The Erasmus Journal for Philosophy and Economics (EJPE) is a peer-reviewed bi-annual academic journal supported by the Erasmus Institute for Philosophy and Economics, Erasmus School of Philosophy, Erasmus University Rotterdam. EJPE publishes research on the methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of inter-disciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>.

EDITORS
Måns Abrahamson
Lennart Ackermans
Annalisa Costella
Erwin Dekker
Savriël Dillingh
Akshath Jitendranath
Vaios Koliofotis
Marina Uzunova
Erica Celine Yu

ADVISORY BOARD

EDITORIAL ASSISTANTS
James Grayot, Chiara Stenico
## TABLE OF CONTENTS

### ARTICLE

*I Choose for Myself, Therefore I Am:*
The Contours of Existentialist Behavioral Economics  
*MALTE DOLD AND ALEXA STANTON*  
[pp. 1–29]

### SPECIAL ISSUE

Introduction: The Philosophy and Economics of Pandemics  
*THE EDITORS*  
[pp. 30–32]

Neighbors Help in a Pandemic  
*NORA MILLS BOYD AND MATTHEW DAVIS*  
[pp. 33–46]

Mandated Shutdowns, the Ratchet Effect, and *The Barstool Fund*  
*JEFFREY CARROLL*  
[pp. 47–57]

Pandemic Windfalls and Obligations of Justice  
*BRIAN BERKEY*  
[pp. 58–70]

Governing Life and the Economy: Exploring the Role of Trust in the Covid-19 Pandemic  
*JOELLE M. ABI-RACHED AND ISHAC DIWAN*  
[pp. 71–88]

How Economists Ignored the Spanish Flu Pandemic in 1918–1920  
*MAURO BOIANOVSKY AND GUIDO ERREYGERS*  
[pp. 89–109]

Three Ways in Which Pandemic Models May Perform a Pandemic  
*PHILIPPE VAN BASSHUYSSEN, LUCIE WHITE, DONAL KHOSROWI, AND MATHIAS FRISCH*  
[pp. 110–127]

Uncertain Policy Decisions During the Covid-19 Pandemic  
*MALVINA ONGARO*  
[pp. 128–137]
How to Handle Trade-Offs in Pandemics

*KRISTER BYKVIST*  
[pp. 138–151]

Collective Responses to Covid-19 and Climate Change

*ANDREA S. ASKER AND H. ORRI STEFANSSON*  
[pp. 152–166]

Vaccine Refusal Is Not Free Riding

*ETHAN BRADLEY AND MARK NAVIN*  
[pp. 167–181]

INTERVIEW

Past and Future of Humanomics:  
A Conversation with Deirdre Nansen McCloskey

*DEIRDRE NANSEN MCCLOSKEY AND PAOLO SILVESTRI*  
[pp. 182–209]

INTO THE ARCHIVES

Mathematical Psychology

*JAN TINBERGEN*

*Translated and annotated by CONRAD HEILMANN, STEFAN WINTEN, RUTH HINZ, AND ERWIN DEKKER*  
[pp. 210–221]

No Envy: Jan Tinbergen on Fairness

*CONRAD HEILMANN AND STEFAN WINTEN*  
[pp. 222–245]

BOOK REVIEWS

Massimiliano Vatiero’s *The Theory of Transaction in Institutional Economics: A History*

*DOUGLAS W. ALLEN*  
[pp. 246–250]

Hélène Landemore's *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*

*ERICA YU*  
[pp. 251–257]

PHD THESIS SUMMARY

Galbraith’s Integral Economics (1933–1983)

*ALEXANDRE CHIRAT*  
[pp. 258–263]
I Choose for Myself, Therefore I Am: The Contours of Existentialist Behavioral Economics

MALTE DOLD
Pomona College

ALEXA STANTON
Pomona College

Abstract: Behavioral economics and existentialism both present informative perspectives on human choice. We argue in this article that the dialogue between the two approaches can enrich the current debate about the normative implications of behavioral economics. While behavioral economics suggests that our capacity to choose is constrained by cognitive biases and environmental influences, existentialism emphasizes that we can (and should) treat ourselves as free and ‘becoming’ beings in spite of the many constraints we face. Acknowledging these two perspectives in the form of a theoretical synthesis—which we propose to call existentialist behavioral economics—provides us with reasons why we should protect our choices ‘as our own’ and how doing so may be more difficult than we anticipate. It also provides a framework to analyze the threat of identity-shaping social and technological developments, such as preference-altering nudges and artificially intelligent prediction algorithms.

Keywords: agency, authenticity, behavioral economics, nudging, choice, existentialism, freedom, identity, meaning

JEL Classification: B41, D63, D91

I. INTRODUCTION
Behavioral economics has been steadily on the rise in recent decades (Geiger 2017). It has been quite successful in exposing many of neoclassical economics’ limitations in explaining and predicting human choice (Bernheim, DellaVigna, and Laibson 2018). More specifically, behavioral economics has helped to refine models of individual choice based on
empirical evidence that challenges the axioms of neoclassical rationality (Sugden 2018, 7–14). Since at least the publication of the bestseller *Nudge*, by Thaler and Sunstein in 2008, there has been a lively debate about the normative implications of behavioral economics. At its core lies the question of how economists can reconcile the insights about the ‘messiness’ of people’s revealed preferences with the logic of traditional welfare economics that assumes welfare to be the satisfaction of rational, well-ordered preferences. While the literature in *behavioral normative economics* is rich and diverse (for an overview, see Bernheim 2016, and Sunstein 2020), there are two main approaches.

The first approach, *behavioral welfare economics*, is most prominently advocated by Thaler and Sunstein (2008), as well as by Bernheim and Rangel (2007, 2009). It aims to define welfare as the satisfaction of preferences that are context-independent and time-consistent. While observed choice is sometimes not a proper indicator of individual welfare, and can be difficult to parse, behavioral welfarists assume that deep within people hold ‘true’ preferences; that is, preferences they would reveal if they “paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control” (Thaler and Sunstein 2008, 5). Such ‘true’ preferences are assumed to be well-ordered in that they fulfill the standard axioms of neoclassical rationality (Infante, Lecouteux, and Sugden 2016). It is the task of economists to identify choice frames in which individuals hold those ‘true’ preferences and take them as normative input for policymaking.

The second approach, which has been forcefully defended by Robert Sugden (2004, 2018), aims to respect individuals’ choices without referring to the underlying structure of preferences. It is grounded in the conviction that the behavioral welfarists lack a psychologically plausible theory that could explain the origin of rational preferences (Infante, Lecouteux, and Sugden 2016). Sugden argues for an alternative normative standard which he calls the *opportunity criterion*. This criterion says that, when comparing social states, economists should not be concerned with the rationality of individual preferences but should instead seek institutional arrangements that foster people’s opportunities, that is, the size of their choice sets.

Much has been written about the advantages and pitfalls of these two approaches (see, for example, McQuillin and Sugden 2012; Dold and Schubert 2018; Lecouteux and Mitrouchev 2021). In this article, we do not want to dissect the intricacies of those approaches. However, we want to point
out that both approaches suffer from a blind spot in that they do not systematically address the issue of agency freedom—individuals’ capability to make choices that they perceive to be authentic and meaningful. In doing so, both approaches underestimate the importance of the nature or quality of the choice process and, for instance, sidestep the important issue of whether a person has the capacity to reflect upon and possibly revise the values and beliefs that motivate her actions (Dold and Rizzo 2020). Also, the degree to which the social environment shapes and constrains individuals’ processes of ‘becoming’ is not sufficiently addressed in either approach (Dold, van Emmerick, and Fabian 2020). We think that this blind spot can be illuminated by the existentialist perspective we develop in this article. Such a perspective shifts the informational basis of normative economics away from welfare and opportunity toward authenticity and meaning. In doing so, we do not take a stance on the relative merits of behavioral welfare economics and the opportunity approach. Instead, our approach runs parallel to these as another ethical approach to behavioral economics.

Throughout the article, we treat existentialism as a branch of ethics that treats agency freedom and the quest for authenticity as central to human existence (Olafson 1967; Webber 2018). The notion of authenticity helps identify situations in which humans are ‘unfree’ (Golomb [1995] 2012; Crowell 2020). At its core lies the question: do I succeed in being the author of my life, or am I mainly a product of the situational framing or the wider social environment I find myself in? Taking this existentialist perspective can enrich positive economic theorizing by shedding light on a wide range of behaviors that traditional choice theory has trouble explaining, such as “mountaineering” (for the latter, see Loewenstein 1999, 318). In this article, we illustrate ways in which the existentialist

---

1 The capability approach of Amartya Sen (1985, 1999) is the closest approach in normative economics to the account we sketch in this article. Sen is famous for highlighting the important normative role of agency freedom and for arguing that a person’s sense of agency is both intrinsically valuable and of central importance to her well-being. We deviate slightly from Sen in our definition of agency freedom. For Sen, “[agency freedom] is the freedom to achieve whatever the person, as a responsible agent, decides he or she should achieve”. It refers to “what the person is free to do and achieve in pursuit of whatever goals or values he or she regards as important” (Sen 1985, 203). It is remarkable that the Senian perspective has gotten little traction in the debate on the normative implications of behavioral economics in recent years. For exceptions, see Davis (2011, chap. 8) and Dold, van Emmerick, and Fabian (2020). For an excellent philosophical discussion of Sen’s idea of agency freedom, see Binder (2019).

2 By ‘becoming’, we mean the process through which a person’s preferences change over time. Depending on her level of self-consciousness, a person might or might not be aware of her own process of ‘becoming’.
perspective can enrich the discussion about the *normative* implications of behavioral economics.

Our argument is based on insights from empirical psychology that highlight individuals' inherent desire for *authentic, meaningful choices* which, in turn, has a considerable impact on human behavior and economic activity (Loewenstein 1999; Karlsson, Loewenstein, and McCafferty 2004; Phelps 2013). Authentic and meaningful choices are those that we can feel are in line with our developing values and the narratives we are constructing of our lives. There is evidence that a person's perception of meaning and authenticity is essential for her physical and mental well-being (Ryan and Deci 2006, 1566). When people perceive that they fulfill meaningful roles in their families and communities that align with their overall values, they show a higher ability to cope with uncertainties, withstand unhealthy temptations, endure hardship, and work more steadily toward difficult goals (Routledge and Bitzan 2020, 2).

Some readers might wonder why we have chosen to bring together such disparate strands of literature. There are two primary reasons we believe that a dialogue between *existentialist philosophy* and *behavioral economics* is fruitful. First, we think existentialism and behavioral economics are effective complements. Both begin with the same object of investigation, *choice under uncertainty*. Behavioral economics uses the tools of decision theory and existentialism uses philosophical reasoning. In doing so, the starting point for both is the acknowledgment that human decision-making is bounded by unknown outcomes and ambiguous probabilities. Moreover, where existentialism may sometimes assume an extreme conception of individual freedom (Webber 2018, 2), behavioral economics often leans toward emphasizing our cognitive flaws and the situational influences that trigger “systematic mistakes” (Bernheim and Rangel 2007, 16). Hence, the two have much to gain from each other. A combined perspective allows us to conceptualize individuals who are, at the same time, internally and situationally constrained, but who are also striving for freedom and self-realization.

Second, existentialism and behavioral economics occupy parallel positions within their respective disciplines. Just as behavioral economics emerged as a response to the formal rigidity of rational choice axioms in neoclassical economics, existentialism can be seen as an antagonist to

---

3 In the existentialist tradition, the idea that decisions must be made without any clear understanding of the consequences is poignantly captured by the statement: ‘Life can only be understood backwards, but must be lived forwards’ (a quote often attributed to Søren Kierkegaard).
analytic philosophy’s focus on formal properties of language (as exemplified in Wittgenstein’s *Tractatus Logico-Philosophicus*) and the idea that “logical analysis is the main business of philosophy” (Bertrand Russell, quoted in Jones 2009, 13). Existentialists like Heidegger were seeking “to free philosophy from logic” (Matthews 2003, 171).4 Both Heidegger and the French existentialists (de Beauvoir, Camus, Merleau-Ponty, Sartre) wanted to shift philosophy’s focus away from ‘narrow’ problems of language and logic to ‘wider’ reflections of the actual *lived experience* of people in an uncertain world (Kaufman 1956, 12). Central to existentialism is the thought that people do “not have an inbuilt set of values that they are inherently structured to pursue. Rather, the values that shape a person’s behavior result from the choices they have made” (Webber 2018, 4). This is very similar to the idea in behavioral economics that people don’t ‘come to market’ with a set list of master-preferences but often construct them in the process of choosing (Lichtenstein and Slovic 2006).

The structure of this article is as follows. In section II, we discuss some of the core tenets of existentialist philosophy: the notions of freedom, agency, and authenticity. Section III sketches the contours of what we call *existentialist behavioral economics*. We explore both the work of earlier behavioral economists (in particular, Kahneman and Tversky) and a newer, more philosophically inclined branch of behavioral economics. Section IV presents the two core elements of existentialist behavioral economics: constrained freedom and the pervasive influence of the social environment on individual choice. Sections V and VI illustrate practical implications of existentialist behavioral economics.

II. EXISTENTIALISM: A BRIEF OVERVIEW

Existentialism evades precise definition, particularly because so many of its leaders refused to accept the label (Kaufman 1956, 37). Existentialism’s very DNA contains within it a resistance to labels; the existentialists celebrate individuality and self-definition, to lump them into a single mass is to violate their very credo (Irwin 2015, 11). They also frequently disagreed with each other and even with their past selves. For example, both Sartre and Heidegger experienced a “turn” later in life, by which their philosophies transformed into something very different from their original iterations (Irwin 2015, 14). While Sartre’s *Being and Nothingness*

---

4 Kaufman (1956, 12) goes even further and identifies as the “heart” of existentialism an anti-system sensibility, a protest against academic philosophy that is caught in the “iron cage” of reason which fails to capture humans’ lived experience.
clearly takes inspiration from Heidegger's *Being and Time*, Heidegger is famously quoted as exclaiming of *Being and Nothingness*, “How can I even begin to read this rubbish!” (Bakewell 2016, 201). The two also held radically different political views, as did many of the philosophers within the tradition.

Despite all of this, we will use the term ‘existentialism’ as a blunt instrument. When we reference what ‘the existentialists think’, we mean the recurring patterns of existentialist thought, at all times asking our readers to keep in mind that existentialism is a colorful, diverse set of ideas. The reason we have chosen to do it this way is because, despite their disagreements, the existentialists share fundamental, core ideas and it is this core to which we wish to attach ourselves.\(^5\)

### II.I. The Existentialist Core

The existentialists are united in rejecting a perspective in which philosophers are seen as neutral observers who seek a general, formal account of what it means to be human (Crowell 2020). The existentialists “are opposed to any system which abstracts from the hopes and fears of the individual or which attempts to fit man into some sort of impersonal schematism” (Roberts 1952, 469). Existentialism is grounded in *lived experience* and in the acknowledgment of the individual holding an essentially subjective, not a bird’s eye view. In existentialist thought, “genuine philosophizing must well up from a man's individual existence and address itself to other individuals to help them to achieve true existence” (Kaufman 1956, 23). Thus, it is a philosophy designed for one’s life as one leads it, diverging significantly and intentionally from the rigid theorems and tight structures of logical-formal inquiry.

To this end, one hallmark of existentialist writing is that much of it is done through story. Nietzsche and Kierkegaard rely on metaphor and allegory, Sartre wrote many plays, and Camus is perhaps most famous for his novel *The Stranger* ([1942] 1970). People understand their life through narrative, through the “drama of decision-making” (Harari 2018, 55), and existentialists employ literature to explore this human drama, to spotlight subjective experience and to give voice to those parts of our lives which we do not fully understand. The existentialists remind us vigorously that

---

\(^5\) Since the purpose of this section is to focus on existentialism’s core tenets, we will not enter into subtleties about different strands within existentialism. See Kaufman (1956), Crowell (2012), and Flynn (2009) for excellent discussions of commonalities and differences between existentialism’s key thinkers.
“man is not primarily a thinker! […] He is a volitional actor, a being who makes choices and lives by them” (Killinger 1961, 306).

