument works, hopefully with more rhetorical forces.

The real interest of this case lies in sincere voting. We have defined sincere voting as the strategy that maps all voting scenarios to a sincere voting behavior. In a sincere voting behavior, the vote casted by the voter coincides with her true preference. There are several alternative yet

equivalent definitions for sincere voting. One of the most common definitions best draws out the moral concern.

Sincere voting is commonly defined as to vote as if one's own vote dictates the result. It should be apparent that this definition identifies the same mapping from voting scenarios to voting behaviors as mine. However, it seems morally problematic, if we evaluate it from the moral perspective I proposed in the last section, namely the way in which a voter has other voters in view. Sincere voting, so formulated, explicitly requires one to behave as if she is the only voter there is. All other voters do not even factor into a sincere voter's deliberation at all. Individuals who follow this solipsistic voting strategy, thus, do not even have others in view. The moral failure of sincere voting, so understood, goes beyond that of pivotal voting. While pivotal voters view other voters as mere obstacles, "sincere" voters, who vote as if their vote dictates the result, do not have others in view at all.

This failure of sincere voting is best brought out in the case of heterogeneous preferences. In the classic CJT, thanks to the homogeneity of values, a voter has all other voters in view in virtue of having herself in view. The failure, however, becomes self-evident in the presence of heterogenous values, where having oneself in view and having others in view do not coincide perfectly. It's worth emphasizing the significance of this failure, since the case of heterogenous values, as Rawls suggests, is the central case of democracy.

Have philosophers in the end failed to justify themselves in their own (viz. moral) terms? I do not think so. I end the paper with a response in defense of sincere voting and some general big-picture remarks.

In this paper, voting behaviors and voting strategies are characterized in extensionalist terms, in line with the formal treatments of CJT-type results. As I mentioned before, the intensional

structure of voting strategies (i.e., as a function from voting scenarios to voting behaviors) is meant to capture certain intentional content of voting. But it does not do so perfectly, especially for sincere voting, which has a minimal functional structure. Such extensionalist formulations are suitable for the purpose of game-theoretic strategic analysis but fall short for moral discussions. Consider the following two sincere voters, who arrive at the same voting strategy through dramatically different processes of thinking.

Sincere Voter 1: "I only care about me and what's best for me. I would vote dictatorially if I could. Sadly, I'm stuck in a democracy with a whole bunch of losers. The best I can do is to vote for what I want and hope my choice is decisive."

Sincere Voter 2: "I respect my fellow citizens and want to do right by them in the voting booth. How should I take their views into account? Well, the best way to do so would be to give them equal weight. But that is already done by the voting procedure. Given that other citizens have equal say in the outcome, securing a kind of formal respect, is there any other way I need to respect them in my choice substantively? Well, we all want the same thing (even if we disagree about what the best way is to get it), so I should at least try to cast my vote in a way that maximizes the chance we get our collectively valued outcome. I know from the CJT that casting my vote sincerely is a way of furthering that end. So, I should just cast my vote sincerely—that's the best way of maximizing the chance that we all get our collective wish."

Both Voter 1 and Voter 2 arrive at the same voting strategy — sincere voting. But their reasoning for it is vastly different. Intuitively, Voter 2 is trying her best to do right to her fellow citizens, while Voter 1 is not attempting at all. If anyone is trying to have other citizens in view, Voter 2 is clearly one of them. It's seemingly wrong to condemn Voter 2 for not having others in

view because she decides to vote sincerely, while Voter 1 is clearly guilty as charged. However, this difference cannot be reflected from the intensional structure of sincere voting.

This example shows that a further level of nuance, concerning the voter's intention and thought process, is needed for doing moral assessment properly. The fact that the moral judgments concern actors' intentions on top of their thinly described behaviors should not be new to anyone.

Following this line of thought, one may further respond in defense of pivotal voting that there could be cases where pivotal voters have other citizens in view in the correct manner, similar to what Voter 2 does above. I do not deny this possibility. It's conceivable that one can come up cases by carefully calibrating the voter's intention so that many different voting strategies can meet the requirement of having other voters in view as fellow citizens. But certainly, the task is easier for some strategies than others. As I said, the intensional structure of strategy captures some, although not all, intentional content of voting. For example, it's seemingly impossible to calibrate a case for the strategy where one votes based on the number of red cars seen on the voting day.

2.7 Final Remark

In this paper, I defended philosopher's favoritism in sincere voting on moral grounds. Philosophers should not be shy to bring up moral considerations in the context of the CJT, or more generally, in the context of epistemic democracy. The epistemic aspect of democracy is, after all, only one aspect of democracy. It must eventually be put together with other aspects.

The limitations revealed at the end of the last section, I believe, comes out of the extentionalist abstractions of voting used in CJT-type results, and more generally game theoretic analyses. Seeing this limitation prompts us to think more about what needs to be done to better put

together different perspectives together, as different perspectives are often discussed in different languages, under different assumptions and with different abstractions.

3.0 The Ideal of Information Efficiency of Financial Markets

This paper studies the ideal of fundamental valuation efficiency (FV-efficiency for short) of financial markets. Existing studies of the information efficiency of financial markets are mostly empirical and focus on information arbitrage efficiency (IA-efficiency for short). The FV-efficiency, albeit widely appraised by investors and researchers, has been rarely studied carefully. One reason behind is that it's particularly challenging to measure the fundamental values of financial assets at any given point of time. This, however, does not excuse us for not carefully examining and reflecting on this ideal. I hope that this paper shows philosophical examinations and reflections can indeed contribute to a better understanding of this ideal, and potentially help achieve it.

This paper has four sections. The first two sections are mostly expository. In §1, after a brief general introduction on the epistemic dimension of financial markets, I introduce and compare two different notions of information efficiency in financial markets: information arbitrage efficiency (IA-efficiency for short, in §1.1) and fundamental valuation efficiency (FV-efficiency for short, in §1.2). In §2, I take a deep dive into FV-efficiency, which is the focus of this paper. After defining the key concepts in §2.1, I illustrate the key conceptual apparatus, fundamental value, with concrete examples in §2.2. Overall, my exposition of the ideal of FV-efficiency focuses on what information should (and should not) be aggregated in efficient prices in an ideal market.

The latter two sections are more philosophical and exploratory. I tackle two crucial questions about the ideal of FV-efficiency: (1) how FV-efficiency can be achieved in presence of unfettered individual preferences and (2) whether the ideal of FV-efficiency is adequate.

I discuss the first question, in §3. After a brief introduction of anomalies to FV-efficiency (§3.1), I use the clever-ticker effect to motivate my discussion of how the ideal of FV-efficiency could be achieved in presence of unfettered individual preferences (§3.2). In §3.3, I argue against the idea that FV-efficiency can be achieved by market mechanisms alone. Rather, FV-efficiency can only be achieved via proper regulations. In §3.4, I leave a general philosophical remark on financial regulation, inspired by recent developments in social epistemology. Namely, it's wrong to assume that only fundamental investors can improve the FV-efficiency of financial markets.

I reflect on the second question in §4, where I discuss the recent movement of ESG (Environmental, Social, and corporate Governance) investing. According to some advocates of ESG investing, an efficient market should reflect ESG impacts of companies in addition to their fundamental value. In §4.1, I distinguish a few different types of ESG investors, which is important for understanding the challenge to the ideal of FV-efficiency presented by ESG investing. Then I try to sketch a new ideal according to the new paradigm of ESG investing (§4.2). In §4.3, I discuss some epistemic consequences of the new ideal and the new paradigm. I present two worries on processing and aggregating (non-pecuniary) ESG preferences in financial markets alone. I suggest some alternative institution, namely democracy, which may be more appropriate to treat certain aspects of ESG preferences.

While all my arguments are suggestive and far from convincing, I hope that they demonstrate how the study of financial markets, a major epistemic institution of modern society, could benefit from philosophical reflections, and how the study of financial markets is of philosophical interest and significance.

3.1 The Information Efficiency of Financial Markets

Markets are central institutions in modern societies. The primary function of markets is to provide venues in which individuals and collective agents exchange goods and services. In addition to this function of accommodating exchanges, F.A. Hajek (1945) first noted that markets also serve an epistemic function of aggregating, compressing, and communicating information in society. Various kinds of information about a good, including how much it's desired by the society (i.e., demand), how difficult/easy it is to make (i.e., associated technology), how much of it is available (i.e., supply), are aggregated into its price via market transactions.

Take oil (petroleum) for example. Events that influence the demand and the supply of oil happen around the world on a daily basis. Some new oil fields have been discovered. Some oil-producing countries went to war. Some new technologies (say nuclear fusion power) that will eventually make oil obsolete have made some progress. And so on. All such information gets aggregated and reflected in the price of oil that is easily accessible to everyone. One doesn't have to become an expert on oil technology or on oil-producing countries' politics to make informed decisions that involve oil, say whether to buy a bigger car. All it takes is the price of it.

This epistemic function of markets is even more crucial for financial markets, where financial securities (stock, bond, etc.) are traded. Fama (1970), the *locus classicus* on market efficiency, starts as follows:

> THE PRIMARY ROLE of the capital market is allocation of ownership of the economy's capital stock. In general terms, the ideal is a market in which prices provide accurate signals for resource allocation: that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any

time "fully reflect" all available information. A market in which prices always "fully reflect" available information is called "efficient."

According to Fama (1970), the ideal of financial markets is one where its epistemic function is fully served. Fama calls such an ideal market “efficient”.

In this passage, two related yet distinct functions of financial markets are mentioned. In the first sentence, Fama talked about the “allocation of ownership of the economy’s capital stock”, which we may call *the allocation function* of financial markets. In the second sentence, Fama switched to *the epistemic function*, that prices provide accurate signals for resource allocation. Each of the two functions is associated with a different notion of efficiency.