Existentialists explore possible responses to the hardest challenge of what it means to be human: to live with the recognition of death, the absurdity of life, and the “benign indifference of the universe” (Camus [1942] 1970, 120), which does not reveal any objective meaning to us. This thought is powerfully articulated by Albert Camus in his famous article, The Myth of Sisyphus ([1942] 1955). Camus compares the human condition to Sisyphus’ fate: each of us is consigned to push our ‘burden’ up a hill only to lose control of it once again, a task void of meaning. Further, if Sisyphus turns inwards, he does not discover any immutable moral law that could serve as a substitute for life’s absurdity (Webber 2018, 14). Paradoxically, this situation creates a radical freedom for him: if neither life’s meaning nor his own nature is fixed, he has the freedom to make himself, to become different from what he is through his own choices. Sisyphus cannot change the task assigned to him by the gods, but he is still free to choose what those circumstances mean to him. The world will not become meaningful on its own nor will it listen to his laments. But he can become resolute, accept ‘the courage to be’ and march up the hill ‘in spite of’ the awareness of his finite existence and the seemingly absurd situation he is thrown into (Tillich 1952). Consequently, rather than giving in to dark cynicism or acquiescence in light of the universe’s indifference, existentialism can be seen as a philosophy of exploration, creativity, and energetic resistance. Existentialists contrast this emphasis on self-exploration with potential forces of unfreedom, showing that in order to be free, we must understand the ways in which we are not (Roberts 1952, 471). The notion of authenticity plays an important role in clarifying situations of ‘unfreedom’.

II.II. Unfreedom and Authenticity
The French existentialists are famous for placing freedom and responsibility at the core of their philosophies. But it was German philosopher Martin Heidegger who particularly emphasized the notion of authenticity. In his 1927 magnum opus, Being and Time, Heidegger describes what the

---

6 Khawaja argues that it is actually “the idea of personal authenticity [which lies] at the center of existential thought” (2016, 24). Having given up the philosophical quest for objective goals, existentialists think it is still possible for an individual to make a distinction between living ‘as herself’ and living ‘in the eyes of others’. In this sense, living authentically means to follow “the norm of self-identity, tied to the project of self-definition through freedom, choice, and commitment” (Crowell 2020).
individual is and what she must resist in order to become her authentic self. Heidegger uses his own terminology, unsatisfied with the limits imposed by a language laden with preconceptions. Thus, he labels his conception of individuals’ existence as “Dasein”, loosely translated as “being-there” (Bakewell 2016, 60). When Heidegger references ‘being-there’, he refers to a type of Being far and above our typical use of the word. Lowercase being generally refers to any living thing; but Being means something much more expansive. For the purposes of this article, Being can be thought of as ‘consciousness’.

By ‘being-there’, Heidegger refers to an existentialist concept, that of being “situated” or “embodied” (Crowell 2020). Various existentialist thinkers have their own interpretations of this concept, but the general thrust is that individuals are, in Heidegger’s terms, “always already”: we are always already within a situation or circumstance and cannot detach our experience, our identity, or our perceptions from this situated position (Wrathall 2002, 220). We belong to a world that precedes us, and we are always enmeshed within it. There is no “view from nowhere” (Nagel 1986), no detached objectivity available to us (Bakewell 2016, 65).

Heidegger argues that constraint is a necessary part of what it means to be human and, consequently, understanding constraints is necessary to live authentically. Part of those constraints is “das Man”, translated as “the they” (Bakewell 2016, 78) or “the one” (Dreyfus 1995, xi). ‘The one’ is a term that describes situations in which a person follows common practice—she does what ‘one’ does; for instance, the person could be ‘eating as one eats’ (for example, with fork and knife) or ‘driving as one drives’ (for example, on the right side of the road). As both examples indicate, conforming to ‘the one’ is not necessarily negative. It can save time and energy if a person does not have to determine how to go about routine tasks. Moreover, conforming to ‘the one’ fulfills a positive coordinative function as it establishes reliable expectations of what other people do. It can, however, become problematic when a person does not just conform to common practices in certain parts of her life, but when she uncritically internalizes the norms and values of her social surroundings. Then, conformity becomes conformism; her actions are succumbing to the roles and patterns laid down by ‘the one’.

For Heidegger, ‘the one’ is not merely external pressure (for example, social conventions) but also something that lives within us (for example,
emotions such as self-doubt or shame) that we must struggle against (Bakewell 2016, 78). In order to achieve authenticity, we must distinguish the line between the functional aspects of ‘the one’ and the conformist influence that lives both inside and outside us. Authentic Being is not a refusal of ‘the one’, it does not demand that we extract ourselves from our emotions and our socio-cultural embeddedness. Instead, it demands that a person develop an attitude toward and consciously commit to certain elements of herself and her socio-cultural surrounding that are meaningful.

Consider the following example: a person who simply does action $\phi$ because this means following the prevailing norm in her peer group—‘it is what one does’—is living inauthentically. However, if this very same person does $\phi$ because it reflects values she identifies with, then—following Heidegger’s logic—we can say that she acts authentically. In the latter case, the person will perceive doing $\phi$ as a meaningful act, as being in line with who she is or wants to be. If a person has the capacity of committing to authentic choices on a continuous basis, she is depicting agency freedom. In this sense, agency freedom requires an acknowledgment of the socio-cultural constraints, but demands also that we rebel against these limitations when they threaten our capacity of making authentic choices. For existentialists, the rebellion is particularly important when the constraints are the result of ourselves or those around us, rather than the result of an indifferent universe we cannot control.8

III. TWO STRANDS OF BEHAVIORAL ECONOMICS

While behavioral economics has been instrumental in exposing many of neoclassical economics’ limitations in explaining and predicting human behavior, we argue in this section that the mainstream view within behavioral economics, which we will call strand one, is still limited by a mechanistic view of individuals. However, a new subfield of behavioral economics, which we will refer to as strand two, has begun to adopt a more nuanced vision of individuals qua economic agents. While still limited by the

8 The existentialists do not pretend that following this call for personal authenticity is easy. Accepting responsibility for oneself and defining this self through free choice in the face of an unstructured, unconcerned universe can be paralyzing. In fact, this is the basis for Sartre’s novel Nausea ([1938] 2013): the main character finds himself confronting the contingent quality of life and experiences an overwhelming sensation of sickness because of it (Bakewell 2016, 100–105). Yet, one must be willing to confront these fears, otherwise they will forever be suffocated by the constant assault of life against them, never finding fulfillment.
logic of economic models, this more recent strand of behavioral economics privileges a dynamic view of choice and identity, which has productive commonalities with existentialist thinking.

**III.I. Neoclassical Economics and Its Discontents**

Neoclassical economics has been the dominant economic paradigm for decades and can be characterized by the combined assumptions of maximizing utility, market equilibrium, and stable preferences, which are “used relentlessly and unflinchingly, [and] form the heart of the economic approach” (Becker 1976, 5). Individuals are assumed to be rational; that is, they act according to a set of “integrated preferences” (Sugden 2018, 7) that are stable and internally consistent. Integrated preferences are assumed to be context-independent, meaning they are unchanged by situational influences, such as informational framing or fleeting emotions (Sugden 2018, 5). This understanding of the individual is obviously simplistic, yet the simplicity is believed to be a large part of its appeal. Neoclassical “parsimony” (Tirole 2002, 636) makes modelling easier in that it abstracts from random deviations in individual rationality, which is a crucial virtue of economic models that seek to give a general account of average human behavior in market settings. Behavioral economists argue, however, that neoclassical economics’ explanatory and predictive power is handicapped by its misunderstanding of individuals’ systematic deviations from rationality. In his magnum opus, *Thinking, Fast and Slow*, Daniel Kahneman (2011) details the core results of his research with Amos Tversky on systematic biases and heuristics in decision-making, instructing the reader to imagine that our minds operate under two (metaphorical) mental systems, System 1 and System 2. System 1 thinks fast. It jumps to conclusions and, as a result, makes biased judgments. Yet, it allows us to react quickly and to conserve cognitive energy in situations that cannot be solved with hedged bets and rough estimates. System 2 thinks slow. It demands more of our cognitive resources but enables us to solve complicated problems, which require care and deliberation.

One example Kahneman uses to illustrate System 1’s hasty thinking is the anchoring effect. It describes the tendency of individuals to base their estimates on an available cue (the ‘anchor’), for example, an initial piece of information or a random number. Kahneman describes an experiment in which participants were asked to guess whether the value of an item $X$ was higher or lower than a number provided by the experimenter. The item $X$ included the height of Mt. Everest or the number of member
states in the United Nations. The median guess of participants who were shown a high initial anchor was much larger than that of participants shown a lower anchor (Jacowitz and Kahneman 1995, 1163). Such an effect occurs even when the anchor is completely unrelated to the estimation task. Critcher and Gilovich (2008) found that estimations of athletes’ performance can be anchored by their jersey’s number and that sales forecasting is influenced by a product’s model number. Despite the fact that the anchor is unrelated to the task in both cases, System 1 still grabs onto it as an unconscious shortcut and systematically biases participants’ guesses.

The anchoring effect is merely one of many ways that individuals do not act like the optimal problem-solvers neoclassical economics assumes them to be. Other systematic deviations from rational choice include people’s assessment of gains and losses as to reference points, their over-weighing of small and under-weighing of large probabilities, and their choices being affected by how options are framed (Kahneman 2011). In general, behavioral economics in the tradition of Kahneman and Tversky has been a powerful rebuttal to the neoclassical assertion that perfectly rational, utility-maximizing agents are good proxies for average human decision-making.

### III.II. Strand Two Behavioral Economics

In recent years, a second type of behavioral economics has emerged that puts more emphasis on preference change and social—not just situational—aspects of decision-making: “strand two behavioral economics” (Hoff and Stiglitz 2016, 28). It explicitly models the individual as an en-culturated actor who is shaped by her own choices and socio-cultural surrounding. If the main goal of Kahneman and Tversky’s program (‘strand one behavioral economics’) was to contest neoclassical hyperrationality, then strand two’s goal is to contest neoclassical stasis.

In neoclassical economics, “neither individuals’ utility functions nor their preferences are changed by choices they make” (Davis 2003, 49). The self is what it is, and it stays that way: “In the standard analysis of choice,

---

9 Hoff and Stiglitz state that strand two behavioral economics “is an attempt to broaden economic discourse by importing insights into human behavior not just from psychology, but also from sociology and anthropology” (2016, 25). We want to point out that they particularly stress the idea from sociology that the social structure influences individual identity. However, they do not investigate to the same degree the question of how individual action transforms social structure. For a discussion of a comprehensive sociological approach to identity, see Davis (2011, 79–81, section 4.3.1).
individuals can never choose anything other than their most preferred option as determined by their preferences and feasible set. Preferences, then, determine choice” (Davis 2003, 49). In this sense, behavioral change is understood as the outcome of the mechanistic interplay between one’s stable preferences and changes in prices or income.10 There is no room to account for a person’s freedom to hold different preferences in the future.11

While ‘strand one behavioral economics’ of Kahneman and Tversky allows for distorting factors in the choice process (for example, anchoring or framing affects), it still subscribes to a static view of human nature: individuals are ultimately constituted by their core of ‘true’ preferences. The benchmark for ‘good’ choice is the satisfaction of those true preferences that are assumed to be stable and integrated (Rizzo and Whitman 2020, 80). It neither explains the psychological foundation of those ‘true’ preferences, nor does it allow for the possibility that a person’s ‘true’ preferences are endogenous to her social environment or that they may change over time (Infante, Lecouteux, and Sugden 2016). This means that traditional behavioral economics portrays a simple, dualistic view of what constitutes human choice: there is a stable inner ECON that is struggling to come to the surface, but situational influences (such as arbitrary anchors or informational framing) are throwing HUMANS off course.

Hoff and Stiglitz (2016) contrast this static understanding with strand two thinking which seeks to explore how social identity impacts and is impacted by economic behavior. They begin with an idea familiar

---

10 Neoclassical economists often defend a weak version of preference stability, that is, preferences are assumed to be stable for the time period for which economists analyze individual behavior (see, for example, Varian [1987] 2014, 118). While some models assume stable life-time preferences (for example, life-cycle models of consumption and saving), many economists do not defend this stark notion of preference stability. However, while being an important nuance, we think this weaker version of preference stability still misses something crucial: neither neoclassical economics nor strand one behavioral economics models individuals’ processes of becoming different nor acknowledges the inherent endogeneity of human preferences. While they allow for ad hoc preference changes, their models typically do not explain mechanisms through which people develop different preferences over time. The existentialist perspective we stress in this article acknowledges the ubiquity and importance of preference change.

11 Some might argue that freedom means nothing but acting according to one’s given preferences. However, we deem this a philosophically unsatisfying understanding of freedom. Accepting such an understanding of freedom, it seems difficult to grasp situations in which, upon reasoned reflection, an individual wants to hold different preferences in the future. One can introduce stable meta-preferences that help explain the motivation for this preference change, but this just pushes the mechanistic interplay up one level of abstraction. For a discussion of the problems of this conceptual move, see Lewis and Dold (2020).
to traditional behavioral economics: priming. They discuss an experiment in which the participant’s prime is their own identity. A group of bankers were told they would be playing a nation-wide coin-flipping game and were instructed to flip a coin and report the results, with each “winning toss” worth $20 (Hoff and Stiglitz 2016, 34). However, some of the bankers were primed beforehand to recall their role as banker while others were not. Individuals who were primed to think of their banker identity acted less ethically than those who were not primed. The researchers explain, “the context of the moment of decision [the identity prime] influences choices even when the context should be transparently irrelevant to the decision” (Hoff and Stiglitz 2016, 26). Thus, in this case, the way that individuals thought of themselves, that is, how close-at-hand their identity as banker was, altered their choice behavior. Put existentially: how subjects believe that they ‘should’ act is shaped by their understanding of how ‘one’ acts in a certain role (here: the role of a banker).

While this experiment applies notions of identity to economics, it still does not represent a significant departure from traditional behavioral economics. Yet Hoff and Stiglitz go further, suggesting that identity is not only relevant as an external, temporary influence on choice but as an internal, long-lasting one as well, manifesting through “cultural mental models” (2016, 25). This position is a radical departure from the typical economic perspective. To evidence their point, Hoff and Stiglitz cite an experiment performed in India where both low-caste individuals and high-caste individuals were asked to play a two-player assurance game in which each anonymous player chooses to ‘hunt hare’ or to ‘hunt stag’ over a series of rounds. Hunting hare offers a low reward consistently, regardless of the other player’s action, whereas hunting stag offers a high reward only if the other player also chooses hunting stag. Low-caste individuals were more likely to choose stag even after being ‘burned’ in the first round by the other player choosing hare. Alternatively, high-caste individuals were more likely to switch to hunting hare after an unsuccessful first round, exhibiting a “retaliatory” response (Hoff and Stiglitz 2016, 39). The authors posit that this discrepancy can be explained by high-caste individuals choosing to protect their ego and sense of status, noting that “a survey provided supporting evidence: high-caste men were much more likely than low-caste men to think it appropriate to beat up an individual who, because of adverse or ambiguous circumstances, had caused one a loss” (Hoff and Stiglitz 2016, 39).
This research implies that the social (not just the situational) context, in which identity is expressed, affects the choices people make. Hoff and Stiglitz elaborate on this idea by proposing a model of the self as a set of potential selves, each of which will be differentially triggered according to context. While a person may become more like one of these identities over time, she will not solidify into a self that is more ‘real’ or ‘true’ than the other selves she could have become. Her identities and preferences are not subject to some ultimate destination, though they are moderated by previous choices and the circumstances in which the person finds themselves.

If we left the discussion here, strand two behavioral economics would seem as deterministic as strand one; it only ascribes a greater role to a person's inculcated identity in shaping her choice patterns. However, Hoff and Stiglitz argue that the reverse is also true: an individual's choices will influence her identity. They theorize that individuals engage in a process of “learning by consuming”, in which current consumption influences downstream choice (2016, 31). By this process, preferences are not static; rather, old preferences inform the development of new preferences, and the choices a person makes in the past will necessarily affect the ones they make in the future. For instance, a person who starts consuming jazz music will end up with a preference for live jazz concerts that she did not have at the outset. This means that strand two behavioral economics acknowledges that preferences are not only endogenous to the social environment somebody is enculturated in, but that they are also changing as a function of one’s past choices.

IV. Elements of Existentialist Behavioral Economics
Strand two behavioral economics paints a more dynamic picture of individual choice and identity that comes closer to the existentialist understanding we sketched in section II of this article. However, it is important to note that it is not yet describing the freedom and agency existentialist philosophers had in mind. While Hoff and Stiglitz make clear that in their

---

12 Admittedly, the difference between the situational and social context is somewhat blurry. What we mean by situational context is the formal, abstract properties of a decision situation (such as the wording of a choice problem). In contrast, the social context of a decision situation is characterized by socially constructed identities, social interaction, internalized norms, etc. Heidegger’s 'being-there' (see section II.II) can be understood as ‘being situated’ in a social context. This broader understanding of ‘being situated’ comes close to the conceptualization of a situation in modern sociology (see, for example, Collins 2014, 3–7).
account “[man] is a social being. He is social not just in what he does, but also in the lenses through which he perceives himself and the world”, they also argue that “[man] does not choose the lenses that he uses. They are socially acquired and socially activated” (2016, 39). Their account underscores how impactful the social environment is for the development of individual identities. Yet, it does not allow for individuals to step back from their environment and their past choices and rank the various selves or social identities they are exposed to. This is a result of the fact that strand two’s conceptualization of the individual still conceives of social identities as parts of the individual’s utility functions. It does not leave room for intentional choices about what ‘type of utility functions’ a person wants to hold in the future. So, there is a tension between existentialism’s emphasis on active self-creation and strand two’s emphasis on the dominant influence of the mental model (‘the lens’) that comes with one’s social environment. In what follows, we want to sketch the contours of existentialist behavioral economics that build on elements of Hoff and Stiglitz and combine them with the economic approach of a pioneer of strand two thinking, James M. Buchanan.

**IV.I. Constrained Freedom**

According to Buchanan ([1979] 1999), it is essential that economists acknowledge that real-world people learn, grow, and seek to change themselves. Buchanan ([1979] 1999, 247) criticizes traditional economics for not acknowledging change as the individual’s defining feature, arguing that people are crucially motivated by ‘becoming’: individuals have the ability to shape themselves and make choices in pursuit of the identity they wish to achieve. They may not always have the same preferences, or even understand fully what these preferences are, so they will value the freedom to pursue these preferences as they arise and shift. Part of this venture involves for Buchanan a recognition that individuals cannot always be best understood through their utility functions or given preferences. By this view, individuals are defined, at their very core, by an ironic lack of definition. They are uncertain and opaque, even to themselves, and always in flux.13

---

13 More recently, Hargreaves Heap has argued that “it is not just that we have well-defined preferences but do not reveal them because our choices are guided by heuristics and biases; we sometimes simply do not have the well-defined preferences in the first place” (2017, 253). For example, some preferences are determined through the process of choice itself (Slovic 1995; Lichtenstein and Slovic 2006).
Like in existentialist philosophy, the individual in Buchanan's framework has the freedom to self-create but they are not unconstrained, free to become anything. Instead, Buchanan points out, humans are “natural and artifactual” ([1979] 1999, 247), both determined and free. As natural beings, we are limited by the confines of our body, our family, our culture, and our circumstances. Yet, we are also artifactual; we can construct ourselves, pursue different ends, change our mind, resist, and engage in processes of becoming different. Because individuals are not entirely determined, there exists “a large set of possible persons that one might imagine himself to be, or might imagine himself capable of becoming” (Buchanan [1979] 1999, 250). People are not subject to any teleology but are instead the agents of their own experiences and identities; their intentional choices direct them into one or another version of themselves. This point adds an important nuance to strand two behavioral economics in that it argues that an individual’s self-identity is the result of their purposeful choices that impact and are impacted by their environment.

**IV.II. The Influence of ‘The One’ and the Power of System 2 Thinking**

Buchanan acknowledges that individuals’ capacities to reflect upon and direct their own processes of becoming are shaped by their socialization and upbringing. He ascribes an important role to education. It can “provide persons with both an array of imagined prospects and some means of valuation” (Buchanan [1979] 1999, 254) that lie at the heart of the capacity to form one’s own identity. However, Buchanan believes that adults, by and large, possess the necessary motivational and cognitive means to intentionally choose different paths of becoming (Dold 2018). In doing so, he might have underestimated the powerful role of cultural lenses and cognitive models individuals often adopt unconsciously and uncritically—a feature that both strand two behavioral economics and existentialist philosophy highlight.

If it is true, as Hoff and Stiglitz (2016, 51) argue, that our environment has the power to become a dysfunctional or dominating influence on our processes of identity formation, then existentialists are right to warn us from das Man. As mentioned above, we don’t want to reduce das Man to its negative dimension only. There can be situations where das Man helps us coordinate our actions with others and see the world through a window of collectively shared norms (think, for instance, of the shared norms of the scientific community). However, we think that strand two

---

14 The idea that humankind has no teleology is a core feature of Sartre’s argument in *Existentialism Is a Humanism*, see Sartre ([1946] 2007).

15 As mentioned above, we don’t want to reduce das Man to its negative dimension only. There can be situations where das Man helps us coordinate our actions with others and see the world through a window of collectively shared norms (think, for instance, of the shared norms of the scientific community). However, we think that strand two
the identity that their context brings forth from them. If one wishes to self-create and avoid passively assuming the “they-self” (Bakewell 2016, 78), then individuals must cultivate their own self-understanding, uphold the capacity to creatively imagine a different future for themselves, and critically assess the ways in which their environment presses upon them. As Hoff and Stiglitz argue, the experience we have in the world is the result of the lens through which it is filtered. But as Buchanan argues, we also have the freedom and the burden to build ourselves the lens that will help us best cope with our world and allow us to choose the life that we find meaningful.

How can we shape this lens? We think that behavioral economic research gives us a partial answer, in particular Kahneman’s distinction between ‘thinking, fast and slow’. The existentialist individual, though free and self-created, will not be immune to impulsive reactions or to expressions of self which may not be entirely her own. To resist the world all the time would be exhausting and lead to diminishing returns, especially in cases where das Man fulfills socially productive, coordinative functions. Thus, social influence cannot and should not be banished entirely, and it will likely emerge most frequently through the operation of the fast and unconscious System 1. In order to cultivate a more secure, self-directed identity, the existentialist individual can look to her slower System 2. Here, the individual can learn to prepare for and moderate the outbursts of System 1 such that slowly, with time, she will learn to identify and reconsider das Man, channelling it to the extent that it is conducive to her own agency freedom.

Some might argue the radical freedom proposed by existentialists is nothing but the philosophical embodiment of neoclassical rationality. However, we think that such a position ignores that System 2 thinking does not need to coincide with integrated preferences; it only describes the mode of critical-creative reasoning, not its outcome in terms of formal properties of an individual’s choice patterns. Authentic choices require becoming aware of oneself and the world around us, it is the ability to step back from one’s ‘given’ preferences by means of System 2 thinking. Recalling Hoff and Stiglitz’s mental models, identity colors the world we see as well as our participation in it; if we want to make choices that are meaningful and authentic, we must be aware of the kind of lens we are behavioral economics can help us identify situations where das Man over-determines individual choice and suppresses an individual’s awareness of what it means to choose creatively and freely.
working with and aim to transform that lens in cases where we sense a lack of agency freedom.

V. TWO APPLICATIONS: NUDGING AND ARTIFICIAL INTELLIGENCE
We believe that the dialogue between strand two behavioral economics and existentialist thinking we sketched in the previous section leads to a more realistic understanding of individual choice, which stresses both the normative significance of and practical obstacles to agency freedom. As we argue in this section, that dialogue can also help us understand some of the dangers in recent policy innovations like nudges and artificial intelligence.

V.I. Existentialist Behavioral Economics and Nudging
The most prominent example of behavioral welfarism is the policy paradigm of libertarian paternalism which aims to honor freedom of choice while helping individuals satisfy their ‘true’ preferences (Thaler and Sunstein 2008). It does so through the use of ‘nudges’, that is, alterations in decision frames that influence behavior by making certain choices easier or more salient. For example, the manager of a college cafeteria might decide to put apples on an eye-level shelf at the cashier and candy bars on a top shelf in the back of the room in order to nudge students into picking the apple. While nudges can be an effective tool to achieve given policy goals, such as achieving a higher degree of sustainable energy consumption (Kaiser et al. 2020) or tax compliance (Holz et al. 2020), from an existentialist perspective, there is reason to question the underlying narrative that tries to justify interventions on the grounds of ‘true’ preferences.

One such reason is that, as we have already discussed, individuals may not even have given preferences, instead using their choices to experiment and determine what to value (Hargreaves Heap 2017). Thus, “the challenge is to find economic institutions that can be recommended to individuals who do not know what they prefer”, and whose so-called ‘true’ preferences are frequently opaque (Sugden 2018, 15). One solution behavioral welfare economists have used to remedy this lack of information and assess whether or not an enacted policy (like a nudge) actually satisfied an individual’s ‘true’ preference is the “as judged by themselves” (AJBT) criterion (Paul and Sunstein 2019). AJBT states that an action taken to improve someone's welfare can be judged as successful or not successful only from the point of view of the targeted individual. Since the nudge
might alter a person’s revealed preferences, Paul and Sunstein (2019) advocate that nudged individuals must assent to the nudge post hoc; that is, after being nudged, they approve of the outcome achieved. For example, if an individual is automatically enrolled in a 401(k) retirement program and then, after the fact, reports she is satisfied with this outcome, then the AJBT criterion deems default-enroll justified.

Yet this solution is imperfect, particularly because people become different in the process of choosing. Choices are ‘transformative’ (some more than others), and this quality threatens AJBT’s applicability. If the individual post-choice is much changed from the individual pre-choice, then post-choice self-evaluation will not capture the opinions of the pre-choice self. Imagine someone trying to decide whether or not to have a child. If they become a parent, they will be fundamentally changed and now occupy a different identity than if they remained childless. In one sense, the difference between the pre-knowledge and post-knowledge selves may not matter; if the pre-choice self no longer exists, perhaps the only experience that matters now is that of the post-choice self. However, this does not address the issue in its entirety because even if the post-knowledge self is satisfied, they cannot know the experience of the counterfactual childless scenario and therefore do not know if their choice was the optimal one. This is of course an inherent feature of individual decision-making in an uncertain world, and its philosophical intricacies may not be of much significance for personal well-being. However, when economists make policy recommendations for a large population (for example, in terms of incentivizing having offspring through child allowances), those philosophical intricacies need to be addressed since the standard for decision-making has to be higher in politics, and preference change has severe consequences for the application—and policy prescriptions—of stated preference approaches in cost-benefit analysis.16

V.II. Sisyphus Confronts AI

Artificial intelligence (AI) is a broad term, referring generally to machines that can perform problem-solving tasks at or above the level of a human (Bringsjord and Govindarajulu 2020). In this section, we want to focus on narrow AI. This is the AI used to predict our choices and recommend things to us, whether in the form of products for purchase, video

---

16 For a discussion of ways to address preference change in cost-benefit analysis, see Adler and Posner (2000, 1114ff.). See Mitrouchev and Buonomo (2020) for a recent discussion of the issue of multiple-selves models in normative economics and the question of which self in time should be given normative authority.
streaming, or partners on dating sites. Prediction algorithms may seem a small issue relative to hot-button issues in artificial intelligence like the future of work, the ethics of self-driving cars, or the defense against a superhuman consciousness. However, from an existentialist behavioral economic point of view, these technological questions should not obscure AI’s immediate, everyday impact on our agency freedom. In fact, we believe that prediction algorithms’ apparent low risk and widespread appeal makes recognizing their potential pitfalls all the more pressing.

The primary appeal of artificially intelligent prediction algorithms is that they can track a myriad of past individual choices, making connections and recognizing patterns that individuals themselves cannot remember. Moreover, prediction algorithms can make use of ‘datification’, which compounds observations about one individual with the observations of millions of others.\(^{17}\) In this way, AI can identify the beats per minute a person exercises to most effectively, locate exactly which sub-sub-genre of comedy makes her laugh the hardest, or expose destructive health habits she never recognized herself. AI also saves transaction costs associated with choice, like the time and effort required to gather information and compare options. In this sense, AI frees up cognitive resources by outsourcing elements of a laborious decision-making process. Consequently, a large part of AI’s allure is that it unburdens us from many of the responsibilities of choice, making our lives simpler, easier, and more comfortable.

However, as pointed out by existentialism, choice is how individuals navigate the world and construct their identities; if they are not making their own active choices, do they relinquish the very means they have for creative personal growth, undermining their own agency? Considering the logic of existentialist behavioral economics, we think that there are (at least) two reasons for why the answer might be ‘yes’.

The first reason is that artificially intelligent algorithms do not just help us satisfy our given preferences in a more effective way. Instead, especially in those situations where individuals do not have well-defined preferences yet, there is a risk that we substitute our own preference formation process with that of an AI algorithm. This substitution may be benign or even beneficial so long as it is accompanied by individuals’

\(^{17}\) ‘Datification’ refers to the process of “collecting and analyzing data about Internet users” and “feeding such data back to users, enabling them to orient themselves in the world” (Kennedy, Poell, and van Dijck 2015, 1).
System 2 thinking and active choosing. However, we can identify Heidegger’s logic of *das Man* in modern algorithms. AI is powerful in part because it uses the will of the crowd to determine and inform our particular preferences. Based on data from millions of users, the algorithm tells us what ‘one’ should choose in a certain area of life, for example, what ‘one’ should do to become fit or healthy.\(^{19}\) Strand two behavioral economics shows us that this problem is a real one: we are very much susceptible to ‘the herd’. Our preferences and identities frequently reflect the environment we find ourselves in and we adopt a self that accords with the other selves we see. This is by no means a new phenomenon, but the addition of AI expedites and expands the traditional process. Where before preferences spread locally, they now spread globally; where before they spread in months or years, they now spread in days or weeks. There will be less space for differentiated local pockets of taste and opinion to emerge and challenge each other; instead, preference adoption can become much more universal and, therefore, much more threatening to authentic individual choices.

The second reason is that the delegation of active choice to AI algorithms might lead to an erosion of our ability to experiment with different identities and reflect on our preference formation process. AI’s personalized recommendations may stunt individual growth by biasing us toward preferences we revealed in the past. Challenging our own prevailing tastes and opinions can be difficult and uncomfortable, often requiring conscious effort on the part of the individual. If AI is designed to give us what we craved in the past, then it will likely reinforce and amplify specific elements of our identities. In addition, the fewer choices we make for ourselves, the less trained we are in making intentional choices and engaging in System 2 thinking. We can think of an individual’s capacity to make active choices as a muscle that atrophies in the long run if not used frequently enough (Rizzo and Whitman 2020, 252–253; Schubert 2021).

While an individual might actively choose to restrict her choice set knowing that choice overload could impede processes of self-experimentation, a systematic outsourcing of active choice via AI might make it more likely in an event that requires serious effort and perspective-taking.

---

18 Grayot (2020) argues that System 2 thinking might also be susceptible to post-hoc rationalizations and other miscarriages of reasoning. So even this point is contestable.