The concept of Pareto efficiency is associated with the allocation function of markets. An allocation is called “Pareto efficient” if there is no alternative feasible allocation that makes some individual better off without making someone else worse off. The epistemic function of markets, on the other hand, is assessed by its “information efficiency”, which as defined in the passage above, concerns whether prices fully reflect available information. Stiglitz (1981) presents models where these two notions of efficiency come apart. A financial market can be efficient in one sense without being efficient in the other sense.

There are, in fact, more than two senses in which a financial market can be efficient. Tobin (1987) distinguished four different senses of market efficiency: information-arbitrage efficiency, fundamental-valuation efficiency, full-insurance efficiency, and allocation efficiency (which Tobin calls “functional efficiency”). Among them, both information arbitrage efficiency (IA-efficiency, for short) and fundamental-valuation efficiency (FV-efficiency, for short) concern the epistemic function of financial markets.

In the rest of this section, I introduce and compare IA-efficiency (§1.1) and FV-efficiency (§1.2). In the next section, we will take a closer look at FV-efficiency (§2).

3.1.1 Information-Arbitrage Efficiency

Information arbitrage efficiency (IA-efficiency) is the most studied sense of efficiency in Finance. The affirmation of IA-efficiency of financial markets is known as the Efficient Market Hypothesis (EMH). Since the publication of Fama (1970), an enormous literature developed around the EMH and is still growing today. In 2013, three financial economists shared the Nobel Memorial Prize in Economics for their contributions to this literature. Among them, Eugene Fama is a leading advocate of the EMH, Robert Shiller wrote extensively against it and Lars Peter Hansen sits somewhere in-between.

The IA-efficiency is defined as follow[28]:

> A financial market is information arbitrage efficient if and only if the returns[29] of financial assets are not predictable from publicly available information.

One may find this condition counterintuitive upfront. What's behind it is the information processing mechanism of financial markets. If the future return of some asset is predictable from

---

[28] This definition of IA efficiency is known as "semi-strong efficiency" in the EMH literature. This terminology comes from Fama (1970). Weak efficiency requires returns not predictable from historical prices. Strong efficiency requires returns not predictable from all information, public or private.

[29] The return of a financial asset captures all pecuniary benefits associated with this asset. For any asset, its price change is always part of its return. For some assets, returns have other parts in addition to their price changes. For example, US bond issues coupon payments. Stocks issue dividends.

some piece of information, one could make a profit by taking advantage of it. Doing so will move the market price accordingly, which eliminates the opportunity for profit and hence incorporates this information into the market price. If market prices fully reflect all publicly available information, no one will be able to predict future returns based on any publicly available information, since any such information must have already been reflected in the price. Therefore, the more efficient the market is, the more unpredictable the returns are.

What makes IA-efficiency so popular in academic Finance is that it has abundant empirically testable implications. One common route of testing is to assume that the sequence of returns of an asset forms a random walk whose mathematical properties can be tested statistically. Also, one can assign structure on various information sets, including historical prices and other publicly available information, to test statistically whether returns are predictable from these information sets.

The empirical study of the return predictability has lasted many decades. It has taken a few turns in the process. Early studies, which are well summarized in Fama (1970), found a lack of predictability in returns. From the 1980s to 2000s, researchers found many reliable predictors for returns. These predictors are known as anomalies in Asset Pricing. Lo (2000) gives a good survey of this period of research, including major anomalies. More recently, most predictors discovered in earlier times cease to work (Welch & Goyal, 2008), as trading technology develops (Chordia et al., 2014) and investors incorporate academic research into their practice (McLean & Pontiff, 2016), or so the authors argue. Overall, well-developed financial markets (e.g., major U.S. markets) seem to be IA-efficient for the most part of time.

With the empirical studies in view, I'd like to leave two remarks about the ideal of IA-efficiency from a theoretical point of view.

First, IA-efficiency concerns publicly available information of all types. The scope is not constraint to information about the fundamental values (which we will define later), or any other aspects, of financial assets. According to this ideal, any piece of information that could be used to predict future returns will soon get processed and aggregated into the efficient price (i.e., prices in an efficient market) by the market.

Second, IA-efficiency is a relatively weak notion of efficiency. The unpredictability of returns is only an indirect measure of how well financial markets are doing their jobs. We expect an ideal financial market to correctly price financial assets. The unpredictability of returns is necessary for pricing correctly, but not sufficient. (More on this later.) Consider the financial crisis of 2008, a major failure of U.S. financial markets. The financial crisis started from the incorrect pricing of subordinated mortgage bonds and their derivatives. The market prices of certain derivatives from mortgage bonds fail to reflect their real values. The market prices eventually get corrected, but the process of correction leads to devastating social and economic consequences. When we reflect on this failure, we really do not care about whether the returns are predictable or not during that period. The market failure of 2008 is primarily a failure for the market prices to reflect the real values of some financial assets. This ideal of correct pricing is captured in the fundamental valuation efficiency.

3.1.2 Fundamental Valuation Efficiency

Fundamental valuation efficiency (FV-efficiency) is defined as follows:

> A financial market is fundamental valuation efficient if and only if the price of the financial asset always equals its fundamental value, which is the rational expectation of its discounted future cashflow.

This definition is apparently highly idealized, best revealed by the phrase "always equals" in it. Of course, no actual market is FV-efficient in a tenseless manner. Also, no actual market is perfectly FV-efficient. The real question for actual markets is how FV-efficient it is during a certain period. When we discuss actual markets, we have to replace "always" by some period of time, and "equals" by "closely approximates". But this precise formulation suffices for our theoretical discussion of ideals for financial markets. So, we will stick to it.

In the next section, I will take a closer look at the key concept, "fundamental value", in this definition. Before that, I'd like to leave some remarks about FV-efficiency while boxing the notion of fundamental value.

First, unlike IA-efficiency, FV-efficiency only requires some but not all information that could potentially lead to market movements to be reflected in efficient prices, namely, information that concerns the fundamental value of assets. We may call such information "fundamental information". Non-fundamental information, according to this ideal, should not be reflected in efficient prices. I elaborate more on this point below in §2.2 and §2.3.

Second, FV-efficiency is generally considered a stronger notion of efficiency. It's clear that IA-efficiency does not imply FV-efficiency. We can imagine a financial market where prices move randomly but fail to track the underlying fundamental values. Shiller (1979, 1981) argue that U.S. bond market and U.S. stock market of that period are such examples. In particular, Shiller shows that the prices in these markets are much more volatile than and thus fail to track corresponding fundamental values. The cryptocurrency market, if ever IA-efficient, will be another example according to its critics who think that cryptocurrencies do not have fundamental

values at all. J M Keynes, a renowned economist and an active market participant, also suggested that actual markets of his time are IA-efficient but not FV-efficient.[30]

On the other hand, IA-efficiency directly follows FV-efficiency as defined above. In short, rational expectation is, by definition, the best forecast based on publicly available information. If the price of an asset always equals the rational expectation of its discounted future cashflow, no one can outperform the market with publicly available information. One may want to switch to a less idealized definition of FV-efficiency in discussing actual markets. It's possible to formulate a less idealized version of FV-efficiency such that IA-efficiency does not follow from it. But even under that definition, a market that is FV-efficient but not IA-efficient will be truly unusual. I will return to this point in §2.3.3, after we take a closer look at FV-efficiency in the next section.

3.2 The Ideal of Fundamental Valuation Efficiency

The concept of fundamental value and FV-efficiency are central to theoretical and practical understanding of financial markets. For example, Mishkin (2007, pp. 150-2), one of the most widely used textbooks on financial markets, defines market efficiency as FV-efficiency. Warren Buffett, the most successful fundamental investor of our time, says that fundamental value "offers the only logical approach to evaluating the relative attractiveness of investments and business".[31] The very idea of FV-efficiency is also at the bottom of economic models of financial markets,

[30] So argued by Wang (1985, p. 346). See also Tobin (1984, p. 7).

[31] See The Essays of Warren Buffett (2019, Cunningham eds.), Chapter VI. B. "Intrinsic Value, Book Value, and Market Price". Buffett calls what I call fundamental value "intrinsic value".

where investments are modelled as sequences of future payments and market participants make decisions by forming rational expectation of them.

In this section, I take a closer look at the ideal of fundamental valuation efficiency (FV-efficiency). I first define fundamental value in general terms (§2.1) and then illustrate those definitions with bond (§2.2.1) and stock (§2.2.2).

After seeing examples of fundamental information, I switch to a discussion about non-fundamental information in §2.3.1. and how ESG information, a specific example of non-fundamental information, could bring challenges to the ideal of FV-efficiency (§2.3.2). I defend the ideal against the challenge of ESG information for the ideal case where non-fundamental information is internalized as fundamental information via proper regulation and leave some suggestive remarks about non-ideal cases. I conclude this section with remarks on regulations and efficiency.

3.2.1 The Definition of Fundamental Value

Here I define fundamental value in general terms. These definitions can be difficult to process for readers without previous background. But they will be illustrated with concrete examples in the next section.

The fundamental value of a financial asset is the *rational expectation* of its *discounted future cashflow*. This definition has three key components: ‘rational expectation’, ‘discounted’, and ‘future cashflow’. They are defined as follows.

Rational expectation was first proposed as an equilibrium concept in Macroeconomics by John Muth (1961) and later adopted in the study of Financial Markets by Robert Lucas (1978).[32] For our purpose, it's sufficient to understand the rational expectation of an economics variable as the leading forecast (i.e., the best guess) of that variable based on all publicly available information.[33] The rational expectation as defined is not an expectation of any specific person, but an expectation formed from all publicly available information. This definition of rational expectation naturally connects FV-efficiency with IA-efficiency. The fundamental value of a financial asset is thus by definition the best guess of the discounted future cashflow of that asset. Therefore, in an FV-efficient market, no one can outperform by forming a better guess of discounted future cashflows, although this does not rule out the possibility of outperforming by other means.