19 As a relatively benign example, consider Google’s algorithm, which uses artificial intelligence to determine which search results to prioritize. This is necessary to Google’s functionality but leads to an inevitable framing such that Google has considerable influence over what qualifies as first-page material and, consequently, over what we all see.
that individuals will not be equipped with the skills of *creative imagination* and *evaluation* that Buchanan ([1979] 1999, 254) identified at the heart of an individual’s authentic process of becoming. As AI gets better at knowing what individuals will likely consume, individuals will have fewer incentives to explore beyond their comfort zone and try something unfamiliar or untested, yet this is precisely how humans learn and improve. There is neurophysiological evidence that we learn predominantly from making mistakes, not from confirmed expectations (Schultz 2016). And in line with the existentialist credo that *lived experience* lies at the heart of personal growth, individuals need “to feel responsible for [their] actions if those actions are to be a source of learning” (Hargreaves Heap 2017, 257). If we allow artificial intelligence to do our learning for us, we’re not really learning, nor will we be prepared for the problems AI cannot solve for us.²⁰

**VI. AGENCY-FRIENDLY INSTITUTIONS**

Admittedly, to claim one’s agency is difficult. One reason is, as strand two behavioral economics shows, we are constantly pressed upon by situational and environmental influences. Existentialist behavioral economics does not dispute that these constraints exist or are meaningful. It rather argues that we, both as individuals and policymakers, should think about strategies to reduce their dominance. Family, upbringing, culture—we cannot escape any of them. Yet this is exactly the existentialist point: a fundamental retreat into social isolation is neither possible nor desirable, but an existentialist economic lens can help us identify both enabling and oppressive forms of identity-shaping influences. Furthermore, it can help us envision institutional reforms—for instance, in education or work environments—that foster the idea of individuals as claimants of their own freedom.

Consider the example of ‘boosting’: boosts have been proposed as an alternative to nudging. Unlike nudges, which are goal-directed and exploit behavioral biases (for example, the status quo biases in the case of defaults), boosts “foster people’s competence to make their own choices—that is, to exercise their own agency” (Hertwig and Grüne-Yanoff 2017, 202).

²⁰In addition, some critics who believe that only narrow AI is possible base their arguments on existentialist grounds (see, for example, Wheeler and Di Paolo 2014). They argue that AI cannot simulate the ‘being-there’ part of the human experience. And it is this lack of capacity to truly mimic the human experience that makes AI decisions alien to the authenticity of the individual.
In general, boosts are interventions that focus on creating and promoting individuals’ cognitive and motivational competencies. They either target competencies directly (for example, by teaching decision strategies or changing procedural routines) or indirectly by altering the choice environment (for example, by changing the informational representation). Furthermore, instead of externally redesigning the choice environment for individuals based on an agent’s assumed ‘true’ preferences, boosts provide individuals with the ability to redesign their proximate choice environment (Samuli and Hertwig, forthcoming). For instance, agency freedom in cyberspace can be boosted when users are given the option to see the full history of a post or to customize how their news feed is designed and sorted (Lorenz-Spreen et al. 2020). Such boosts can increase agency freedom by helping individuals become aware of the power of situational influences not just in cyberspace, but in other areas of their lives as well.

If strand two behavioral economics is right in emphasizing that cultural and environmental influences are powerful identity-shaping forces in our lives, then social institutions play a large role in determining the kind of people we become. In acknowledging the reflexivity between social structure and individual preferences, the existentialist perspective shifts the institutional focus from outcomes to the rules and processes that performatively shape choices. At the same time, the existentialist perspective we have described in this article asks from economists a certain degree of epistemic humility. Hargreaves Heap rightly notes that economists cannot know in advance in any detail what agency freedom will consist of for a particular person; it is an individual capability, “a state that does not depend on some set of final outcomes (like the experience of friendship or love), it is a state that depends on the character of the route taken to it” (2020, 117). Institutions play an important role in determining ‘the character of the route’. They can foster individual self-reflection and active choice, or they can nudge and steer people. The existentialist perspective we developed in this article defends the former and challenges the latter. Agency-friendly institutions “would seem naturally to be concerned with the conditions (for example, the educational system, the media, the family, the vibrancy of the arts world) that support reflection on what preferences to hold” (Hargreaves Heap 2013, 996). Naturally, the repeated exchange of goods and services on markets can also

21 While Sunstein (2013) makes the distinction between means-oriented and ends-oriented paternalism and argues that only the former is legitimate, in practice, it is often difficult to discern the difference between means-paternalism and ends-paternalism.
contribute to self-reflection and individual processes of preference learning. By means of ‘experiments of living’, markets can be seen as arenas that foster agency freedom and complement the aforementioned socio-cultural institutions (Delmotte and Dold, forthcoming).

VII. CONCLUSION

Artificial intelligence and nudging are but the newest in a long line of technological innovations that promise to help individuals make better choices. They offer to unburden us from a laborious and error-prone decision-making process. Existentialist behavioral economics can provide us with a conceptual approach that helps us assess what the impact of those external decision devices might be. This approach does not call for us to become luddites that shun social and technological innovations altogether. Instead, it invites us to weigh the costs and benefits of delegating active choices. Choice is difficult, and nudging and AI purport to assume that burden for us. But if we use them unreflectively and unflinchingly, we do so at a potential cost to our agency freedom.

We hope that philosophically inclined economists and philosophers of economics find some merit in the dialogue between existentialism and behavioral economics we sketched in this article. We believe that the core tenets of existentialism can add to contemporary behavioral economic thinking by offering a richer understanding of human choice that conceptualizes individuals as being in flux and motivated by their quest for meaning and authenticity. At the same time, we think that behavioral economics does provide the existentialist perspective with a much-needed empirical grounding that rightly highlights how pervasive the influence of situational factors and cultural mental models on human choice is. The integrated perspective of existentialist behavioral economics conceptualizes individuals as beings who are constrained but at the same time motivated to develop preferences that they can fully embrace as their own and whose satisfaction leads to meaningful choices. Such a perspective invites behavioral economists to move away from efficiency and preference satisfaction as the benchmarks of policymaking. Instead, it opens a much-needed discussion about the character of institutions that constrain or enable individuals’ agency freedom.
REFERENCES


Institute for Global Innovation and Growth Research Brief, June 2020. North Dakota State University, Fargo, ND.


Malte Dold is an Assistant Professor in the Economics Department at Pomona College in California. Previously, he spent two years as a post-doctoral fellow at New York University. He holds a master’s degree in Philosophy and Economics from the University of Bayreuth, and received his PhD in Economics from the University of Freiburg. His research lies at the
intersection of behavioral economics, philosophy of economics, and history of economic thought. 
Contact e-mail: <malte.dold@pomona.edu>

Alexa Stanton graduated from Pomona College magna cum laude in 2020, with a major in Philosophy, Politics, and Economics (PPE), and a minor in Computer Science. In her thesis, Alexa examined the links between existentialism and behavioral economics, and how these links can inform perspectives on artificially intelligent prediction algorithms. She now works as a consultant at Transom Consulting Group. 
Contact e-mail: <alexastanton@gmail.com>
Introduction: The Philosophy and Economics of Pandemics

In response to the ongoing Covid-19 pandemic, the *Erasmus Journal for Philosophy and Economics (EJPE)* invited scholars to reflect on the philosophy and economics of pandemics, in general, and on the current pandemic, in particular. We welcomed short, focused contributions—from methodological, ethical, public-policy, and historical perspectives—that target particular aspects of the pandemic, or of pandemics generally, and that articulate a single, incisive idea.

The result is this special issue, comprising ten articles—four by special invitation (Joelle M. Abi-Rached and Ishac Diwan, Krister Bykvist, Andrea S. Asker and H. Orri Stefánsson, Ethan Bradley and Mark Navin) and six through open submission. All accepted articles went through our external peer-review process.

This issue is organized around four central themes. The first theme deals with the *effects* of the pandemic—such as vulnerabilities at the individual level (Nora Mills Boyd and Matthew Davis) and at the national level (Jeffrey Carroll)—and the *obligations* that arise from them (Brian Berkey). Did the deterioration of non-market, neighborhood relationships exacerbate the negative impact of the pandemic? *Boyd and Davis* argue that it did, which is all the more reason, they conclude, to cultivate fair-weather local networks. Crises, Robert Higgs once observed, are a sort of ‘ratchet’—they propel an indefinite expansion of government authority. But there is a way of escaping the current pandemic’s ratchet, *Carroll* argues, by substituting private charitable funds for government subsidies. While Carroll’s argument for private charitable funds is of a consequentialist bent, a different pandemic effect, *Berkey* argues, creates—on a broader range of views about distributive justice and at least for some individuals—obligations to donate: some have benefited from economic ‘windfalls’ and that benefit grounds a corresponding obligation.

A second set of papers deals with *responses* to the pandemic—those by governments to the current pandemic (Joelle M. Abi-Rached and Ishac Diwan) and those by economists to past pandemics (Mauro Boianovsky and Guido Erreygers). *Abi-Rached and Diwan*’s contribution is motivated by an empirical puzzle: while the current discussion of public responses...
to the pandemic is often grounded in a trade-off between ‘lives and livelihoods’, the data suggests that this trade-off does not exist between countries. Trust in government, Abi-Rached and Diwan argue, has a role in explaining this puzzle and they elaborate, more precisely, what this role may be. Boianovsky and Erreygers are motivated by a different puzzle: in stark contrast to the overwhelming response of contemporary economists to the current pandemic, the ‘silence of the economists’ during the 1918–1920 Spanish flu pandemic seems deafening. After documenting the phenomenon, the authors propose a number of answers to the question: ‘Why this silence?’.

Third, a set of methodological papers reflects on how we ought to think about and model the pandemic—they deal with performativity (Philippe van Basshuysen, Lucie White, Donal Khosrowi, and Mathias Frisch), uncertainty (Malvina Ongaro) and how uncertainty complicates the moral trade-offs we face (Krister Bykvist). Performativity is front and center in van Basshuysen, White, Khosrowi, and Frisch’s contribution. Epidemiological models, they argue, are just as performative as economic models. This might complicate forecasting, but, at the same time, it opens up new evaluative criteria—that is, epidemiological models can be assessed not just epistemically, but also on their behavioural impact. Uncertainty, Ongaro argues, is not a unitary concept—the pandemic has revealed and exacerbated at least three types of uncertainty we may face. And there is good democratic and epistemic reason to conclude that this, in turn, calls for a much more inclusive type of collective decision-making. Given such varied uncertainty—and fundamental ethical disagreement—how should policymakers handle difficult moral trade-offs? Not by the usual ‘apply your favourite moral principle’ approach of moral philosophers, Bykvist argues; rather, we need to supplement democratically approved ethical platforms with ethical frameworks based on a more local, domain-restricted type of ‘moral modeling’.

Finally, a pair of articles warns against drawing quick similarities between features of the current pandemic and other phenomena, such as climate change (Andrea S. Asker and H. Orri Stefánsson) and free-riding on a public good (Ethan Bradley and Mark Navin). The large-scale coordinated response to the current pandemic, some commentators claim, is reason to be optimistic about a similar response to the threat of climate change. Not so fast, Asker and Stefánsson respond. Not only is the individual decision problem in the two cases significantly different, but also government responses to the current pandemic suggest that, if anything,
coordination in the collective action problem of climate change will continue to fail. Contrary to much current discussion, BRADLEY AND NAVIN argue, vaccine refusal is not a case of free-riding; (most) vaccine refusers, the authors contend, fail to satisfy both subjective and objective criteria of classic free-riders. And if that is the case, then policymakers need to treat vaccine refusers accordingly.

In the midst of multiple types of uncertainty, one thing, we believe, is certain—the discussion on the current pandemic, and its past siblings, will be ongoing. We hope readers enjoy these texts and take them as a starting point for continuing this discussion. In the meantime, we thank the authors and the reviewers for making this special issue possible. We further gratefully acknowledge the impeccable, as always, editorial assistance by James Grayot and Chiara Stenico. Finally, we thank the Erasmus Institute for Philosophy and Economics and the Erasmus School of Philosophy at the Erasmus University Rotterdam for their continued support of the EJPE.
Neighbors Help in a Pandemic

NORA MILLS BOYD
Siena College

MATTHEW DAVIS
Siena College

Abstract: The degradation of non-market relationships has rendered individuals unnecessarily vulnerable in disasters, including the global pandemic. While local networks of community-based aid that emerge in response to disasters improve the efficacy of response, they tend to be short-lived. This is unfortunate, since the existence and strength of such local networks prior to the onset of disasters not only boosts the efficacy of response but also contributes to the well-being of individuals and communities in non-disaster times. Therefore, individuals ought to establish and strengthen fair-weather local networks of non-market relationships—that is, cultivate neighbor relationships.

Keywords: pandemic, social capital, mutual aid, disaster, care, local networks

JEL Classification: A13, B54, D64, H12, H84

I. INTRODUCTION

It is said that neighbors used to visit each other in the evenings to tell stories (Berry [1988] 2017, 107). That neighbors would help one another to cultivate the land, to build. Such cooperation and neighborliness now strikes many as quaint. ‘Community’ today has been taken to mean something placeless, something digital (Bradshaw 2008). Geographically dispersed communities certainly have benefits—they are effective at connecting individuals in minority subcultures, crowdsourcing resources at immense scales, and building international social movements. Yet, living without strong local social networks in addition to these ‘post-place’ communities has serious downsides, as the global Covid-19 pandemic has rendered painfully clear. With travel and economic intercourse restricted, people were thrown home. They were abruptly stuck ‘in place’ where local

AUTHORS’ NOTE: We are grateful for the helpful comments of two anonymous reviewers.
social bonds have atrophied under decades of neglect. They found themselves asking: Do we even know our neighbors? Who can we rely upon? How can we help?

This paper argues for increased emphasis on rebuilding fair-weather local networks of non-market relationships as one important pathway to improving disaster resilience. We argue that the degradation of local non-market relationships has rendered individuals unnecessarily vulnerable in disasters, including in the global Covid-19 pandemic. In 2020, needs precipitated or made salient by the pandemic spurred a heartening, if harried, attempt to improvise the required local networks (Pascoe and Stripling 2020). Volunteer brigades delivered food to isolated elders (Tiratelli and Kaye 2020). Regional people-power was rallied to construct vegetable beds in urban backyards for physical and psycho-spiritual sustenance (Soul Fire Farm 2021). Networks of mutual aid sprung up all over the world (Sitirin and Sembrar 2020). However, as we will discuss, emergent disaster communities like these are often temporary, so there is an opportunity to further increase the efficacy of disaster responses by building and strengthening fair-weather local networks of non-market relationships (see also Pitas and Ehmer 2020).

Our paper characterizes these relationships as relationships between *neighbors*. As we employ the concept here, to be someone’s neighbor means to be related to that person by certain kinds of social ties characterized by care, good will, and generosity. Importantly, neighbors also share geographical proximity—they are neighbors in part because they live in the same neighborhood. Note, however, that we endorse an understanding of ‘neighbor’ that is not restricted to long-term residents of a neighborhood, but is rather inclusive enough to embrace new-comers and persons experiencing homelessness.

Relatively little has been written about the ethics of the neighbor relationship in this rather everyday sense. Nearby themes are addressed in the robust philosophical literatures on care ethics (for example, Noddings 1984; Held 2006) and the philosophy of friendship (for example, Badhwar 1993; Lynch 2005). The subject of neighbors has enjoyed extensive treatment in philosophical and theological scholarship related to the biblical injunction to love thy neighbor as thyself, as expressed in the parable of the good Samaritan. This scholarship has probed the interconnected themes of love, self, ‘Other’, ethics, and God (compare Kierkegaard [1847] 1962; Žižek, Santner, and Reinhard 2006). Seeking to address the question *who is my neighbor?*, philosophers have asked whether my neighbor is
my ‘mirrored image’, someone I could never fully and completely understand, a ‘monster’, and so on. Such questions need not be settled in order to motivate the kind of mundane ethics of neighborliness that we argue would improve disaster resilience. We offer the following rough-and-ready characterization:

A neighbor
1. takes respectful interest in their neighbors,
2. is generous to their neighbors without being patronizing,
3. is friendly, sociable, and considerate of their neighbors,
4. is engaged, according to interest and ability, in the ordinary physical upkeep of the neighborhood and well-being of its residents,
5. takes some opportunities to contribute to improving the neighborhood and the well-being of its residents,
6. offers extra help, according to ability, to neighbors and the neighborhood in times of extraordinary need.

The ethics of neighbor relationships can be connected to the philosophical debate about preferential moral attitudes, such as a parent’s specific concern for the well-being of their child, or of the special regard that citizens of the same nation may have for one another. Philosophers have asked whether relationships of physical proximity ought to have special moral significance. Waldron, for instance, asks whether moral concern properly diminishes according to distance, and if so, whether “distance” amounts to “sheer geography” (2003, 333). Insofar as neighbor relationships involve shared geography, one might worry that our argument implies moral disregard for non-neighbors. This is not our intent. Our account is compatible with individuals having substantive moral obligations to distant agents. Rather, we aim to highlight ways that cultivating relationships between neighbors can contribute to disaster resilience. While geographical proximity is not sufficient to characterize the relationship of being neighbors in our sense, ‘sheer geography’ does make a difference to the opportunities for providing and receiving aid, especially in times of disaster.

In the following section, we discuss failures of disaster response and in section III we argue that local networks of non-market relationships (that is, networks of neighbors) have some advantages over market-based relationships or state-governed aid for delivering effective responses. Section IV describes the unfortunate decline of non-market relationships and section V highlights the temporary nature of emergent responses. Section
VI provides a few specific suggestions for increasing neighborliness, which we argue would improve effective responses to disasters, including pandemics.

II. Failures in Responses

In times of disaster—including pandemics—pre-existing social problems are exacerbated. Individuals and communities already excluded from the formal economic and power structure find themselves even more stranded (Thomas et al. 2013). In addition, new problems arise. Fundamental assumptions that underlie functioning markets and the effectiveness of centralized government response may no longer hold, resulting in problems that require rapid adaptation. In particular, as we will discuss, the efficacy of disaster response increases when there is good flow of information between those with needs and those with the capacity to meet those needs, when flexibility is possible in the nature of the response, and when the agents involved in response efforts stand in relationships of mutual care. These features of disaster response are more challenging for markets and centralized governmental responses to achieve than local social networks.

Uncertainty is a fundamental attribute of crisis. During times of extraordinary need, information about who is in need, what their needs are, who is in a position to contribute to meeting the needs of others, and so forth, is critical to mobilizing effective responses. In the disruption to daily life brought on by a pandemic or other disaster, normal communication channels and infrastructure may be disturbed, leaving centralized authorities without the necessary information. Of course, this information is accessible to those who are themselves experiencing need or who have aid capacities. The ability to obtain this information is weakened the farther an individual or organization is from the experience of need and aid capacities along geographical and social lines. The 1995 Chicago heat wave provides a powerful example: deaths were concentrated among elderly individuals who were socially isolated (Klinenberg 2003). Local governments and other organizations that might have helped did not know where the need was or that it even existed.

Markets also suffer from problems caused by incomplete information. Textbook models of supply and demand assume that both buyers and sellers have full information about goods and services, and have the ability to write enforceable contracts specifying every detail of the transac-
tion. While this assumption is rarely (if ever) met during ‘normal’ situations, disasters are particularly characterized by uncertainty. For example, a contract to deliver meals to hurricane survivors cannot account for all possible logistical difficulties. Contracts drafted in information-poor contexts (‘incomplete contracts’) may go unfulfilled as new difficulties come to light that make the contract unprofitable or even impossible to fulfill. After Hurricane Maria, numerous failed Federal Emergency Management Agency (FEMA) contracts came to light, including a contract for meals that had delivered only 50,000 out of the 18.5 million meals contracted for when terminated (Mazzei and Armendariz 2018). Flexible contracts are also by nature incomplete. Incomplete contracts function only when the parties are responsive to non-contractual features like reputation and social norms, including reciprocity. The greater the missing information or necessary flexibility, the more incomplete the contract—and the more that prosocial norms and behaviors will be necessary for market exchange to function (Bowles 1998).

As disaster conditions are likely to change as the situation unfolds and more information is revealed, flexibility and speed are key aspects of effective responses. Community-organized aid can be more agile and adaptive than the lumbering and homogenous bureaucratic machinations of the state. Due to the uncertain nature of a disaster, often it will not be clear what aid is needed when, and how that need changes over time. Aid provided by governments or large non-profits may be limited in scope, as when one organization provides housing, another food, et cetera. Informal networks of mutual aid have an advantage in this respect. With no rigid area of focus, such networks can adapt their efforts to needs as they arise: grocery delivery, rides to medical appointments, help changing light bulbs, et cetera. When aid has to be provided via a legislative or bureaucratic process, it will often be slow and reactive. For instance, although in response to the Covid-19 pandemic, the United States Congress passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act expeditiously, further aid waited nine months. Hurricane Katrina is perhaps the most prominent example of a delayed government response, with various levels of government deflecting responsibility, causing a failed response that culminated in the resignation of FEMA director Michael Brown. Lack of a bureaucratic hierarchy was one of the reasons that Occupy Sandy was able to provide aid faster than government authorities following Hurricane Sandy (Feuer 2012).
Markets and governments are also significantly limited in their ability to furnish caring responses to disasters, including pandemics. In a pandemic, many required actions are taken not necessarily to protect oneself, but to protect others. Wearing a mask provides more protection from the mask wearer than for the mask wearer. For young, healthy individuals, the lesser risks of Covid may not justify behavioral changes purely on the basis of self-interest. Yet when the potential for infecting others (a classic economic externality) is taken into account, the behavioral choices become more clear. Caring creates benefits for society as well as individuals in the caring relationship, which means that markets will underprovide care (England, Budig, and Folbre 2002). While government programs can provide substantial material aid, they are also often motivated by paternalism and a punitive mindset, imposing constraints that may actively harm recipients (Davis 2019). The government aid relationship is one-directional; individuals cannot reciprocate or show gratitude to the direct source of aid.

III. SOCIAL CAPITAL AND NON-MARKET RELATIONSHIPS IN DISASTER RESPONSES

Social scientists have long investigated the role and significance of social capital, and thus, much of the available evidence that is most relevant to our argument pertains to social capital. Social capital can be thought of as one product of interpersonal relationships, including non-market relationships. While market relationships are characterized by self-interest, non-market relationships cover a wide range of human activities, from raising children to communal worship. Non-market relationships have in common that they do not maximize exchange value; the agents are not necessarily acting out of raw self-interest. ‘Social capital’ serves as an umbrella term, encompassing several different aspects of relationships characterized by reciprocity and the propensity of people who know each other to help one another. An individual’s social capital is a function of their social network (in some cases, the social capital is described as belonging to the social group itself). Individuals may have low social capital if they lack connections to others, or if their network includes others with little ability or desire to help. In both cases, the manifestation of social capital is in the actual assistance or resources given, be it help moving, access to job opportunities, or mutual aid.

Meyer (2018) provides a review of the empirical literature supporting social capital as a key factor in disaster resilience. Following Hurricane
Katrina, a number of improvised actions by local groups organized responses when local institutions were disrupted. For instance, through grassroots organizing drawing on existing groups (such as Food Not Bombs, street medics, and Indymedia), the Common Ground Collective helped mobilize essential supplies, health care, and information channels in New Orleans without the blessing or support of the state (Crow 2014). After Hurricane Sandy, the Occupy movement—a preexisting movement already practicing mutual aid—provided rapid assistance where both the government and large charity groups failed. FEMA itself has recognized the unique capabilities of local communities in the implementation of the Whole Communities program. In fact, disaster experience can also create social capital as new communities unite around a social identity born of the shared experience (Ntontis et al. 2020), a point we elaborate upon below.

For those typically excluded from market and state solutions, social ties may be the main form of assistance (Braun and Assheuer 2011). ChapPELL et al. (2007, 352–353) found that among Hurricane Katrina survivors in Mississippi, 32% claimed friends and relatives; 23% religious organizations; and 9% strangers, acquaintances, or “other” as their most important source of emergency aid. In contrast, federal aid was the most important source of emergency aid for 14%; military aid for 9%; and state or local government aid the most important source for 2%. When asked to report all sources of aid, more individuals reported receiving aid from strangers, acquaintances, or ‘other’ sources than from the federal government (40% vs. 37%).

Non-market relationships help not only in responding to crises, but in preventing them. Strong social ties increase the likelihood of mitigation efforts (see Meyer 2018, 269), as predicted by research showing that solidarity increases provision of public goods. Individuals who reported having more neighbors they exchanged greetings with and that they could rely on for help reported greater intention to wear masks, receive a vaccination, and wash hands in a hypothetical future outbreak of influenza (Chuang et al. 2015), and greater trust in others was associated with greater intent to be vaccinated in the H1N1 pandemic (Rönnerstrand 2013, 2014). Early research shows an association between higher social capital and fewer Covid-19 cases (Makridis and Wu 2021; Fraser, Aldrich, and Page-Tan 2020). Similarly, social capital has been associated with psychological resilience when disasters strike. Noel, Cork, and White (2018)
review the literature and find that higher levels of social capital are associated with better mental health outcomes, particularly lower post-traumatic stress.

We have appealed to social scientific evidence regarding social capital out of necessity. However, our emphasis will ultimately rest on the value of neighbor relationships rather than social capital per se (see section VI). The concept of social capital does not necessarily connote geographically local relationships. Social capital can also be formed from market or government interactions. In contrast, neighbors are individuals living in geographic proximity to one another who form certain kinds of non-market ties and participate in a certain kind of social relationship characterized by care, good will, and generosity. Having neighbors is therefore more specific than being rich in social capital, since social capital could come in many different forms, may not be place-specific, and could involve market relationships. This difference suggests that future social science research on disaster resilience may benefit from disaggregating the influence of neighbor networks from social capital broadly construed.

IV. NETWORKS OF NON-MARKET RELATIONSHIPS ARE ON THE DECLINE

We have argued that local networks of non-market relationships expedite effective responses to disasters. Unfortunately, as the market economy has grown, more aspects of production have entered the market sphere and weakened networks of non-market relationships (Tittenbrun 2017). As more relationships between individuals have come to be characterized by market exchange, the scope of relationships that build and sustain interpersonal trust and care has been reduced. Household ‘reproductive labor’, including child care and food preparation, is increasingly provided by the market instead of family. Even areas formerly handled by the state, such as the provision of public goods, have increasingly been shifted to market provision via ‘public-private partnerships’ (examples include privately managed toll highways and water systems). Market exchange is characterized by an impersonal nature; indeed, as markets expand and become more globalized the opportunities for repeat interaction that could cultivate a caring relationship are reduced. As a result of shifting the practice of care to the market, individuals have fewer opportunities to practice care and develop valuable social capital (Ciscel and Heath 2001). Individuals routinely caring for community members also appears to be in decline—even between 2003 and 2019, the share of individuals reporting care duties for a non-household member on a given day
dropped from 15.7% to 10.5%, and average hours per week caring for a non-household member dropped by 32% (U. S. Bureau of Labor Statistics, n. d.). The overall decline in social capital has been famously documented by Putnam (2000).

In addition to reducing the scope of non-market relationships, market expansions may incur broader effects on norms and values. The benefit of acquiring a reputation for trustworthiness or fairness declines as these traits are less used (Bowles 1998). Market expansions and the shift away from local businesses to national or multinational firms accelerates the decline in social capital (Heying 1997; Clark and Record 2017; Goetz and Rupasingha 2006). Indeed, the process of market expansion includes finding replacements for the very social capital it weakens, which can result in apparent GDP growth alongside decreases in social capital (Bartolini and Bonatti 2008). Tsakalotos notes that market expansion “makes alternative conceptions much more difficult to conceptualize, let alone carry out” (2004, 29). Once the market and state are the dominant ‘solutions’ in a society, they will more and more appear the only solutions. The existence of non-hierarchical, non-market mutual aid between neighbors in the face of these countervailing forces is a testament to just how beneficial such relationships are.

V. EMERGENT RESPONSES

‘Emergent social capital’ has been described as a temporary phenomenon arising during disasters. Solnit (2010) poignantly describes the temporary solidarity often found in disaster situations. This phenomenon has been called “catastrophe compassion” (Zaki 2020, 588).

In direct response to a disaster, emergent local aid groups can form, such as those that rescued survivors and fought fires after the 1995 Kobe Earthquake (Aldrich 2011). Over the medium-term, while it may be true that funding and certain kinds of information can be readily shared across great distances insofar as the will exists to do so, there are physical aspects of disaster response that, by their very nature, must be accomplished by people on site. Child care, nursing the ill, fetching food and medicine, and the irreplaceable value of in-person presence and social interactions, are just a few examples. Especially in circumstances where travel is restricted—as is often the case following natural disasters, and has been enforced as a matter of policy in the Covid-19 pandemic—local people will be the ones who will ultimately perform this work. If they do
it voluntarily and bypass the market and state by organizing themselves, they can often do it particularly effectively.

However, even when non-market relationships are built or strengthened in direct response to disaster conditions, experience has shown that as the acute conditions ebb, these spontaneous relationships of mutual aid will be overpowered by impersonal and voracious market forces (Ntontis et al. 2019, 2020). Fledgling mutual aid initiatives are unlikely to survive the full-scale return of market-based interactions precisely because those initiatives were reactionary. When the next disaster arrives, individuals can hope for another temporary resurgence in mutual aid. But without concerted effort to entrench and normalize local networks of non-market relationships, mutual aid will remain fringe and occasional. This is to our detriment, since emergent mutual aid initiatives are better equipped to succeed in their aims if they can draw on extant networks of trust and reservoirs of information gathered by established social infrastructure (Jun and Lance 2020).

VI. CULTIVATING NEIGHBOR RELATIONSHIPS

Taken together, our arguments thus far show that individuals are unnecessarily vulnerable to disasters, including disease outbreaks such as the catastrophic global Covid-19 pandemic. Although there is evidence that local networks of non-market relationships aid in effective disaster response, the stability and strength of such networks has generally been declining. When mutual aid networks emerge in response to disasters, they are effective, but unfortunately short-lived. In contrast, in circumstances where networks were already in place before the onset of the disaster, those contributing to disaster response could draw on resources, familiarity, trust, and pre-existing infrastructure. As we have seen, the benefits that local networks of non-market relationships have in disaster response hinge at least in part on their access to the right sort of information about needs and capacities, flexibility unhindered by rigid bureaucracy or large-scale coordination, and the sort of care that accompanies social interactions. This suggests a path forward: building and strengthening fair-weather local networks of non-market relationships; that is, cultivating neighbor relationships.

Individuals can endeavor to cultivate neighbor relationships by introducing themselves to others in their neighborhoods, inviting them to social events, and organizing block parties or community dinners, to mention just a few examples. As Helm (2008) has argued, participating in a
shared activity intentionally directed at an aim that the agents involved care about is closely related to mutual affection. Thus, cultivating social relationships among neighbors could also involve working together on projects of mutual interest, such as tending a community garden, neighborhood repair and beautification projects, and grassroots activism to address issues of mutual concern. It could also involve doing one another favors (minding children), extending friendly gestures (offering help with projects), and giving gifts. Obviously, these activities are more easily and safely accomplished in fair-weather/non-pandemic times, which speaks to the importance of proactively cultivating neighbor relationships.

While cultivating neighborliness in these sorts of ways may seem commonsensical, building social relationships among neighbors during ‘normal’ times may require intentional action. Pew Research has found that across community types, a greater percentage of older adults than younger ones reported feeling supported in their communities, and older adults are more likely than younger ones to know their neighbors (Parker et al. 2018, 65, 77). Younger adults may eventually build support among their neighbors as they age themselves. However, prudence cautions against taking this possibility for granted. For some communities, aversion to bringing issues perceived as “political” into the local social dynamics may be an obstacle to building effective and lasting networks of mutual care (Grayson 2020, 28). While non-market relationships need not constitute overtly anti-market activism in order to improve disaster resilience, building mundane trust among neighbors may lay groundwork for further political reflection and action, which may in turn break down obstacles to neighbor relationships. Similarly, now-entrenched social norms characterizing neighbors as “eyes on the street” rather than, say, folks who organize together, may have to shift first via intentional steps, like creating conditions conducive to chance encounters in the neighborhood (Halegoua and Johnson, forthcoming, 13).

We have argued that building networks of neighbor relationships will bolster the capacities that neighborhoods have to respond to disasters effectively. Disposing oneself to one’s neighbors and neighborhood in the ways articulated above would likely be associated with increased access to information of particular relevance to disaster response and with mutual care among neighbors. In virtue of the informal nature of the social networks thereby established, neighbors would retain the capacity to respond nimbly. Finally, we want to stress that cultivating neighbor relationships is valuable for its own sake. Being in relationship with one’s
neighbors (much like belonging to friendships) is enriching, regardless of the occurrence of disasters. Although cultivating such relationships may have become unfamiliar, we believe it would be well worth it—and, as Berry reminds us, there are evening stories waiting to be told.