The future cashflow of a financial asset is the sequence of future cash[34] payments that a financial asset *entitles* its owner to. Such entitlements consist of contractual rights (e.g., for bond)

---

[32] See Sargent, T. J. (2017) for a detailed history of rational expectation models in Macroeconomics. See Delcey, T., & Sergi, F. (2022) for a detailed history of the association between efficient market hypothesis and rational expectation.

[33] This captures the general idea of rational expectation, albeit being short of a proper definition. This general idea gets explicated as equilibrium prices in economic models where every market participant is rational, in the sense that they optimize according to some specified utility functions. An early example is Lucas (1978). A more recent example is Farboodi & Veldkamp (2020), which we will discuss in the second part of the paper.

[34] Sometimes, a financial asset entitles its owner to non-cash payments, e.g., some other financial assets, which in turn entitle the owner to other cash or non-cash payments. Such chains can be very long in reality but will eventually bottom out with cash payments. Cash has a very special status in the financial system and is taken as a primitive here. For some interesting discussion on the nature of cash and payments, see Gleeson (2019).

or ownership claims (e.g., for stock). By focusing on the future cashflows, the definition of fundamental value employs a common yet important abstraction of financial assets. Namely, a financial asset can be identified by its future cashflow. Other attributes of a financial asset, by this definition, do not factor into its fundamental value and thus should not be reflected in the efficient prices according to the ideal of FV-efficiency. We will reflect further on this abstraction and its implications later in this section.

The last piece in the definition is "discounted". Discounting is the process of determining the present value of a payment or a stream of payments that is to be received in the future. The discounted future cashflow of a financial asset is, thus, the present value of the stream of future payments entitled by this asset. Putting these pieces together, the fundamental value of an asset is the leading forecast of the present value of future payments this asset entitles its owner based on all available information.

Now I have defined what fundamental value is in general terms. In the next section, I illustrate these definitions with two common types of financial assets, bond and stock. It will help us better understand what fundamental value is and what information plays a role in the determination of fundamental values.

3.2.2 Two Examples

In this section, two common types of financial assets, bond and stock, are used to illustrate the definition of fundamental value. The example of bond presents some important factors in determining the fundamental value of financial assets and thus gives us a hold on what fundamental information is. The example of stock, on the other hand, highlights a common yet crucial

assumption behind the ideal of fundamental information. Namely, a financial asset is identified by its future cashflow.

3.2.2.1 2.2.1 Bond

Let's start with the simplest possible example, a nominal bond with no coupon payment.[35] For a nominal bond, there is no uncertainty with respect to the structure of its future cashflow. When the bond is issued, when and how much the issuer should pay the bond owner is determined upfront as part of the bond. For our example, let's say that the bond has a face value of $1000 and its maturity date is August 1st, 2024. That is, this bond entitles its owner to a simple payment of $1000 on August 1st, 2024.

It is possible for the holder of this bond to sell the bond and collect a cash payment before August 1st, 2024. Let's say that she sells the bond on February 1st, 2024, for 950 USD. This payment on February 1st, however, is not part of the future cashflow of this bond, because the bond does not entitle its owner to this payment. This payment comes out of an exchange between the owner who sells the bond and the buyer who pays for it. To qualify as a part of the future cashflow

[35] A bond represents a loan made by the bond owner to the bond issuer. A nominal bond is a bond whose payments are determined at its issuance in terms of cash. A bond has a face value, which is the money amount due to the bond owner on the maturity date of the bond. By convention, U.S. bonds often pay a coupon rate, which is a regular payment of a fixed percentage of the face value paid before its maturity. For example, a 10-year bond with $1000 face value and 5% coupon rate entitles its owner to receive $50 dollars each year on the coupon date and $1000 on the maturity date, which is 10 years after the issuance of the bond. All relevant dates and rates are determined at the bond's issuance. In our example, I assume that the coupon rate is 0 for simplicity. The bond thus only carries one payment at its maturity.

of a financial asset, the payment must come from either a contractual right (in the case of bonds) or an ownership claim (in the case of stocks) that is part of the financial asset.

After clarifying the future cashflow of a bond, we consider four factors that play crucial roles in calculating the fundamental value of a bond, which is the leading forecast of its discounted future cashflow.

The first factor is *credit risk*. The bond issuer may not be solvent when the bond matures. When that happens, the bond owner will not receive the payment entitled by the bond in its entirety. This is the credit risk of bonds. The credit risk concerns the solvency of the bond issuer on the maturity date. Any information that concerns the issuer's future solvency will go into the rational expectation of this factor. Clearly, for a bond with face value of $1000, if one expects the bond issuer to be solvent with 0.9 probability on the bond's maturity, one will assign a mathematical expectation of $900 for this bond.[36] This, however, assumes that the individual is risk-neutral, which is rare among actual investors. The credit risk of a bond must filter through investors' aggregate risk attitude before it gets properly reflected in equilibrium prices.

This brings us to the second factor, Investors' *risk attitude* (or *risk preference*). Risk preference is a familiar concept from decision theory and microeconomics. In the context of financial markets, investors' aggregate risk preference is often measured by the *risk premium* offered by risky assets. US treasury bonds are generally considered to be safe assets since they are credit-risk-free. If U.S. treasury ever ends up not having enough U.S. dollars to meet its obligations, it always has the U.S. central bank, which can create U.S. dollars out of thin air, as its

[36] This assumes that the bond holder will either get paid in full or not get paid at all. In reality, the bond holder may receive partial payment of the principal value of the bond.

last resort. Bonds issued by other issuers, e.g., U.S. corporations, do not have this safe net. Being risky in nature, these other bonds must offer some extra returns (i.e., to promise a higher interest rate) to attract investors. This extra return is known as a risk premium. How much extra return is needed for each unit of extra risk? For a single investor, this depends on this investor's risk attitude. Some investors are more willing to take risks for higher return, known as risk-seeking, and would require a less risk premium. Others are more risk-averse and thus will only be moved by a higher risk premium. When calculating the fundamental value of a bond, we need to factor in the aggregate risk attitude of all investors in the market, in addition to the credit risk of the bond. This factor does not concern the underlying asset per se but depends on the preference of market participants regarding risks. Needless to say, this aggregate preference of market changes over time.[37]

The third factor is *inflation*, which is the decrease of the purchase power of money. Historically, the purchase power of the U.S. dollar has been going down. Based on this fact, one should expect $1000 in the future to have less purchase power than $1000 now. Even for a U.S. treasury bond with no credit risk, one would like to discount this change in purchase power when calculating the present value of its future payments. A few different metrics are used to measure inflation. Each metric is essentially a basket of goods selected by certain criteria. The purchase power of a certain amount of dollars is measured by how many of these baskets those dollars can buy.

[37] It's impossible to measure this aggregate preference from individual preferences. Instead, this risk attitude of the market is often measured by the market's risk premium, which is the average amount of extra return provided by risky assets (e.g., stock, corporate bonds) compared to safe assets (e.g., treasury bond).

The last factor I consider here is investors' *time preference*, which is sometimes known as the real interest rate. Suppose that there is no credit risk associated with the bond and that there is no change in purchase power between now and the bond's maturity. I would still prefer to have the same purchase power available now than later, and thus prefer 1000 USD now over 1000 USD in the future. There are a few related yet different factors behind the overarching concept of time preference. First, there is the psychological effect that humans tend to prefer consumption now over future consumption. Second, having the same purchase power available earlier makes more options open to me. By committing my money to a bond, my options are limited as I can only access its purchase power after the bond matures. If I choose to keep the money instead, I can use it any time from now on. Third, committing my money to a bond not only limits my options to consume, but also limits my options to invest. I lose the opportunity to put this money into other investments if I commit my money to a bond. This is known as the opportunity cost of investment. These three factors are intimately connected yet different in subtle ways. It suffices for our purpose to put them together under the generic term "time preference". Like the market's risk attitude, investors' time preference is determined by the aggregate preference of investors, which change over time.[38]

Now we have in view the four most important factors that influence the fundamental value of a bond. In fact, these factors apply to almost all investments. Inflation and time preference

[38] Since treasure bond does not have credit risk, a nominal treasure bond is discounted by inflation and time preference alone. Taking inflation away from that interest rate, what's left is called the real interest rate, which is a proxy of time preference. Many governments, including the U.S., also issue inflation-linked bonds, whose interest rates are adjusted for inflation. The interest rate of inflation-linked bonds is also a measure of real interest rate and investors' time preference.

simply apply to any future cashflow. Risk preference plays a role for any risky asset, which is basically anything other than treasury bonds. Credit risk is somewhat specific to bonds and will be replaced by other factors for other assets (e.g., company fundamentals for stocks). These factors are closely related to each other. For example, credit risk cannot be priced without knowing or assuming investors' risk preferences. Estimating any one of these factors is extremely challenging. Any information that helps form a rational expectation of any one of these factors is considered "fundamental information".

In the case of bonds, fundamental information includes (1) information that helps predict the bond issuer's future financial standing, (2) information about investors' aggregate risk and time preferences, and (3) inflation. Inflation is itself a complicated topic, which I will not touch on in this paper. What I want to emphasize is (2) investors' aggregate risk and time preference. This type of information is often neglected in the discussion of fundamental value, as people tend to focus more on the asset itself. But an efficient or equilibrium price cannot be formed without this information. An FV-efficient market is a market where this information is aggregated in prices since prices cannot be formed without it.

3.2.2.2 2.2.1 Stock

Compared to bonds, the structure of the future cashflow of stocks is much more complicated. This complex nature of stocks makes some abstractions in the ideal of FV-efficiency more visible. Getting clear about these abstractions prepares us for the discussion of non-fundamental information in the next section. On the other hand, once we understand the structure of stocks, discounting a stock is not too different from discounting a bond, at least in theory.

A share of the stock of a company is a share of the company's ownership. For return-seeking investors, the primary sense of 'the ownership of a company' is the claim to the company's

equity. The equity of a company is what is left of the company's assets (e.g., cash, equipment, properties, etc.) after deducting the company's liabilities, which are debts of various formats, including loans, costs owed to suppliers, salaries owed to employees, products owed to customers, etc.