REFERENCES


**Nora Mills Boyd** is an Assistant Professor of Philosophy at Siena College. She holds a PhD in History and Philosophy of Science from the University of Pittsburgh.

Contact e-mail: <nboyd@siena.edu>

**Matthew Davis** is an Assistant Professor of Economics at Siena College. He received his PhD in Economics from American University.

Contact e-mail: <medavis@siena.edu>
Mandated Shutdowns, the Ratchet Effect, and The Barstool Fund

JEFFREY CARROLL
University of Virginia

Abstract: Perhaps the most contentious part of the response to the Covid-19 pandemic has been the decision by governments to mandate—or effectively mandate—the shutdown of certain businesses. The justification for doing so is broadly consequentialist. The public health costs of not shutting down are so great that potential benefits from allowing businesses to open are dwarfed. Operating within this consequentialist framework, this paper identifies an underappreciated set of social costs that are a product of the present public policy that pairs mandated shutdowns with government subsidies. Such policy is prone to being an instance of what Robert Higgs calls the ratchet effect. Given that ratchets tend to be both costly and sticky, it is best to avoid allowing them to come into existence. This paper identifies a way of circumventing this particular ratchet; namely, by replacing governmental subsidies with support from private charitable funds like The Barstool Fund.

Keywords: consequentialism, pandemic, Robert Higgs, ratchet effect, shutdowns

JEL Classification: D7, H11, H2, H51, H84, I1

I. INTRODUCTION

One of the most controversial aspects of the response to the Covid-19 pandemic has been the decision by governments to mandate—or to enact stay at home orders that effectively mandate—the shutdown of certain businesses deemed high-risk for transmission of the virus. Such businesses include restaurants, bars, gyms, movie theaters, casinos, museums, and libraries, amongst others. The rationale given for doing so is consequentialist: the costs resulting from further spreading the disease
outweigh the benefits from allowing business owners to decide to operate as they best see fit.¹

But, public policy of this sort has detrimental effects on the businesses, business owners, and employees. These effects are particularly pronounced on small businesses.² Requiring that a small business close—or effectively close by, say, prohibiting indoor dining—leads to employees being laid off or fired. In the worst cases, the loss of revenue necessitates that some businesses close for good.

Many proponents of shutdowns find this consequence unacceptable. In the United States, the proponents’ solution is to provide governmental aid to the businesses in the form of PPP loans and to the out-of-work employees in the form of stimulus checks. However, if the shutdown persists sufficiently long, then the aid to both the businesses and the individuals will run out and another round of aid will need to be administered. For said proponents, such an ongoing cycle can be (indeed, needs to be) justified until the pandemic ends.

Yet, if we are consequentializing (as the given rationale suggests we are), this is a very costly route to go. One cannot consider only the near-term costs of the policy. Both the near-term and long-term costs of taking decision-making authority out of the hands of private citizens need to be factored into the calculus.³

This paper focuses on the cumulative costs that result from the government mandating (including effectively mandating) a shutdown.⁴ In particular, it identifies the public policy response of shutting down businesses and subsequently using governmental aid to keep the affected businesses and individuals afloat as an instance of what Robert Higgs (1987) calls the ratchet effect. Section II describes the ratchet effect. Section III, drawing on the work of Nathan Goodman, Christopher Coyne, and

---

¹ I do not commit to consequentialist or utilitarian public policy. My concern is with what policy conclusions follow if we embrace the broadly consequentialist and utilitarian framework that is used in popular discourse to justify mandatory shutdowns. For a thorough defense of utilitarianism as the proper policy lens, see Goodin (1995).

² This is not to say that it does not hurt larger businesses and corporations. But large chains like Chipotle and Dominoes have done quite well during the pandemic because they have the capital and infrastructure to more easily and effectively transition to being fully to-go and delivery establishments than, say, your favorite local Mexican restaurant or hometown pizza place.

³ This essay focuses on the costs created by the pandemic public policy. For an excellent essay on why the costs would actually be lower in the absence of governmental intervention, see Leeson and Rouanet (2021).

⁴ For more on hidden costs, see Bastiat ([1964] 1995). Relatedly, for a discussion of the complex implications of pandemic public policy on gender equality, see Alon et al. (2020).
Abigail Devereaux (2020) as well as the related work of Coyne and Yatsyshina (2020), makes the case that the pandemic policy is prone to becoming a ratchet. Section IV considers the normative implications of a ratchet within a broadly consequentialist framework and states why it is worth taking steps to avoid this outcome. At the conclusion of Section IV, one may believe that a choice must be made between two undesirable alternatives: either opening society and accepting the costs associated with the amplified spread of the virus, or closing society and accepting the costs involved with the ratchet. Fortunately, this is a false dichotomy. Section V looks at a way of avoiding the ratchet; specifically, by relying on private, non-governmental agents to voluntarily support the shutdown businesses and out-of-work employees. The example of The Barstool Fund is considered and the reasons for opting for a private initiative like it over governmental aid, whenever possible, is discussed.

II. THE RATCHET EFFECT

In Crisis and Leviathan, Robert Higgs (1987) makes the case that much of the growth of government is explicable in terms of a pattern of crisis and response. Following a crisis, government is larger than it would have been had the crisis not occurred (though not as large as during the crisis). Crises are a sort of ‘ratchet’ that propel the expansion of government forward indefinitely.

But in what respect is it larger? That is, how is the size of government measured? The way to measure the size of government, says Higgs, is in terms of how much authority it has, not in terms of activity which “is secondary and derivative” (1987, 33). After a national emergency, the relevant sense in which government grows is in terms of the amount of authority it has over something like economic decision-making.

A logical next question is: Why does governmental authority expand? Part of this expansion is a result “of the ‘hard residues’ of crisis-spawned institutions (for example, administrative agencies and legal precedents)” (Higgs 1987, 58–59). Once an institution exists, it is hard to get rid of it in its entirety. Another part is a result of the shift in ideology following a crisis as “the events of the crisis created new understandings of and new attitudes toward governmental action” (Higgs 1987, 59). In short, people, in some sense, want the emergent institutions.

Still, an account of how the ratchet effect occurs is owed. Higgs explains it in the form of a five-stage sequence: (1) pre-crisis normality; (2) expansion; (3) maturity; (4) retrenchment; and (5) post-crisis normality.
The narrative underlying this progression of stages goes as follows. There is a base rate at which government is expanding. Then, a crisis ensues resulting in government growing at a much higher rate because “an insistent but ill-defined public [demands] that the government ‘do something’ about the crisis” (Higgs 1987, 73). But doing something is costly and no one (or no institution) wants to be responsible for said costs. So, government “takes steps to conceal the true costs” by substituting “a (cost-hiding) command-and-control system of resource allocation for the (cost-revealing) market system” (Higgs 1987, 73).

After this expansion levels off (that is, reaches maturity), then it retrenches, but to a level that is higher than it would have been had the crisis not occurred because some of the emergent apparatus persists. Some of the aforementioned ‘residue’ remains. This residue also makes it easier for additional expansions in the future.

The aim of the next section is to provide reason to believe that the public policy response to the Covid-19 pandemic, particularly as it has been rolled out in places like the United States, is prone to following the pattern of stages identified by Higgs in his articulation of the ratchet effect.

**III. PANDEMIC PUBLIC POLICY AS AN INSTANCE OF THE RATCHET EFFECT**

There are numerous examples of governmental growth following infectious disease crises. One example is segregation in Cape Town, South Africa that was, in part, a response to the bubonic plague, while another is the use of precedent from a legal case concerning compulsory vaccination for smallpox to justify a ruling for non-volitional sterilization. Is the public policy response to the Covid-19 pandemic—specifically, the part involving the mandated shutdowns of businesses and the subsequent governmental aid—susceptible to being another?

For the answer to be yes, it needs to be the case that (i) expansion occurs in response to the mandated shutdowns and (ii) retrenchment leaves residue that makes the size of government larger than it would have been had the pandemic not occurred. The expansion of government in other areas has been documented (e.g. Habeeshian and McDade 2020; Kallingal 2020; Neil 2020; Speri 2020; Tuccille 2020). The germane

---

5 A more thorough discussion of these examples can be found in Goodman, Coyne, and Devereaux (2020).

6 Popular discussions of the policy response to the Covid-19 pandemic as an instance of the ratchet effect are available, such as Rozmajzl (2020) and Goodwin (2020).
challenge is to outline the propulsion mechanism for this domain-specific ratchet.

Fortunately, the logic of this particular ratchet is not too mysterious or difficult to understand. Individuals are prohibited from working. Meanwhile, the need to make a living does not go away. Likewise, small businesses are prohibited from opening. Meanwhile, the need to make a profit does not go away. So, both individuals and small businesses need to be exogenously supported. This support comes from government. Stimulus checks to individuals and PPP loans to businesses are used to keep both afloat. Yet, as the pandemic persists and the decision to mandate the closure of certain businesses continues, the initial support gets exhausted and additional support becomes necessary. Government must then distribute a new round of support. This cycle repeats, presumptively, until the pandemic ends—and it is unknown when that will be.

The underlying logic makes for an indefinite cycle of expansion. Once the pandemic is reined in, the shutdown orders are able to be lifted. This means that individuals and businesses can resume normal operation. But it does not mean that all the governmental apparatus instantaneously disappears. Residue remains. This residue may be in the form of new governmental agencies persisting or previously existing agencies becoming bloated. Additional staff may need to be hired to adequately rollout the aid distribution. The addition or expansion of these agencies as well as the personnel staffing them are a social cost, albeit a dispersed one. With each successive round of governmental aid, the social costs increase as does the overall cost-trajectory as a result of additional excessive residue remaining.

Moreover, there are potentially even more social costs in the form of interventions in the future that are, at least partially, parasitic on the present policy response to the Covid-19 pandemic. The ideological shift and the presence of the residue, together, increase the potential for future interventions that may themselves be socially costly.

IV. CONSEQUENTIALISM, COSTS, AND THE PANDEMIC POLICY RATCHET

Having made the case that this policy response is an instance of Higgs’s ratchet, it is now time to state the case why this is undesirable in normative terms. One mode of critique would be to challenge the policy response from the vantage point of an alternative moral theory (say, a deontological one that finds the shutdown unjustifiable because of its restriction on individual rights). I believe a more effective critique is one
that operates within the same broadly consequentialist framework that proponents of the shutdown defend. The specific justification seems to be that the net costs of mandating a shutdown is lower than the net costs of allowing these industries to proceed unfettered. The underlying supposition is that certain types of businesses are sites of significant transmission of the virus and the consequences of transmission are decisively costly.

I in no way doubt that there would be significant public health costs involved in allowing these businesses to operate without restriction. ⁷ But I also believe that the costs resulting from the implementation of a mandatory shutdown are under-calculated. To be clear, I do not claim to show that the costs of not shutting down are less than the costs of shutting down. Rather, I merely seek to show that costs of not shutting down have been under-calculated in the comparative analysis. This is because the long-term costs of the ratchet effect are not being factored in. The near-term costs, such as the loss of work for individuals and the significant debts incurred by businesses, have been well-rehearsed. But there are longer-term social costs such as those resulting from each successive round of stimulus and from the institutional residue left as a result of this particular public policy. And these longer-term costs have not been adequately considered. ⁸

What are these longer-term costs? In their paper “Pandemic Police States”, Coyne and Yatsyshina (2020) catalog costly expansions that may linger. Using Coyne and Yatsyshina’s analysis as a point of departure, I propose the following three general categories under which residual costs may fall for this domain: publicizing, enforcing, and punishing. First, government must make it known that a shutdown will be occurring. This requires publicizing the rules which will be the responsibility of someone(s) or some institution(s). Second, government must enforce the shutdown. This involves both monitoring and administering. Governmental actors are needed to scout out which parties are failing to comply with the shutdown and to ticket those in violation. Third, government must hold rule-breakers accountable. There needs to be a set of governmental

---

⁷ There also important health costs that are more personal from shutting down. See Mulligan (2020).
⁸ There are also prospective costs from future interventions that are made possible by the present policy response. While I will not discuss this in detail, my argument is only strengthened to the extent that such interventions are costly as a result. Thank you to an anonymous reviewer for prompting this point.
representatives that collect the debts incurred by the violators or seek further punishment for those that refuse to pay.\(^9\)

It is far from obvious that all the additional actors needed by government to perform these tasks will be relieved of their roles after they are completed. More likely is that many will remain on payroll or be reassigned to a different division, thereby, expanding the size and cost of government (as well as potentially facilitating future interventions). These long-term costs can continue to accrue (perhaps, indefinitely?) in a way that limited-term public health costs do not.

V. Rectifying the Ratchet: The Barstool Fund as Case Study
At this point, one may plausibly think that consequentialist public policy is between a rock and a hard place. Either mandate a shutdown of businesses for the sake of containing Covid-19 while supporting those impacted by providing governmental aid—accepting the costs of the impending ratchet, or allowing businesses to open so that the impending ratchet can be avoided—accepting the public health costs of further spread of the virus. Neither option seems very appealing.

Fortunately, this is a false dichotomy. One needn’t choose between these two options. There already exists a third way as evidenced by The Barstool Fund. At the time of this writing, the comedy-sports media company Barstool Sports has raised nearly 40 million dollars from over 200,000 donors in order to be able to support over 300 small businesses for the remainder of the pandemic. Its founder, Dave Portnoy, in concert with Liz Gonzalez, an employee of the company, screened through applications and allocated funds to small businesses impacted by the policy response to the pandemic. If a business receives funding, it is sufficient to get them through until the relevant governmental restrictions are lifted.

Thus, the third way is that if small businesses are not permitted to remain open, then private funds can support them. Having the support come from private funds like The Barstool Fund is preferable to support that comes from government because it prevents ‘residue’ being left behind in the form of bloated governmental agencies. This is because the costs are internalized by the private company. In this particular case, the operation of the fund is being done by people who are already employees of the company. So, there is no additional overhead or bloat. Furthermore,

\(^9\) Like was the case with the Atilis Gym, ran by Ian Smith, in New Jersey (Bellano 2021).
in virtue of the support being provided by donations to a private fund, the support comes from wholly voluntary sources and sidesteps thorny issues that arise for governmental aid that comes from taxation or increasing the national debt.

Of course, there are limits to the extent to which private charitable funds can support small businesses. There is both a conceptual and an empirical point to address. The conceptual point is that it should not be a strike against a private alternative that it cannot support all small businesses because, surely and hopefully, there are limits to what government can support, as well. The relevant comparison is not between a limited charitable fund and an unlimited government. Such a comparison makes it seem as if the support provided by government comes deus ex machina, when it actually is a dispersed social cost. It is just one that is (perhaps, purposefully) hard to precisely define.\(^{10}\) Such a comparison is equivalent to comparing nonideal capitalism with ideal socialism.\(^{11}\) One must compare like-to-like. This means that while charitable funds are limited, the full costs of unbounded governmental support must be considered. Doing so makes the latter decidedly more difficult to justify in consequentialist terms. Hence, there is a compelling pro tanto case to be made that it is less costly for private funds to support businesses and individuals than it is for government to provide blanket aid.

A plausible conclusion would be that private funds, not government, should support businesses and individuals impacted by a mandated shutdown. But this may bump up against empirical realities. The number and power of private funds will vary by country.\(^{12}\) Those with less of a charitable tradition may not have anywhere near the private support needed to support the businesses that are forced to close. What follows from this? It is not that that empirical realities necessitate that aid come from government, since private funds are too limited. Rather, the inference that should be drawn is that private funds, whenever possible, should support those impacted by the shutdown. While private funds may not be able to support all in need, it is preferable to have private support when possible.

A relevant corollary is that the proliferation of private funds would be a superior alternative to resorting to government aid. Put differently,

\(^{10}\) Higgs (1987) offers an explanatory logic as to why the cost-concealing apparatus of government is opted for over the cost-revealing mechanisms of the market.