The entirety of a company's equity will only be paid to its shareholders at the company's liquidation. In a liquidation, all the company's assets are sold, all the company's liabilities are settled, shareholders receive whatever is left of the company and the company gets shutoff entirely. The future cashflow of a share of stock is typically not a one-time payment, but a stream of payments known as dividends. Dividends are payments a company issues to its shareholders as a return on their investment at the company's discretion. Dividends are typically paid in cash, although some company occasionally issues non-cash dividends.[39] We shall focus on the primary case of cash dividends.

The fundamental value of a stock share is thus the rational expectation of its discounted future dividends. Dividends by nature involve greater uncertainty, which makes both the fundamental value and the market price of stocks much more volatile than those of bonds.

The stream of dividends of a company can potentially be infinitely long. A company may last forever and keep paying dividends to its shareholders *ad infinitum*. The stream will only come

[39] Sometimes, companies pay dividends in stock shares. If a company issues its own shares to its shareholders, it's known as a stock split, where each shareholder gets some number of shares from each share they own. No real payment is involved in a stock split since it does not change the distribution of ownership among shareholders. A shareholder who owns 2% of the company still owns 2% after the stock split. More rarely, a company may decide to give shares of another company owned by itself to its shareholders. In this case, some payments are made from the company to its shareholders in the form of stock shares.

to an end if the company goes bankrupt or gets liquidated. At that point, the shareholders will either receive one last dividend payment in the case of liquidation or nothing in the case of bankruptcy. Bankruptcy is the worst case possible for a company's shareholders. The limited liability status of modern companies exempts shareholders from being responsible for a company's liability that exceeds its assets.

No single individual can live long enough to benefit from all potential future dividends.[40] How should I value payments that will only be made after my death? In reality, very few investors will hold a stock forever. Sooner or later, they will sell their shares on the market. Let's say that an investor bought a stock at time t for price p. Instead of holding it forever, the investor sold it later at t' for price p'. If the market is FV-efficient, p (and p') would equal the rational expectation of discounted future dividend payments from t (and t') onward. Thus, what the investor did in this process is that she purchased a stream of dividends at t for p, collected payments between t and t', and cashed out all dividends after t' at once for p'. One way to make sense of the process is to think that the investor belongs to a particular generation. She bought the stock from an earlier generation and collected dividends from t to t', before selling it to the next generation investors who can and are willing to hold the stock and collect dividends from t' to some t''.

Apparently, this process is only possible under the assumption of market efficiency. Being able to sell the stock at/near its fundamental value in the future is crucial for stock investors since it's impractical to collect all future dividend payments. In this sense, stock investment has the

[40] But some institutions, e.g., a national pension fund, can. This why institutional investors often have a longer investment horizon than individual investors and are more willing to hold investments for longer time.

assumption of FV-efficiency at its very basis.[41] If the market is not FV-efficient at all, it would not make sense for most investors to even consider investing in stocks.

Unlike bonds whose future payments are determined upfront, when and how much dividends to pay are at the discretion of the company on a rolling basis, depending on the success or failure of the business. Most big companies issue dividends on a regular basis in proportion to their profits. But it's also possible for a (sometimes publicly listed) company to never issue any dividend from its start to its end, in which case shareholders lose all their investments. Or the company could issue a one-time dividend at its liquidation. Information concerning the future dividends of a company, including its profitability, revenue, assets, growth potential, etc., is known as the fundamentals of a company (or *company fundamentals* for short). Such information, of course, is crucial for determining the company's fundamental value, and is part of the fundamental information that should be reflected in the efficient price in an FV-efficient market. In addition to company fundamentals, factors we discussed earlier, including investors' risk preference, investors' time preference and inflation also apply to stocks.

The case of stocks highlights a crucial abstraction behind the notion of fundamental value. Namely, a financial asset is identified by its future cashflow. Other aspects of the financial asset are abstracted away. This is most obvious in the case of stocks. A share of the stock of a company is often more than a mere claim on the company's future dividends. Stockholders jointly own the company and can participate in the decision-making process of the company. Modern corporation

[41] This is also where liquidity comes into the picture. Liquidity, although not the focus of this paper, is widely used as a metric of market quality, in addition to (FV-)efficiency. While (FV-)efficiency concerns whether I can sell my shares near its fundamental value, liquidity concerns how much I can sell without disrupting the market.

is the main way in which production is organized in our society. A company provides jobs, produces consumer products, and makes impacts on individuals and society in numerous ways. It's much more than a mere profit-generating and dividend-issuing machine. However, all attributes that do not help estimate a company's future dividends are abstracted away in the ideal of FV-efficiency. All that matters for the company's fundamental value is its future dividends. I will revisit the implications of this abstraction in §4 where I discuss the recent development of ESG investing.

3.2.3 Fundamental and Non-Fundamental Information

We now have a grasp over what fundamental information is. In general terms, fundamental Information about a financial asset is whatever information that contributes to the formation of the rational expectation of its discounted future cashflow. Other information that does not contribute as such is non-fundamental information.

While many attributes of a stock contribute to the determination of its fundamental value, some attributes are apparently irrelevant. Such attributes are non-fundamental information that should not be aggregated into efficient prices according to the ideal of FV-efficiency. In the next section, I start with a concrete example where some non-fundamental information, namely stocks' ticker symbols, is in fact reflected in actual markets. This anomaly is apparently caused by investors' revealed preferences over stocks with clever ticker symbols. I discuss how the ideal of FV-efficiency can be achieved in the presence of such non-fundamental preferences.

3.3 Anomalies: Challenges in Realizing the Ideal

A failure of market efficiency is known as an anomaly in asset pricing. In this section, I reflect on a specific anomaly to FV-efficiency, known as the ticker effect. Using this example, I attempt to show that the achievement of FV-efficiency is far from automatic. Proper regulations are needed for the ideal to come true.

I start with a general discussion of anomalies to FV-efficiency in §3.1. In §3.2, I present the clever-ticker effect, which motivates us to reflect on how FV-efficiency can be achieved in presence of individuals' unfettered preferences, like the clever-ticker preferences. In §3.3, I discuss certain market mechanisms that are supposed to help realize the FV-efficiency. I argue that these mechanisms are not adequate in themselves. Proper regulations are needed to realize FV-efficiency. In 3.4, I leave a general remark on regulating financial markets toward FV-efficiency inspired by recent developments of social epistemology.

3.3.1 Anomalies to FV-efficiency

An extensive empirical literature has developed around anomalies to IA-efficiency, which are relatively easy to find in theory. All it takes is to find some predictor that can predict the returns of some assets over a relatively long period of time. This is more difficult than it sounds in practice, since any such anomaly is an opportunity for profit that every investor actively searches for.

Anomalies to FV-efficiency are even more difficult to study. There are some obvious cases where prices fail to approximate fundamental values. The financial crisis in 2008 is a good example. But other than such dramatic failures, subtle failures of FV-efficiency are elusive since it's impossible to measure the rational expectation of the discounted future cashflow of some asset

independent of its market price. All attempts to empirically test the FV-efficiency of actual markets are done indirectly. For example, Shiller (1979, 1981), which I mentioned earlier, argues that actual markets are not FV-efficient based on the observation that prices are much more volatile than underlying cashflows from a retrospective perspective.

Earlier (§1.2), I mentioned that IA-efficiency follows from FV-efficiency as defined. If so, every failure of IA-efficiency is also a failure of FV-efficiency. Let's revisit this point before we move to specific examples in the next section.

IA-efficiency says that returns of financial assets are not predictable from publicly available information. Generally speaking, returns of a financial asset come from two sources: price change and the realization of its future cashflow (e.g., payments from bonds or dividend from stocks). In an FV-efficient market, the future cashflow of an asset is imminently tied to its price. We may think that the realization of future cashflow is just a kind of price change in an FV-efficient market. The maturity of a bond is nothing but a price change to its face value. The issue of dividends from a stock merely reflects the fact that one of its future payments is dated today. Thus, all returns in an FV-efficient market come from changes in prices, which are equivalent to changes in fundamental values.

The fundamental value is itself a best guess formed from publicly available information. A better guess of a best guess is not possible unless some new information arrives. Thus, changes in fundamental value are not predictable in an FV-efficient market. Therefore, an FV-efficient market as defined is guaranteed to be IA-efficient, as price changes and returns in such markets are predictable. Another way to think about this is that an FV-efficient market as defined always reflects all and only publicly available fundamental information. An FV-efficient market will only

move when some new fundamental information arrives. The arrival of new fundamental information is not predictable. If it's predictable it's not new.

Such an idealized definition of FV-efficiency, of course, does not hold in any actual market. Some famous anomalies to IA-efficiency, which are also anomalies to FV-efficiency, include: the size effect, which says that small capitalization stocks tend to outperform large capitalization stocks by a wide margin over the period of a year; the January effect, which says that stocks that underperformed in the fourth quarter of the prior year tend to outperform the markets in January; the September effect, which says that returns from the stock market are generally lower in September compared to other months in a year. Each one of them has been studied and debated over extensively in the past. I will not touch on any of these effects here. Instead, I will discuss a clear example where some non-fundamental information is reflected in market prices. We may call it "the clever-ticker effect".

3.3.2 The Clever-Ticker Effect

Each publicly listed company has a code for trading, known as its ticker symbol, or simply its ticker. For example, the ticker of Apple Inc. is 'AAPL', the ticker of Microsoft is 'MSFT', the ticker of Walmart is 'WMT', etc. Researchers have found several features of stock tickers to be correlated with stock's performance. Apparently, stocks whose tickers are pronounceable, actual words in English, or simply clever tend to outperform the market average in return.