\(^{11}\) Brennan (2014) and Freiman (2017) charge Cohen (2009) with making such an unfair comparison. The general point that behavioral asymmetry needs to be assumed in comparative analysis is found in Buchanan and Brennan (2000).

\(^{12}\) Thank you to the editors for raising this point.
public policy should be looking at ways to incentivize increasing the number and expanding the size of private funds. The solution in a slogan is ‘more private funds, not more government’. An interesting implication worth further exploration is that in cases where private funds are insufficient to support those in need, it would be preferable to mandate charity rather than levy a tax as it eliminates the tendency for lingering, unnecessary residue, and reduces the propensity for future interventions.13

The upshot of this section is that one is not forced to choose between increased public health costs that result from a full-fledged opening and increased economic costs that result from restrictions on businesses.14 One who thinks that the public health costs of opening are sufficient to justify mandating a shutdown should not couple that conviction with governmental aid. Doing so creates the condition for a thriving ratchet. Given the difficulty of rectifying a ratchet once it emerges, it is best to avoid bringing it into existence in the first place. Hence, private aid is preferable and public policy should seek ways to incentivize it and expand its extent so that governmental aid does not need to be resorted to or is resorted to as little as possible.

VI. CONCLUSION
This paper has made the case that the policy response to the pandemic is an instance of the ratchet effect. Within a consequentialist policy frame, ratchets are problematic because they lead to increased costs over time. But there are emerging alternatives that offer the potential to break the ratchet or prevent one from emerging. Voluntary charitable funds like The Barstool Fund have the potential to support individuals and small businesses during the shutdown without the potential for the problematic institutional bloat that arises with governmental support. In short, private funds can offer the same type of benefits, but with a relative reduction in costs. Though, of course, private funds will be unable to support all in need when government mandates a shutdown. Nonetheless, complete support from private funds can be the aspirational goal.15 Hence, one way

13 I lack the space to offer a full-fledged defense of this implication here, but believe public policy would benefit from a deeper comparative analysis between mandated charity, on the one hand, and tax and redistribution, on the other.
14 There are other middle options, some of which may be worth exploring. My aim in this paper is not to settle on a precise positive proposal. Rather, it is to make the case that costs resulting from shutdowns are far more extensive and tend to accrue much farther into the future than is often recognized. Thanks to Greg Robson for a helpful discussion on this point.
15 On aspirational goals, see Estlund (2014, 2020).
of understanding the argument of this paper is that if we are committed to a mandatory shutdown (which I have reservations about), then if we are committed to supporting those out of work as a result of the shutdown (which, when the prior antecedent obtains, I have reservations about), then the support should come by way of private charitable funds like The Barstool Fund because it reduces the net social costs by reducing the institutional bloat of governmental agencies over time.

**REFERENCES**


**Jeffrey Carroll** is a PhD candidate in Philosophy at the University of Virginia. His research is in social and political philosophy and PPE. More specifically, he often works on methodological issues concerning justice-theorizing and democracy. His published work appears in *The Journal of Politics, Social Theory and Practice, HEC Forum*, and *The Independent Review*. Contact e-mail: <jmcarroll422@gmail.com>
Pandemic Windfalls and Obligations of Justice

BRIAN BERKEY
University of Pennsylvania

Abstract: The Covid-19 pandemic has caused significant economic hardships for millions of people around the world. Meanwhile, many of the world’s richest people have seen their wealth increase substantially during the pandemic, despite the significant economic disruptions that it has caused on the whole. It is uncontroversial that these effects, which have exacerbated already unacceptable levels of poverty and inequality, call for robust policy responses from governments. In this paper, I argue that the disparate economic effects of the pandemic also generate direct obligations of justice for those who have benefitted from pandemic windfalls. Specifically, I argue that even if we accept that those who benefit from distributive injustice in the ordinary, predictable course of life within unjust institutions do not have direct obligations to redirect their unjust benefits to those who are unjustly disadvantaged, there are powerful reasons to hold that benefitting from pandemic windfalls does ground such an obligation.

Keywords: Covid-19, distributive justice, inequality, institutions, poverty, windfalls

JEL Classification: A13, D31, D63, D64, E24, I30

I. INTRODUCTION

The Covid-19 pandemic has caused significant economic hardships for millions of people around the world. In the United States alone, it is estimated that the number of citizens living in poverty grew by 8.1 million between June and December of 2020 (Han, Meyer, and Sullivan 2021).\(^1\)

\(^1\) It is worth noting that the poverty rate in the United States did decline modestly between January and June of 2020 (Han, Meyer, and Sullivan 2021), most likely due to a combination of Earned Income Tax Credit payments, CARES Act stimulus checks, and the expanded unemployment benefits also provided under the CARES Act (Perolin et al. 2020, 4–5). Nonetheless, there was still a significant increase overall between January and December, and the large increase between June and December coincided with a 40% reduction in the unemployment rate (Han, Meyer, and Sullivan 2021). The
This increase is primarily the result of lost income from paid labor,\(^2\) combined with the lack of a government policy response that would sufficiently offset these losses to prevent people from falling into poverty. Many people have lost their jobs, while others have had their hours reduced, or found their opportunities to earn money through work in the ‘gig economy’ curtailed. The economic losses that Americans have suffered as a result of the pandemic have, to a large degree, been experienced by the poor and the working class.\(^3\)

Globally, it is estimated that the pandemic could increase the number of people living below the World Bank’s $5.50 per day poverty threshold by between 200 and 500 million (Berkhout et al. 2021, 24).\(^4\) More generally, there is evidence that the pandemic will increase inequality throughout much of the world (World Bank 2020, xi), and that the poorest people in virtually every country will experience a drop in their incomes (Christensen and Wells 2020, 8–9).

Meanwhile, as has been widely reported, many of the world’s richest people have seen their wealth increase substantially during the pandemic, despite the significant economic disruptions that it has caused on the whole. The world’s billionaires gained nearly $4 trillion in wealth between March and December of 2020 (Berkhout et al. 2021, 23). Elon Musk alone gained nearly $129 billion in those ten months, while Jeff Bezos gained over $78 billion (Berkhout et al. 2021, 23).

In addition, while other well-off people have not experienced the massive gains that billionaires have captured, those in the upper middle class or higher in the income distribution in wealthy countries have largely avoided significant losses from the pandemic.\(^5\) One important reason for this is that these people are more likely to have jobs that they can perform from home, and so are less likely to have lost their jobs or had their hours

\(^2\) More than 93 million unemployment insurance claims have been filed in the United States during the pandemic (Han, Meyer, and Sullivan 2021).

\(^3\) From February to mid-May of 2020, employment among Americans in the bottom 25% of wage earners fell by 35% (Timiraos 2020).

\(^4\) For the lower estimate, see Lakner et al. (2020); for the higher, see Sumner, Ortiz-Juarez, and Hoy (2020, 8).

\(^5\) In the UK, for example, those in the top 20% of the income distribution have saved $30 billion since March of 2020, while those with lower incomes have tended to fall (further) into debt (Berkhout et al. 2021, 24).
or pay cut. While the pandemic may not have increased the income of most of these people, it has benefitted many of them economically, since they have been able to avoid commuting costs, and have been able to save more money in virtue of the more limited availability and attractiveness of, for example, the leisure activities on which they typically spend some of their income.

The pandemic, then, has generated substantial economic windfalls for many of the world’s better-off people, while at the same time causing significant economic hardships to befall many of the worst-off citizens in nearly every country. It is relatively uncontroversial that these effects, which have exacerbated already unacceptable levels of poverty and inequality, call for robust policy responses from governments. Justice requires, for example, that states provide income supplements to those who have been thrust into poverty, and that they take steps to prevent people from becoming homeless (for example, by adopting eviction moratoriums).  

In the remainder of this paper, I argue that the disparate economic effects of the pandemic (that is, windfalls for the already well-off and increased hardships for the poor and working class) also generate direct obligations of justice for those who have benefitted from pandemic windfalls. Specifically, I argue that even if we accept that those who benefit from distributive injustice in the ordinary, predictable course of life within unjust institutions do not have direct obligations to redirect their unjust benefits to those who are unjustly disadvantaged, there are powerful reasons to hold that benefitting from pandemic windfalls does ground such an obligation.

I proceed in the remainder of the paper as follows. First, in section II, I clarify how, for the purposes of the paper, I understand what constitutes a windfall. In addition, I explain, with reference to the two most prominent views about the kinds of factors that determine the justice or injustice of a distribution, the relationship between windfalls and distributive justice. Next, in section III, I argue that on either view about the kinds of factors that determine the justice/injustice of a distribution, there are compelling grounds for accepting that beneficiaries of pandemic windfalls are obligated to redirect their windfall benefits in ways that will improve the lives of those who are unjustly disadvantaged. I conclude, in

---

6 These requirements apply on top of standing requirements to raise citizens out of poverty, ensure decent housing for the homeless, etc.
section IV, by briefly noting the central implications of my argument with respect to the current pandemic.

II. DISTRIBUTIVE JUSTICE AND WINDFALLS

In colloquial terms, windfalls are economic gains that are not the result of careful or strategic planning for the event or events that caused those gains to obtain. Typically, only at least moderately large gains are described as windfalls. If I find a $10 bill while walking down the street, it would seem an exaggeration to claim that I have received a windfall. Windfalls are often unexpected and unpredictable, though they need not be. Examples of economic gains that intuitively count as windfalls include substantial lottery winnings (which are unexpected and unpredictable) and large inheritances (which in some cases are both expected and predictable).

For the purposes of this paper, I will limit my account of what constitutes a windfall to significant economic gains that are not the result of (relevant forms of) careful or strategic planning, and are either: (1) enjoyed by those who were not unjustly disadvantaged prior to the receipt of those gains; or (2) enjoyed by those who were unjustly disadvantaged prior to the receipt of the gains, but large enough that they improve the economic position of the beneficiaries to an extent that they become unjustly advantaged. In addition, within the second category, I take the portion of a person’s gains that constitutes a windfall to be limited to that which makes it the case that she becomes unjustly advantaged on the whole. So, for example, if a person who initially possessed no wealth at all, and was therefore unjustly disadvantaged, wins $500,000 in the lottery, then the amount of her windfall is $500,000 minus whatever amount she ought to have had as a matter of justice initially (so, if she ought to have had $100,000, then she received a $400,000 windfall). This characterization of windfalls ensures that my claim that individuals are obligated to redirect windfall benefits does not imply that anyone will ever

7 Some might worry that counting gains in this second category as windfalls begs the question in favor of my view. But importantly, the fact that one is unjustly advantaged does not by itself imply that one has direct obligations to redirect their unjust benefits to the unjustly disadvantaged. In fact, as my discussion will show, many philosophers reject this view. What does follow, on virtually all views of what is required when some are unjustly advantaged while others are unjustly disadvantaged, is that the state ought to adopt policies that will remedy the relevant injustice, without infringing other requirements of justice (for example, by increasing taxes on the unjustly advantaged and using the funds generated in ways that benefit the unjustly disadvantaged). My argument relies only on this view about appropriate state responses to injustice.
be obligated to redirect resources that they ought to possess as a matter of justice.

In order to clarify how we might understand the relationship between windfalls and distributive justice, it is important to consider what kinds of factors might determine the justice or injustice of distributions. There are two prominent views about this question in political philosophy. On the first type of view, distributive justice is fundamentally about outcomes, and requires that the distribution of resources satisfies a substantive criterion or set of criteria (Cohen 2008, 126). The *sufficientarian* view that a distribution is just if and only if everyone has sufficient resources to live a flourishing life is an example of a view of this kind, as is the *luck egalitarian* view that a distribution is just if and only if there are no inequalities that reflect differences in luck rather than choice.

To illustrate what will count as a windfall on views of this kind, consider the sufficientarian view as an example. On this view, a person has benefitted from a windfall (1) if she already had sufficient resources to live a flourishing life and then experiences a significant economic gain that is not the result of careful or strategic planning on her part, or (2) if she experiences a significant economic gain that is not the result of strategic planning on her part and that moves her from being unjustly disadvantaged to being unjustly advantaged. On plausible views, she will count as unjustly advantaged after receiving a significant gain if, for example, the state would be justified in increasing her tax burden in order to provide resources to those who are unjustly disadvantaged. Importantly, this criterion for when a person is unjustly advantaged can be applied in conjunction with any outcome-focused account of distributive justice.

According to the second type of view about the factors that determine the justice or injustice of a distribution, distributive justice is fundamentally procedural rather than outcome-focused. On views of this kind, justice requires that procedures that meet certain conditions are implemented, and a distribution is just if it results from those procedures being

---

8 Views of this type need not take resources to be the metric or ‘currency’ of justice (Cohen 1989). Instead, they can hold that the distribution of resources must ensure that another currency, such as welfare or capabilities, is distributed in a way consistent with justice.

9 Some philosophers who are sufficientarians or luck egalitarians do not accept an outcome-focused account of distributive justice, and instead hold that, for example, a distribution is just if relevant institutional procedures ensure, as much as possible, that everyone has sufficient resources to live a flourishing life, or that inequalities that reflect differences in luck are limited. For my purposes, the important thing is that views of the first type apply the criteria of distributive justice (whatever those criteria are) directly to distributive outcomes.
properly followed. Robert Nozick’s *libertarian* account of justice in “holdings” (1974, chap. 7) is a view of this kind, as is John Rawls’s “pure procedural” (1999, 74–77) account of distributive justice, according to which a distributive outcome is just, “whatever it is” (1999, 75), so long as the institutions of the “basic structure of society” (1999, 6–9) have adopted policies consistent with his two principles of justice, and those policies have been properly followed.10

On views of this kind, a person has benefitted from a windfall (1) if she already had at least as much as she would have had with just procedures in place, and then experiences a significant economic gain that is not the result of careful or strategic planning on her part, or (2) if she experiences a significant economic gain that is not the result of strategic planning on her part and that moves her from having less than she would have had with just procedures in place to having more than she would have had with just procedures in place.

It is important to note that while I have defined windfalls in terms of economic gains, my argument does not require that the correct metric or currency of justice is one on which economic resources are a direct component (for example, Dworkin’s 1981 *resourcist* view, or Rawls’s 1999 account of primary social goods).11 This is because I have defined windfalls so that only gains in economic resources that leave one with more than she is entitled to as a matter of justice can constitute windfall gains. Even if, for example, welfare is the currency of justice (Arneson 2000), or a component of the currency of justice (Cohen 1989), it is resources that must be distributed in order to ensure that individuals’ entitlements of justice are satisfied—welfare is not directly distributable. Because of this, even if the share of economic resources to which one is entitled, as a matter of justice, is itself determined at least in part by the way in which her welfare would be affected by different economic distributions, it will nonetheless be the case that there is some particular share of resources to which she is entitled. And since the share of resources to which she is entitled will itself be a function, at least in part, of how resource shares would, for her, map onto (expected) welfare levels—if she were to receive a windfall, and therefore have more resources than she is entitled to as a matter of justice, then she would also tend to have more (expected)

---

10 The most compelling evidence that Rawls viewed the relevant procedural requirements as both necessary and sufficient for distributive justice can be found in *Political Liberalism* (1993, 282). For criticism, see Murphy (1998, 287), and Berkey (2015; 2016, 715–718; 2018, 730–732, 744–747; 2021, 183–185, 197–204).

11 I am grateful to an anonymous reviewer for encouraging me to clarify this point.
welfare than she is entitled to.\textsuperscript{12} Because this is the case, my argument will not imply that it is possible for one to receive a windfall and yet have less than one is entitled to in terms of welfare (or any other non-resourcist possible currency of justice).\textsuperscript{13}

It is also worth noting that while my account of what constitutes a windfall relies on a distinction between gains that result from (relevant forms of) careful and strategic planning and those that do not, I do not and cannot here provide a complete account of how this distinction should be understood. And while I suspect that in some cases it will be relatively uncontroversial that gains either are or are not the result of careful and strategic planning, there are a number of issues that a complete account would require taking a position on that would be relevant to assessing more difficult cases. For example, we would need to take a position on whether the fact that one has engaged in careful and strategic planning for purposes other than achieving economic gains makes it the case that gains that one obtains as a result do not constitute a windfall. In other words, we would need to determine whether one must be intentionally pursuing economic gains in particular in order for her careful and strategic planning to render any gains that she obtains exempt from the obligation to redirect windfall gains. In addition, we would need to take a position on whether careful and strategic planning that aimed to generate economic gains in one way, but which happens to generate gains in a different, entirely unexpected and unplanned way, makes it the case that the gains do not count as a windfall. My own inclinations tend toward a fairly

\textsuperscript{12} This is because, all else being equal, increases in resources tend to generate increases in (expected) welfare. The exceptions will be cases in which a person’s gaining in terms of resources would have no effect on her expected welfare level. This is likely true for many, if not most, of the very wealthy. In these cases, those who hold that welfare is the currency of justice, or a significant component of that currency, should be especially inclined to think that those who receive windfalls are obligated to redirect them to people who are unjustly disadvantaged.