Most recently, Baer, Barry, and Smith (2020) found that a portfolio of 22 stocks with clever ticker symbols outperformed the market average in return by a substantial margin from 1984 to 2018. This study is a replication and an extension of the original study by Head, Smith, and Wilson (2009), which tested the same portfolio from 1984 to 2005. 19 stocks out of this portfolio

outperformed the market over the span of 34 years. Here are some sample tickers from the portfolio.

Table 5: Examples of Clever Tickers.

Ticker	Company Name	Primary Business
BABY	Natus Medical	medical products for babies
CASH	Comdata Network	ATM networks
DNA	Genentech	gene research
JOB	General Employment Entrepreneurs	Employment

Baer, Barry, and Smith (2020) also constructed another portfolio of clever tickers based on an online survey. They asked people with little knowledge of the stock market to select 10 "cleverest, cutest, and most memorable" tickers from a list of 69 based on brief descriptions of what the company does like in the table above. They put the 20 most voted tickers into a portfolio and found that it outperformed the market from 2006 to 2018. See Baer, Barry, and Smith (2020) for references to other past research on the effects of ticker symbols.

Unless the fundamental values of clever-ticker stocks systematically outperformed the market average, the extra returns from clever-ticker stocks must reflect a persistent preference for these stocks in aggregate among investors.[42] That is, investors decide to buy more of them, compared to the market average, for reasons independent of their fundamental values. We may call these preferences of investors that concern the cleverness of ticker symbols and are

[42] One may argue that the cleverness of the ticker of a company reflect that company is more creative, open-minded or under better leadership. The cleverness of the ticker is indeed a proxy for the company's fundamental value. While this is certainly an open possibility, it does not hurt my discussion here. The cleverness of tickers is just an interesting yet arbitrary example of non-fundamental information being reflected in the market. Any other example would do the same job.

independent of fundamental values “clever-ticker preferences”. Whether clever-ticker preferences are well-reflected or not, they are genuine revealed preferences of the market that persisted for decades. The clever-ticker effect gives us a concrete example where some non-fundamental information (clever-ticker preferences) gets reflected in actual markets persistently. There is no doubt that the clever-ticker preferences are not supposed to be reflected in an efficient market according to the ideal of FV-efficiency.

The very fact that the ideal of FV-efficiency requires some genuine preferences of market participants not to be reflected in the market may disturb some *laissez-faire* economists. If it’s some genuine preference of investors, why shouldn’t they be reflected in market prices? One may answer that such preferences do not seem to be rational. But what makes a preference rational? All preferences are on equal grounds according to modern decision theory. For example, in Savage’s theory, all preferences are captured by a single binary relation and cannot be distinguished from others.

Following this line of thought, a *laissez-faire* economist may end up rejecting FV-efficiency all together on the ground that it privileges certain preferences over others. I want to make two responses to this concern. The first concerns why do we need an ideal of efficiency and the second concerns the relationship between the ideal of FV-efficiency and individual economic rationality.

The epistemic function of financial markets is to aggregate, compress and communicate information about financial assets via prices to individuals so that they can make better and more informed decisions over investments and other matters. Better individual decisions together lead to better allocations of the economy's capital stock and other resources. There is some information behind every movement of the market. Why did the market plunge so much yesterday and recover

so fast today? Because a lot of investors solder their investments at lower prices yesterday and bought them back at higher prices today. While it may sound trivial, it's a genuine piece of information about the market. If we do not distinguish some set of information from others, we will simply trivialize the notion of efficiency by concluding that the market is just what it is.

It's true that IA-efficiency does not distinguish some sets of information from others. But as I said earlier, IA-efficiency is an indirect measurement of how the market is doing its job. The mere unpredictability is far from a guarantee to be informative. If we want our ideal market to be informative at all, the ideal must be more than IA-efficiency. FV-efficiency provides one demarcation between information that should be reflected in efficient prices from information that should not. Whether this demarcation is appropriate is up to debate. I intentionally picked the clever-ticker effect to start our discussion of anomalies because the clever-ticker preference is intuitively not the kind of information that one expects an ideal market to reflect in efficient prices. I will revisit this question, namely whether FV-efficiency is the correct demarcation, in §4 where I discuss ESG-related preferences, which some people argue should be reflected in efficient prices despite of being non-fundamental.

Now, let's address the potential tension between the ideal of FV-efficiency, which requires only certain preference gets aggregated into efficient prices, and contemporary theories of economic rationality, which does not distinguish among preferences.

FV-efficiency requires that the market in aggregate reflects all and only fundamental information in efficient prices. This requirement is not a requirement for individuals. FV-efficiency does not condemn individuals for having non-fundamental preferences, like the clever-ticker preference, with or without reflections. Instead, the ideal of FV-efficiency only requires that such preference is not present in aggregate through efficient prices. You may naturally wonder

how this ideal could be realized without putting constraints on market participants. This naturally brings us to the next section, where I discuss how the ideal could be realized in the presence of unfettered individual preferences.

3.3.3 Market Mechanism for FV-Efficiency

Traditionally, advocates of market efficiency argue that the ideal of FV-efficiency can be achieved in the long run through market mechanisms. According to these advocates, non-fundamental preferences are idiosyncratic to some small group of investors, known as noise investors. A *noise investor* is an investor who makes decisions based on signals that are independent of assets' fundamental values.[43] The cleverness of tickers is apparently one such signal. In contrast, a *fundamental investor* is an investor who makes investment decisions based on their best estimates of the fundamental value of stocks.[44] In a market where fundamental investors outnumber noise investors, fundamental investors will sooner or later cancel the influence noise investors made in the market.

---

[43] There are a few different possibilities behind why an investor would make decisions based on noises. They could have mistaken noises with true signals. For example, they may have thought that the cleverness of tickers is a genuine proxy for the stock's fundamental value. Or, they could simply have non-fundamental preferences, e.g., a preference to hold stocks with clever tickers. Such preference could be well-reflected or subconscious. The definition of a noise investor does not distinguish between these possibilities, since all that matters is the investor's behavior when analyzing the mechanism of financial markets.

[44] What I call fundamental investors is often called a "rational investor" in the finance literature. I use fundamental investor instead to avoid confusions.

In our example, when clever-ticker investors push up the prices of clever-ticker stocks, fundamental investors will notice that those stocks are priced above their fundamental values. They will correct this mistake with a short position on these stocks or a long position on other stocks, which are undervalued in comparison. Fundamental investors will make a profit when the prices get corrected. According to this story, prices may temporarily depart from fundamental values under the influence of noise investors but will eventually get corrected by fundamental investors.

You may ask what if the market is populated by noise investors? Advocates of market mechanism would answer that it will not happen in the long run. Because (1) financial assets cannot deviate from their fundamental values forever and (2) noise traders will lose money and eventually get wiped out while fundamental investors prosper.

The first point is most obvious in the case of bonds. The price of a bond will return to its fundamental value by fiat at its maturity. When a bond matures, investors will know for sure whether the issuer is solvent or not. As a bond approaches its maturity, more information becomes available, and we expect its price to increasingly approximate its fundamental value. However, this point is less clear in the case of stocks. The investment horizon of stocks is much longer than bonds. One can only be sure of a stock's fundamental value when the company goes bankrupt or gets liquidated. At any time point before that, it's up to the market to decide. Optimists often appeal to a quote from Abraham Lincoln, saying that "you cannot fool all of the people all of the time". Eventually, prices will return to fundamental values.

The second point is an evolutionary point derivative from the first point. If the first point is true, noise traders will eventually run out of their luck and leave the market, while fundamental traders, who follow the "only logical approach"[45] to investment will prosper.

While the market mechanism certainly contributes to the realization of FV-efficiency, it has several apparent weaknesses. Even if the market prices reflect fundamental values in the long run, it may not motivate people to become fundamental investors. Investors may be better off in the short run by following the market trends. Some famous proverbs known among investors say, "trend is your friend", "do not short a bubble, ride a bubble". As John Maynard Keynes famously put, in the long run, we are all dead. We want financial markets to be effective in a timely manner, not only in the long run. Even if we only focus on bonds, whose prices by fiat will return to their fundamental values, a belated correction could turn out disastrous, like in the 2008 financial crisis.

As to the second point, Stiglitz (1989) used a quote commonly attributed to G. T. Barnum, "a fool is born every moment", to highlight the possibility that new noise traders will enter the market as existing noise traders get wiped out. There is no guarantee that the market is dominated by fundamental traders at any specific time point.

Overall, it's unlikely that the ideal of FV-efficiency can realized by market mechanism alone. Regulations are needed for the ideal of FV-efficiency to be achieved in the presence of unfettered individual preferences and sometimes irrational market participants. In the next section, I will give a brief discussion of market regulation from a general philosophical point of view.

---

[45] Buffett, ibid.

3.3.4 Regulation for FV-Efficiency: a note from social epistemology

Financial regulation is an important yet complicated topic that has been studied extensively by economists, legal scholars, government officials, etc. Proper treatment of the topic is clearly beyond the ambition of this paper. Instead, I want to make a general remark from the viewpoint of social epistemology to warn against some potential conceptual pitfalls. The message in a nutshell is that it's wrong to assume that only fundamental investors can improve the FV-efficiency of financial markets. Non-fundamental investors may be able to improve (as well as sabotage) the FV-efficiency of financial markets.

A widely discussed proposal, which gets implemented by the French government in 2012, is a tax on financial transactions, known as a Tobin tax. Tobin tax was originally proposed by James Tobin (1972) as a transaction tax on currency exchange to stabilize exchange rates after the collapse of the Bretton Woods system. Later, Stiglitz (1989) and Summers & Summers (1989) proposed to adopt a similar transaction tax in other financial markets to improve market efficiency. The motivation of this tax is rooted in the mechanism described above concerning noise and fundamental traders. Both Stiglitz (1989) and Summers & Summers (1989) suggest that a transaction tax will discourage noise traders from participating in the market without too much influence on fundamental traders. A Tobin tax thus helps increase the percentage of fundamental traders in the market. This is known as *the composition effect* of the Tobin tax. Ross (1989) and Schwert & Seguin (1993) argue that a Tobin tax will decrease market quality and thereby sabotage market quality. This is known as *the liquidity effect* of the Tobin tax.

Whether either of these effects is real and whether a Tobin tax in fact improves the overall market quality are questions that can only be answered by empirical tests. Colliard & Hoffmann (2017) argues that the French Tobin tax substantially decreased market liquidity without

improving market (IA-)efficiency, by comparing market data before and after the implementation of the tax in 2012. The empirical study of Tobin tax is clearly beyond the scope of this paper. Instead, I want to highlight a potential conceptual pitfall in regulating financial markets based on a well-known lesson in social epistemology.

Some advocates, including Stiglitz (1989) and Summers & Summers (1989), seem to assume that only fundamental investors can help improve the aggregation of fundamental information in financial markets, which explains why they emphasize the importance of the composition effect of a Tobin tax.

Some recent development in social epistemology, however, shows that the epistemic goal of a group is sometimes best served by individuals who are otherwise motivated. The most famous example in philosophy of science comes from the study of the credit economy of science. There, it's argued that the advancement of science, i.e., the pursuit of truth, is often best served by scientists who are motivated by the pursuit of credit and truth, instead of those who only care about truth (Goldman 1992; Kitcher 1993; Strevens 2003, Zollman 2018).

Mayo-Wilson, Zollman & Danks (2011) calls the general phenomenon where rational individuals can form irrational groups and that, conversely, rational groups might be composed of irrational individuals "the independent thesis". Financial markets are indeed perfect examples of the independent thesis. While the epistemic function of financial markets is to aggregate, compress and communicate information, no market participant entered the market for this purpose. They either seek returns from investment or want to use the market to hedge risks they are exposed to. However, scholars of financial markets may still fall into a similar conceptual pitfall as those who assumed that the quest for truth is best achieved by scientists who only care about truth. They may

assume without reflection that the aggregation of fundamental information in financial markets is best achieved by those who make their decisions based on fundamental information.

Farboodi & Veldkamp (2020) presents a way how non-fundamental investors, who make their decisions based on non-fundamental information, can help aggregate fundamental information in the market. They developed a model that includes both noise investors and rational investors. Their noise investors behave in the same way as the noise investors I defined above. Namely, noise investors invest based on signals that are independent of assets' fundamental values. On the other hand, a rational investor does not have to be a fundamental investor. Instead, a rational investor can choose to observe fundamental signals that help estimate the fundamental values of investments, or to observe demand signals that help estimate the noise investors' behaviors. While rational investors who choose the first option are called fundamental investors, those who choose the second option are "demand-driven investors". In actual financial markets, the role of demand-driven investors is played by high-frequency trading companies, who process order flow data and execute large volumes of trades within very short time frames.

While there are many interesting things to be learned from Farboodi & Veldkamp (2020), I only describe how demand-driven investors contribute to the market's FV-efficiency in their models. When there is a price movement in the market, demand-driven investors know better whether it's caused by noise investors or fundamental investors. If the movement comes from noise investors, they will move against the trend since they know that it's a mere noise that doesn't reflect the asset's fundamental value. On the other hand, when the price movement is caused by fundamental investors, they will free-ride fundamental traders' insights in the asset by following this trend. Basically, in this model, demand-driven investors remove noises caused by noise traders

and free-ride fundamental traders' information advantage.[46] Demand-driven investors thus contributed to the improvement of the FV-efficiency in the model.

But we shouldn't be too optimistic about the influence of demand-driven investors either. The model in Farboodi & Veldkamp (2020) has the critical assumption that the behaviors of noise investors are independently distributed across time, which is common knowledge of all market participants in the model. This means that demand-driven investors know that they cannot profit by following the market movements created by noise investors. If we assume that the noises are serially correlated, which seems to be true in actual markets, demand-driven investors may decide to follow and amplify the noises for profit instead of removing them. This will make the market more volatile and less FV-efficient.

Overall, markets are complicated institutions. Recent literature on financial regulations is mostly empirical, while actual regulations are often politically driven. Proper theoretical analysis is rare in recent literature. I hope that my discussion in this section at least gestures in a direction where theoretical analysis could contribute to the literature on financial regulation.

3.4 Is the Ideal Adequate? Challenges from ESG Investing

While the clever-ticker effect motivates us to reflect on how to better achieve FV-efficiency in financial markets, it does not constitute a genuine challenge to the ideal itself. After all, not too many people really think that investors' clever-ticker preferences should be reflected by efficient

[46] Kervel and Menkveld (2019) shows that demand-driven traders in actual financial markets exhibit similar behaviors.

prices in an ideal market. In this section, I discuss a more pressing challenge to the ideal of FV-efficiency presented by the recent movement of ESG investing in stock markets.

Environmental, Social, and Corporate Governance (ESG) investing is a recent world-wide movement in stock markets. The movement of ESG investing encourages investors to invest in companies that are environmentally sustainable and socially responsible. Several different ratings are developed to evaluate the environmental sustainability and the social responsiveness of companies. Stocks with high ESG ratings are known as ESG stocks. Under the influence of the movement, ESG stocks are valued significantly higher than non-ESG stocks globally at this moment.

It's far from guaranteed that a sustainable and responsible company will succeed and generate profits in the future. In fact, being sustainable and responsible often comes at a price that the company must pay. This creates at least a superficial tension between the movement of ESG investing and the ideal of FV-efficiency. To think through this tension, I will start with what motivates ESG investing.

3.4.1 What Motivates ESG Investing?

Researchers divide the motivations for ESG investing into *pecuniary* and *non-pecuniary* ones. An investor may decide to invest in ESG companies out of pure pecuniary reasons. Non-ESG companies are exposed to more environmental and regulatory risks. When environmental risks, e.g., global warming, unfold and governmental regulations become stricter, one may expect ESG companies to outperform non-ESG companies and thereby deliver better returns. Pecuniary ESG investors do not directly care about the ESG-related impacts made by a company. The ESG ratings of companies are only proxies for their fundamental value. Pecuniary ESG investors are

just fundamental investors who take ESG factors into consideration when estimating a company's fundamental value.

An investor could also be motivated by non-pecuniary reasons that have little to do with companies' fundamental values. An investor may simply care about the environment (and society, etc.) and therefore care about the ESG impacts of companies. They want to support ESG companies by holding their stocks. We may call these investors "altruistic ESG-investors", whose preferences for ESG stocks are independent of the stocks' fundamental values and even independent of their own consumptions (narrowly defined). An altruistic ESG-investor does not care if ESG companies will outperform in return in the future.

Another non-pecuniary motivation for ESG investing is known as the warm glow effect in economics. The warm glow effect originally comes out of the economic literature on donation. According to this effect, some people decide to donate to charities because donating evokes an emotional reward or creates a positive social image for the donor. The same idea can be applied to ESG investing. Warm-glow ESG-investors may not directly care about the ESG impacts of companies or expect higher returns from ESG stocks. They decide to invest in ESG companies because it makes them feel good or look good. The warm glow effect is probably more common than many would expect, especially among institutional investors. Recently, many universities announced that their endowment funds will not be invested in non-ESG companies under pressure from their students and society. Similar social pressures may exist for other institutional investors, e.g., pension funds or sovereign funds. These institutional investors, motivated by social pressures, are thus warm-glow ESG investors.

It's debatable whether the motivation of a warm glow is pecuniary or not since such investors are indeed motivated by an interest in their own consumptions (of positive emotional

feeling or positive social image). But the use of terms doesn't really matter for our purpose here. I put the warm-glow motivation and the altruistic motivation under the same label "non-pecuniary" because they both create a challenge for the ideal of FV-efficiency. Unlike pecuniary ESG investors, non-pecuniary ESG investors (including both altruistic and warm-glow ones) exhibit a non-pecuniary (and thus non-fundamental) ESG preference over stocks that is independent of the stocks' fundamental values.

Measuring the actual presence of (non-pecuniary) ESG investors and (non-pecuniary) ESG preferences in markets is difficult since it's difficult to know what investors are thinking when they make decisions. Yoo (2022) argues that both pecuniary and non-pecuniary motivations partly explain the current pricing of ESG stocks. In reality, what motivates an investor to invest in ESG companies is often a mix of the three (if not more) motivations. It's safe to assume that the current pricing of ESG stocks reflects both investors' expectation of their future performance and investors' non-pecuniary (and thus non-fundamental) preferences of ESG stocks.

3.4.2 New Ideal for New Paradigm

The presence of non-pecuniary ESG preferences in market prices is an anomaly to FV-efficiency just like the actual presence of clever-label preferences. Unlike the clever-label effect, it creates a genuine challenge to the ideal of FV-efficiency because many are inclined to think that such preferences should be reflected in efficient prices in an ideal market. In the rest of the paper, unless otherwise specified, I will focus on *non-pecuniary* ESG investors and *non-pecuniary* ESG preferences since they create challenges for the ideal of FV-efficiency.

Many advocates of ESG investing claim that ESG investing is not only a re-discovery of some factors that should be considered in estimating stocks' fundamental values. It's a paradigm

shift. From now on, investors must not only care about the returns of their investments but also their associated social and environmental impacts. If so, the ideal of FV-efficiency must be revised. An ideal market under this new paradigm will not only reflect the fundamental values (defined as rational expectations of discounted future cashflows) of financial assets, but also their associated ESG impacts.

This proposal will not only revise the ideal of information efficiency, but also goes against some traditional wisdom of economics. The environmental and social impacts of a company are typically considered as externalities to its production in economic models. The textbook response to externalities in economics is to internalize it via proper regulations. If the government can properly subsidize companies with positive ESG impacts and penalize companies with negative ESG impacts, the ESG impacts made by a company will be reflected in its future dividends. The ESG attributes of a company are thus internalized as fundamental information that contributes to the determination of its fundamental value. Investors will not have to treat ESG attributes of a company differently from other fundamental information, since ESG-related considerations are taken at a regulation level. The ideal of FV-efficiency can hence be saved.

This traditional solution of externalities is often found too unrealistic. Governmental regulations are often slow and undermotivated. The proposed paradigm shift of ESG investing clearly comes out of dissatisfaction with the current state of governmental regulations. According to its advocates, under the new ideal, companies will be held accountable for their social and environmental impacts. Since the compensations to executive officers of publicly listed companies are often tied to their stock prices, this new paradigm will motivate companies to operate in more sustainable and responsive ways. It will also provide ESG companies easier access to funding, and thereby guide the development of technologies toward a more ESG-friendly direction.

This new paradigm of ESG investing clearly requires market participants to have some specific preferences, namely, non-fundamental preferences over ESG investments, at an individual level. These individual preferences, together with other ESG information, are then aggregated in financial markets. One may reasonably worry how this ideal could be realized, due to theoretical and practical concerns in controlling individual preferences. This will be an interesting question for another paper. This paper studies the financial markets as an epistemic institution. According to the new paradigm, the financial markets are supposed to also aggregate, compress, and communicate ESG information about companies, in addition to their fundamental values. (People will likely redefine what fundamental value is after the shift is completed. For now, let's stick to the notion of fundamental value as defined.) In the remainder of this paper, I reflect on some of the epistemic consequences of this new paradigm.

3.4.3 Are Financial Markets suitable for processing ESG information?

There are two types of ESG information that should be reflected in stock markets according to the new paradigm: the ESG impacts of companies and the (non-pecuniary) ESG preferences of investors.

The traditional ideal of FV-efficiency allows (and in fact requires) ESG impacts to be reflected in prices *qua* predictors for fundamental values. The new paradigm of ESG investing in addition requires companies' ESG impacts to be reflected in prices for their own sakes, independent of their influences on companies' fundamental values. In the following, I exclusively discuss the consequences of requiring markets to reflect ESG impacts for their own sakes, since it is what is new of the proposed paradigm of ESG investing. Just like the price of a corporate bond

cannot be determined by its credit risk alone, the ESG impacts of investments cannot be priced without knowing investors' (non-pecuniary) ESG preferences.

In this section, I reflect on how financial markets could handle these two types of information. I argue that financial markets are capable of aggregating and reflecting companies' ESG impacts under proper regulations but will face certain challenges in processing the (non-pecuniary) ESG preferences of investors. These challenges come out of some important differences between ESG preferences and risk preferences. I shall start my discussion with the ESG impacts of companies.

Evaluating the ESG impacts of a company is a difficult task that requires various kinds of expertise and detailed understanding of the company's operation. In stock markets, most investors rely on self-disclosed ESG reports and third-party ESG ratings to assess the ESG impacts of a company. In this sense, financial markets do not directly generate ESG information, but merely react to it. However, it would be wrong to assume that financial markets do not contribute to the generation of this information. The investors' demand for ESG information financially incentivizes companies and rating agencies to generate ESG information for higher stock prices or subscription fees. Many companies recently started to publish ESG reports under the influence of ESG investing. New rating agencies come into existence to meet investors' needs. Such financial rewards are often more effective than governmental regulations.

However, investors, while interested in companies' ESG impacts, often do not have the motivation or the ability to assess the quality of these reports and ratings. Similar problems concerning the interaction between experts and novices are discussed in the first paper of my dissertation. Self-disclosed ESG reports, as of now, are voluntary and do not require external auditing. Third-party ratings are not well regulated either. If we want the financial markets to

correctly reflect the ESG impacts of companies, governmental regulations over these reports and ratings are essential.

An investor may conduct independent ESG research. But it will only influence the market if it is publicly released or independently discovered by other market participants. This makes independent ESG research different in nature from independent fundamental research. An investor can benefit from her independent fundamental research, even if this research is not known to the public. For example, if an investor finds that a company will outperform/underperform market expectation in the next quarter through independent research, the investor can make a profit when the company indeed outperformed/underperformed, even if she kept this research all to herself. However, if an investor finds that a company is making more positive/negative ESG impacts than the market consensus, she will not benefit from this information unless it's made known to the public by herself, other investors, or the company itself. This may discourage investors from conducting independent ESG research. One potential remedy is to require companies to issue externally audited ESG reports, just like they must issue externally audited financial reports.

The movement of ESG investing has already incentivized listed companies and rating agencies to conduct more ESG research and make available more ESG information. With proper regulation over ESG-related rating and reporting, financial markets are capable of reflecting companies' ESG impacts, just like it's capable of reflecting fundamental information. What I find challenging for financial markets to handle is investors' (non-pecuniary) ESG preferences.

In order for the stock market to reflect companies' ESG impacts in prices for their own sakes (i.e., not as mere predictors of fundamental value), it must also reflect investors' (non-pecuniary) ESG preferences at the same time. This is analogous to the fact that credit risks cannot be priced without investors' risk preferences. However, there are two important differences

between (non-pecuniary) ESG preferences and risk preferences that make the former not appropriate to be aggregated through financial markets alone, or so I argue.

The first difference is that (non-pecuniary) ESG preferences have a moral dimension that is lacking in the risk preference of an individual or of the entire market. Suppose that the financial market as a whole suddenly changes its risk preference in either direction. This sudden change is guaranteed to create significant short-term and long-term consequences. In the short term, it will move the prices of safe and risky assets. In the long term, a more risk-averse market may suffer from slower growth while a more risk-seeking market will exhibit higher volatility. But there doesn't seem to be any direct and obvious moral problem in this change. This is not true for ESG preference. It would be morally wrong if society as a whole decides to ignore environmental issues all together. They would be collectively wronging future generations. This difference may not matter if we only expect the financial markets to aggregate whatever preferences individuals have. But it will be problematic if we expect more.

Financial markets, by nature, can only aggregate individual preferences as they are in a mechanical manner. It cannot alter or guide individual preferences in meaningful ways. In a society populated by individuals who do not care about ESG impacts at all, financial markets will simply reflect the society's disinterest in social and environmental matters and their ignorance of future generations. This goes against the original motivation of the new paradigm of ESG investing, the point of which is to invest in an environmentally, socially and morally responsible way. An alternative epistemic institution that is better at steering the society's ESG preference is democracy. In democracy, individuals discuss their views on how much one should care about ESG impacts with other citizens through various public forums. Such public debates provide opportunities for individuals to reflect and alter their preferences over ESG matters. Eventually,

individual preferences are aggregated through voting. Unlike financial markets, the institution of democracy not only mechanically puts all individual preferences together, but also facilities public discussion and deliberation, which are not possible in financial markets.

The second worry is that aggregating investors' ESG preferences in financial markets may increase market volatility. This worry concerns more directly the mechanical details of the market and market participants. A proper treatment will require developing a mathematical model. I shall only give it a sketch here.

Keynes famously compared financial markets to a beauty contest, where judges are rewarded for selecting the most popular faces among all judges, rather than those they may personally find the most attractive. Keynes' beauty contest certainly captures some aspects of actual financial markets, where many profit by following trends. The concern is that this aspect of financial markets will be amplified if efficient prices are supposed to reflect (non-pecuniary) ESG preferences.

Although actual investors are often motivated by a mix of pecuniary and non-pecuniary reasons, for our purpose, we may assume that some investors are pecuniary and only seek profits, while other investors are non-pecuniary and really care about companies' ESG impacts for their own sakes. Thus, only non-pecuniary ESG investors have non-fundamental ESG preferences. Pecuniary investors only invest in ESG stocks in seek of higher returns.

We have been separating information about ESG impacts and information about ESG preferences in our discussion. In actual markets, investors cannot always distinguish these two, among other factors. When an investor observes an increase in prices of ESG stocks, the observer must make her own judgment whether this increase comes from a change in the stock's

fundamental value, the stock's expected ESG impacts or investors' ESG preferences, among other factors.

The worry is that aggregating (non-pecuniary) ESG preferences in financial markets may create a positive feedback loop. Pecuniary investors, instead of reporting their own ESG preferences, participate in a contest of guessing others' ESG preferences. When pecuniary investors observe an increase in the prices of ESG stocks, they may attribute this change to the change of market's ESG preference and further push up ESG stocks' prices. Other pecuniary investors do not know if this increase comes from a genuine change in market's ESG preference or from movements of pecuniary investors. They may thus follow the lead, which further pushes up ESG stock prices. Non-pecuniary investors, seeking that the market's hype for ESG investment, may further increase their appetite for ESG investing. After all, there is no right answer to how much one should care about ESG investing. This obviously creates a positive feedback loop pushing up the prices of ESG stocks, which increases market volatility.

Similar effects do exist for other aspects of the market. Investors may similarly participate in a beauty contest in guessing the market's risk preference or simply what other investors think will outperform the market average. The concern is that such a beauty contest effect will be significantly amplified in the new paradigm. It's unlikely that a rational investor will become more risk-seeking when other investors do. But it's possible that a non-pecuniary ESG investor becomes more willing to reduce negative ESG impacts when she observes others doing so. After all, there is no correct answer in how much one should care about companies' ESG impacts. This dissimilarity between risk preference and ESG preference makes the beauty contest aspect of the market particularly concerning for the new paradigm.

It's unfortunately beyond the scope of this paper to give a thorough assessment of this worry, not to mention a thorough assessment of the new paradigm of ESG investing. In addition to leaving some suggestive remarks, I'd emphasize the need for such assessments, especially from an epistemic point of view. This new paradigm is clearly assigning some new epistemic tasks to financial markets. Whether financial markets are well suited to completing these tasks can only be determined through careful examination. To the best of my knowledge, an epistemic assessment of the new paradigm of ESG investing has never been conducted.

3.5 What's Next

In this paper, I put the ideal of information efficiency of financial markets under philosophical scrutiny. This ideal, albeit widely endorsed by investors and scholars alike, is rarely articulated and examined systematically. Financial markets are central epistemic institutions in our modern capitalist society. Individual and governmental decisions are often made based on information provided by financial markets. The proper functioning of financial markets is no doubt of central importance for the proper functioning of our society.

In the second half of this paper, I briefly touched upon two related questions: (1) how FV-efficiency can be achieved in presence of unfettered individual preferences and (2) whether the ideal of FV-efficiency is adequate. The first question has received extensive attention from finance researchers in the past. I hope that my writing will interest philosophers in thinking about this question. I believe that there are interesting lessons to be learned by both sides. The second question, to the best of my knowledge, has never been discussed in writing. While my writing is

preliminary and suggestive, I hope that it marks the beginning of my future work in thinking about the ideal of information efficiency.

Bibliography

Alonso, R., & Camara, O. (2016). Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory* 165, 672-706.

Anderson, Elizabeth (2003). Consumer Sovereignty vs. Citizens' Sovereignty: Some Errors in Neoclassical Welfare Economics. In H. Pauer-Studer & H. Nagl-Docekal (Eds.), *Freiheit, Gleichheit und Autonomie*. Verlag Oldenbourg.

Anderson, Elizabeth (2006). "The Epistemology of Democracy." *Episteme* 3(1): 8-22.

Anderson, Elizabeth (2008). An Epistemic Defense of Democracy: David Estlund's Democratic Authority. *Episteme*, 5(1), 129-139.

Anderson, Elizabeth (2012). "Epistemic justice as a virtue of social institutions." *Social Epistemology* 26(2), 163-173.

Aumann, R., & Brandenburger, A. (1995). Epistemic conditions for Nash equilibrium. *Econometrica: Journal of the Econometric Society*, 1161-1180.

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, 90(1), 34-45.

Baer, N., Barry, E., & Smith, G. (2020). The name game: The importance of resourcefulness, ruses, and recall in stock ticker symbols. *The Quarterly Review of Economics and Finance*, 76, 410-413.

Beatty, John (2006). "Masking disagreement among experts." *Episteme: A Journal of Social Epistemology* 3, no. 1: 52-67.

Brennan, G., & Lomasky, L. (1997). *Democracy and decision: The pure theory of electoral preference*. Cambridge University Press.

Brennan, J. (2012). *The ethics of voting*. Princeton University Press.

Chordia, T., Subrahmanyam, A., & Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, 58(1), 41-58.

Colliard, J. E., & Hoffmann, P. (2017). Financial transaction taxes, market composition, and liquidity. *The Journal of Finance*, 72(6), 2685-2716.

Collins, H., Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science* 32(2), 235–296.

Collins, H., Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Collins, H., Weinel, M. (2011). Transmuted expertise: How technical non-experts can assess experts and expertise. *Argumentation*, 25(3), 401.

Delcey, T., & Sergi, F. (2022). The efficient market hypothesis and rational expectations macroeconomics. How did they meet and live (happily) ever after? *The European Journal of the History of Economic Thought*, 1-31.

DeMarzo, Peter M., Ilan Kremer, and Andrzej Skrzypacz (2019). "Test design and minimum standards." *American Economic Review* 109(6), 2173-2207.

Dunleavy, P. (1997). A critique of pivotal choice theory. *L'Année sociologique (1940/1948-), Troisième série, Vol. 47, No. 2, Choix rationnel et vie politique*, pp. 55-83

Farboodi, M., & Veldkamp, L. (2020). Long-run growth of financial data technology. *American Economic Review*, 110(8), 2485-2523.

Feddersen, T., & Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica: Journal of the Econometric Society*, 65(5), 1029-1058.

Fricker, Miranda (2007). *Epistemic Injustice: Power and the ethics of knowing*. New York: Oxford University Press.

Gleeson, Simon (2018). *The Legal Concept of Money*. Oxford: Oxford University Press.

Goldman, Alvin (1992). *Liasons: Philosophy Meets the Cognitive Sciences*. Cambridge, MA: MIT Press.

Goldman, Alvin I. (1999). *Knowledge in a Social World*. New York: Oxford University Press.

Goldman, Alvin I. (2001). "Experts: Which ones should you trust?". *Philosophy and phenomenological research* 63(1), 85-110.

Goldman, Alvin I. (2010). "Systems-oriented social epistemology". *Oxford Studies in Epistemology* 3, 189-214.

Goldman, Alvin. I. (2011). A guide to social epistemology. *Social epistemology: Essential readings*, 11-37.

Goldman, Alvin. I. (2021). How Can You Spot the Experts? An Essay in Social Epistemology. *Royal Institute of Philosophy Supplements*, 89, 85-98.

Goodin, R. E., & Spiekermann, K. (2018). *An epistemic theory of democracy*. Oxford University Press.

Hardwig, John (1985). "Epistemic dependence." *The Journal of philosophy* 82(7), 335-349.

Hardwig, John (1991). "The role of trust in knowledge." *The Journal of Philosophy* 88(12), 693-708.

Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, 35(4), 519-530.

Head, A., Smith, G., & Wilson, J. (2009). Would a stock by any other ticker smell as sweet? *The Quarterly Review of Economics and Finance*, 49(2), 551-561.

Hedlund, J. (2017). Bayesian persuasion by a privately informed sender. *Journal of Economic Theory* 167, 229-268.

Herzog, Lisa (forthcoming). *Citizen Knowledge: Markets, experts and the infrastructure of democracy*. https://www.rug.nl/staff/l.m.herzog/citizen-knowledge.pdf

Hodgson, G. M. (1994). Some Remarks on 'Economic Imperialism' and International Political Economy. *Review of International Political Economy* 1(1), 21-28.

Hutchins, E. (1995). *Cognition in the Wild*. MIT press.

Huttegger, Simon M., and Michael Nielsen (2020). "Generalized learning and conditional expectation." *Philosophy of Science* 87(5): 868-883.

Kamenica, Emir, and Matthew Gentzkow (2011). "Bayesian persuasion." *American Economic Review* 101(6), 2590-2615.

Kitcher, Philip (1993). *The Advancement of Science*. New York: Oxford University Press.

Ladha, K. K. (1992). The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, 617-634.

Lederman, H. (2018). Uncommon knowledge. *Mind*, 127(508), 1069-1105.

List, C. and Goodin, R.E. (2001), Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy*, 9: 277-306. https://doi.org/10.1111/1467-9760.00128

Malkiel, B. (1992). Efficient market hypothesis. *New Palgrave Dictionary of Money and Finance*. London: Macmillan.

Martini, Carlo (2014). "Experts in science: a view from the trenches." *Synthese* 191(1): 3-15.

Mayo-Wilson, C., Zollman, K. J., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, 78(4), 653-677.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1), 5-32.

Miller, N. R. (1986). Information, electorates, and democracy: some extensions and interpretations of the Condorcet jury theorem. *Information pooling and group decision making*, 2, 173-192.

Mishkin, F. S. (2007). *The economics of money, banking, and financial markets*. Pearson education.

Muth, J. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica* 29: 315–335.

Myerson, R. B. (1998). Extended Poisson games and the Condorcet jury theorem. *Games and Economic Behavior*, 25(1), 111-131.

Nagel, Thomas (1978). *The possibility of altruism*. Princeton University Press.

Nash, J. (1951). "Non-Cooperative Games." *Annals of Mathematics* 54(2): 286–95.

Pedersen, L. H., Fitzgibbons, S., & Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2), 572-597.

Roemer, John E. (2019). *How We Cooperate: A Theory of Kantian Optimization*. Yale University Press.

Rogers, G. (2021). *Speculation: A Cultural History from Aristotle to AI*. Columbia University Press.

Ross, S. A. (1989). Commentary: Using tax policy to curb speculative short-term trading. *Regulatory Reform of Stock and Futures Markets: A Special Issue of the Journal of Financial Services Research*, 19-22.

Sargent, T. J. (2017). Rational Expectations. In *The New Palgrave Dictionary of Economics* (pp. 1-7). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-95121-5_1684-2

Schwert, G. W., & Seguin, P. J. (1993). Securities transaction taxes: an overview of costs, benefits and unresolved questions. *Financial Analysts Journal*, 49(5), 27-35.

Stiglitz, J. E. (1989). Using tax policy to curb speculative short-term trading. *Regulatory Reform of Stock and Futures Markets: A Special Issue of the Journal of Financial Services Research*, 3-17.

Strevens, Michael (2003). "The Role of the Priority Rule in Science." *Journal of Philosophy* 100(2): 55-79.

Summers, L. H., & Summers, V. P. (1989). When financial markets work too well: A cautious case for a securities transactions tax. *Journal of financial services research*, 3, 261-286.

Tobin, J. (1984). On the Efficiency of the Financial System. *Lloyds Bank Review*, July 1984. ISSN 0024-547X.

Udehn, L. (1996). *The limits of public choice: a sociological critique of the economic theory of politics*. Routledge.

Van Kervel, V., & Menkveld, A. J. (2019). High-frequency trading around large institutional orders. *The Journal of Finance*, 74(3), 1091-1137.

Veselova, Y.A. (2016). Computational complexity of manipulation: A survey. *Autom Remote Control*, 77, 369–388. https://doi.org/10.1134/S0005117916030012

Wang, W. K. (1985). Some arguments that the stock market is not efficient. *UC Davis L. Rev.*, 19, 341.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455-1508.

Young, H. P. (1988). Condorcet's theory of voting. *American Political Science Review*, 82(4), 1231-1244.

Yoo, Paul (2022). "ESG Investing: A Tale of Two Preferences". *Kenan Institute of Private Enterprise Research Paper No. 4069015*, Available at SSRN: https://ssrn.com/abstract=4069015 or http://dx.doi.org/10.2139/ssrn.4069015

Zollman, K. J. (2018). The credit economy and the economic rationality of science. *The Journal of Philosophy*, 115(1), 5-33.