\textsuperscript{13} It might be objected that at least some people who have received economic windfalls during the pandemic have nonetheless been made worse off in terms of welfare, so that despite having more resources than they are entitled to, they have less welfare than they did previously, and perhaps less than they were entitled to as a matter of justice. However, while it is likely true that some people who have received windfall economic benefits during the pandemic have nonetheless suffered greater overall welfare losses than most others, the fact that the pandemic has negatively affected the welfare of the vast majority of people makes it the case that those who were quite well off before the pandemic will, on welfarist or partially welfarist views, be entitled to less in terms of welfare than they were previously—there is simply less (potential) welfare that can be achieved in a world dealing with a global pandemic. Because of this, at least many who have received windfalls will, if my argument is right, be obligated to redirect at least some of these gains, even if they have been made worse off in terms of welfare.
broad account of when gains count as windfalls, so that plans that unintentionally produce gains, and plans that produce gains in unplanned ways, do not exempt the gains from the requirement to redirect windfalls. A full defense of this view, however, must be left for another occasion.

III. PANDEMIC WINDFALLS AND OBLIGATIONS OF JUSTICE

On outcome-focused accounts of distributive justice (for example, sufficientarianism), the case for the obligation to redirect windfall gains is fairly straightforward. First, since any redirection of resources from a person who has more than she is entitled to as a matter of justice to someone who has less than she is entitled to would make the resulting distribution less unjust, there is always at least some moral reason in favor of the redirection. In order for it to be permissible to refrain from redirecting, then, a justification is required that is sufficient to counteract the force of the reason in favor. Perhaps the most plausible justification that could be offered for refraining from redirecting resources that one possesses beyond what she is entitled to as a matter of justice is that one has, within an admittedly unjust system, carefully and strategically made choices that resulted in gains beyond one’s justice-based entitlements, in order to limit one’s risk of becoming one of the people that the unjust system allows to be unjustly disadvantaged. If this justification is offered alongside the acknowledgement that one is, along with everyone else, obligated to work to make the system just (or at least less unjust), even if the success of that effort would result in, for example, one’s becoming subject to increased taxation, it is at least not obvious that it should be rejected.  

This justification, however, is not available in the case of windfalls. By their nature, windfall gains are not attributable to careful and strategic planning on the part of those who benefit from them. Instead, they result from simple good fortune. When one experiences good fortune that renders her better off than she is entitled to be as a matter of justice, while others who are unjustly disadvantaged suffer severe hardships, the

14 See, for example, Rawls’s discussion of the “natural duty of justice”, which requires us to “further just arrangements not yet established, at least when this can be done without too much cost to ourselves” (1999, 99).

15 Because on outcome-focused accounts, redirections from the unjustly advantaged to the unjustly disadvantaged necessarily constitute improvements with respect to justice, this justification cannot plausibly succeed in cases in which individuals, even as a result of their own careful and strategic planning within an unjust system, possess vastly more resources than they are entitled to as a matter of justice, while others are unjustly extremely badly off.
reasons for thinking that she is obligated to redirect her windfall benefits seem especially strong. When an event such as a global pandemic (perhaps in combination with inadequate policy responses from governments) consistently delivers good fortune to those who were already very well off (and, on essentially all plausible views unjustly advantaged), while at the same time making millions of poor and working class people worse off than they were (and therefore even more unjustly disadvantaged), the prospects for justifying refusal by the beneficiaries to redirect their windfall gains seem dim. The gains in terms of justice would be too great, and the grounds that they might offer for refusing are too limited and implausible.

It is, however, somewhat less clear that an obligation to redirect windfall gains can be defended within procedural accounts of distributive justice. This is because, on these accounts, redirection (from those with more than they are entitled to as a matter of justice to those with less) that is not accomplished via the procedures that constitute the fundamental requirements of justice need not count as making the resulting distribution any less unjust than the initial distribution. If a just distribution is defined as a distribution that results from the following of just procedures, then even if their tendency to bring about distributive outcomes with certain substantive features is what makes particular procedures the ones required by justice (as is the case on Rawls’s view), actions that bring about similar results by means that do not run through the required procedures, such as direct transfers from well-off individuals to badly-off individuals, cannot, as a conceptual matter, make a distribution less unjust.

Procedural accounts are often motivated by the thought that distributive justice is fundamentally the responsibility of state institutions, and not of agents acting within those institutions, such as individuals or firms (Rawls 1993, 268–269). Proponents of these accounts typically hold that while individuals and other non-state agents are obligated to contribute to transforming unjust procedures into just ones, they are not obligated, as a matter of justice, to redirect unjust advantages that they possess in virtue of the operation of existing unjust procedures to those who are unjustly disadvantaged by the operation of those procedures. This is, on these views, part of what follows from the view that distributive justice is fundamentally the responsibility of state institutions rather than agents acting within them.
As I suggested above with respect to outcome-focused accounts, it might be argued that when individuals are able, through careful and strategic planning, to acquire more in the way of resources within an unjust system than they would have been able to acquire had the system been just, they are permitted to refrain from redirecting what they possess beyond their just entitlements to those who are unjustly disadvantaged. We might think, for example, that when unjust advantages and disadvantages occur as a predictable result of flawed policies implemented by states, individuals’ conscientious efforts to improve their own lives within the unjust system entitle them to keep what they have acquired, at least so long as they also satisfy their obligation to support the institutional changes required by justice.

Even if this line of argument is defensible, however, it does not provide grounds for concluding that individuals are permitted to refrain from redirecting windfall gains. There are at least two reasons to doubt that this extension can be defended. First, windfall gains are not attributable to individuals’ conscientious efforts within an admittedly unjust system, but instead result from (often unpredictable) good fortune. In these cases, the same reasons that explain why policies that tend to bring about certain distributive outcomes are the ones required by justice will also support individual obligations to redirect gains, and these reasons will not be counterbalanced by reasons in favor of a permission not to do so grounded in individuals’ conscientious efforts to realize those gains.

More importantly, however, at least some windfall gains, including those that have resulted from the pandemic, are not outcomes that it is reasonable to expect state policies to be designed to fully prevent ex ante. State policies must be designed to predictably bring about just results, as much as possible, in the ordinary course of life in a society. Of course, generally applicable policies can and should be adopted that aim, broadly speaking, to mitigate the negative effects of unpredictable events such as the pandemic. But no such policies can be expected to be able to fully redress the negative, justice-relevant effects of every possible large-scale unpredictable event or set of events. Instead, as we have seen during the pandemic, states must respond to such events, as they are happening, with policies that aim to limit their negative effects. And implementing such policies takes time, even when states are functioning reasonably well. This leaves those who suffer unpredictable and unjust disadvantages unavoidably waiting for relief, which may or not be forthcoming. Those who enjoy windfall gains can help provide the required relief by
redirecting those gains to the unjustly disadvantaged.16 Because it is not possible for policies to be designed in advance that could reliably and fully redress the negative effects of any unpredictable event such as a pandemic, the reasons that might generally justify attributing exclusive direct responsibility for ensuring distributive justice to states do not apply in cases involving large-scale unpredictable windfall gains for some and significant losses for others.

IV. CONCLUSION

If my argument in the previous section is correct, then those who have received windfall gains as result of the pandemic are obligated to redirect those gains in ways that would benefit the unjustly disadvantaged. Importantly, there are reasons to think that it is not just billionaires such as Musk and Bezos that have received such gains.17 Many of those who are well off but far from billionaires likely have as well. If one has not lost her job or had her pay cut, has enjoyed more limited transportation expenses in virtue of working from home, and perhaps owns some stock, the value of which has been propped up by government policies, while less has been done to protect the poor and working class from the effects of the pandemic, it seems likely that she has received at least some windfall gains, and is obligated, if I am right, to redirect them.

REFERENCES


---

16 As an anonymous reviewer rightly points out, not every case in which one receives windfall gains is one in which individuals will be able to act to aid those who are unjustly disadvantaged before governments can. For example, the stimulus checks that many Americans have received may constitute windfall gains for some, but of course they cannot redistribute those gains until the government sends them the funds. This is, however, an example in which policy likely cannot be calibrated in a short period of time to reliably direct all of the relevant funds in the ways require by justice—a blunt policy that simply sends checks to all citizens with incomes below a certain threshold may be the best that can be done in short order. But because this policy will unavoidably result in windfall gains for some, there are reasons to think that there can be obligations to directly redistribute those gains to those who remain unjustly disadvantaged.

17 While it seems to me quite unlikely that those like Musk and Bezos have received no windfall gains as a result of the pandemic, it is of course difficult to estimate how much of any particular person’s gains during the relevant period constitute windfall gains.


Brian Berkey is Assistant Professor in the Legal Studies and Business Ethics Department in the Wharton School at the University of Pennsylvania, and holds a secondary appointment in the Penn Philosophy Department. He works in moral and political philosophy, including environmental and business ethics, and has published articles on moral demandingness, individual and corporate obligations of justice, climate change ethics, exploitation, effective altruism, collective obligations, and animal ethics. Contact e-mail: <bberkey@wharton.upenn.edu>
Governing Life and the Economy: 
Exploring the Role of Trust in 
the Covid-19 Pandemic

JOELLE M. ABI-RACHED  
Sciences Po, Paris

ISHAC DIWAN  
Ecole Normale Supérieure and Paris Sciences et Lettres

Abstract: When comparing both GDP loss and mortality across countries, it appears that countries that have managed to save more lives during the Covid-19 pandemic have also managed to save their economies better. What accounts for these stark differences in country performances? In this article, we argue that a salient feature of economic and health performance is the degree of trust populations have in their governments. We set up a heuristic analytical framework that models this relation, under particular assumptions about what drives government and individual behavior, in order to better understand the mechanisms that may be at work. We identify three key roles that trust in government may play in enforcing social distancing policies, conveying credible information for individual decision-making, and shaping government attitudes towards risk. We argue that these implications are consistent with the empirical evidence. We also discuss the relevance of other forms of trust, namely, interpersonal trust and trust in science.

Keywords: trust, trust in government, Covid-19, pandemic, economic growth, mortality, trust in science, interpersonal trust

JEL Classification: A13, C7, I15, I18, J18

AUTHORS’ NOTE: Different versions of this paper were presented by Abi-Rached at the London School of Economics and Political Science (LSE), “Lives versus Livelihoods: Evaluating policies to address COVID-19”, hosted by the Global Health Initiative and the Department of Philosophy, Logic and Scientific Method, October 27, 2020, and Sciences Po (Médialab), “Gouverner la vie et l’économie : le rôle de la confiance à l’ère de la COVID-19”, November 3, 2020. Abi-Rached thanks Alex Voorhoeve (LSE) and Guillaume Lachenal (Sciences Po) for their kind invitations and the participants for their helpful suggestions and comments. We also thank the anonymous reviewers as well as the editors of this journal for their generous comments on an earlier draft. Last but not least, we are grateful to Jamal Ibrahim Haidar for his valuable advice and friendship.
I. INTRODUCTION

One salient feature of the Covid-19 pandemic compared to previous ones is that, for the first time, governments around the world have chosen to implement very costly measures, such as national lockdowns. These measures were met with unequal success in reducing Covid-19 related mortality. They have also resulted in economic losses (reflected in large drops in GDP) both through their effects in reducing the demand for goods and services, and because of impediments to production.

Much of the recent literature on the challenges posed by Covid-19 on public policies has focused on the trade-off between economic and health performance. But when we compare performance across countries, it appears that countries that have managed to save more lives during the first wave of the Covid-19 pandemic have also managed to save their economies better. What accounts for this positive relation, and for the stark differences in country performances? In this article, we propose that a striking determinant of both economic and health performances may be the degree of trust populations have in their government.

Historically, from the plague to the more recent coronavirus outbreaks (Severe Acute Respiratory Syndrome or Middle East Respiratory Syndrome) and the Ebola epidemics, public authorities have had recourse to a number of constraining measures (quarantine, isolation, the closure of borders and schools, surveillance, and so on) that restrict the circulation of people and goods as an attempt to reduce the spread of an epidemic (Tognotti 2013; Hawker et al. 2019, 147, 359, and 394). And, as with earlier pandemics (Gilles et al. 2011; Prati, Pietrantoni, and Zani 2011; Quinn et al. 2013; Blair, Morse, and Tsai 2017), several empirical studies on Covid-19 have shown the existence of a positive relation between compliant health behavior (such as respect of stay-at-home orders) and trust in government (Bargain and Aminjonov 2020; Bicchieri et al. 2021; Brodeur et al. 2020; Elgar, Stefaniak, and Wohl 2020). But to the best of our knowledge only a few studies have looked at how trust in government has affected health performances during the Covid-19 pandemic (Elgar, Stefaniak, and Wohl 2020). And no studies have looked at the effect of trust in government on both health and economic performances. Relatedly, the mechanisms underlying this relation have not been conceptualized clearly. Finally, besides trust in government, several other forms of trust come up in the literature with insufficient clarity as to their differential impacts—this is particularly the case for interpersonal trust.
and trust in science (Elgar, Stefaniak, and Wohl 2020; Bicchieri et al. 2021; Borgonovi and Pokropek 2020).

Our goal in this article is to explore how trust in government (TG thereafter) can enable better social distancing policies to ameliorate health outcomes at low economic cost. While we observe correlations between these variables, we do not seek to demonstrate that particular forms of trust cause performance. Such an investigation would be replete with endogeneity issues, unless a convincing identification method is found, which would require data that we do not have access to. What we do instead is develop a coherent and plausible theory about the possible determinants of this relationship, under reasonable assumptions about what drives government and individual behaviors under health risk. The basic logic of the model is simple: the greater TG is, the easier it is for government policy to reduce the incidence of the virus at low cost, and thus the fewer lost lives and less economic losses a country will experience.

The model allows for a logical exploration of the various ways in which TG affects outcomes. We identify three key roles that TG can play in the context of this pandemic: (1) in enforcing social distancing rules, (2) in convincing people of the seriousness of the health risks, and (3) in affecting a government’s attitude towards risk. We illustrate these mechanisms by comparing particular country cases, and we argue that the implications of the theory we present are broadly consistent with the results of a wide range of empirical studies. We conclude with a discussion of the possible effects of other forms of trust on performance.

II. LIVES AND LIVELIHOODS
Now that we have some hindsight of the Covid-19 pandemic that emerged in January 2020, it is possible to compare countries’ performances along both the economic and health dimensions (Figure 1). The data is quite spread out, illustrating a broad range of country experiences. Yet strikingly, there appears to be a highly significant positive relation between economic and health performance (the correlation coefficient is −0.35): countries that have been able to save more lives have also managed to preserve their economy better (and vice versa). This is surprising, as one

---

1 We define social distancing policies as all government-mandated non-pharmaceutical interventions (NPIs) that can reduce the circulation of the virus (including mask wearing, national lockdowns, state-at-home orders, restriction on mobility, and so forth).
could have expected a negative relation, given the focus of much of the literature on the trade-off between the goals of preserving lives versus livelihoods. But while this trade-off must have been present within each country, the evidence does not suggest that it describes well differences across countries. Instead, there must be other factors that explain why some countries have managed to perform well on both goals, while others have not.

There are many possible factors that can affect both economic and health outcomes in the face of the Covid-19 pandemic shock: the extent to which a country is globalized (Bouhaj 2020), demographical characteristics (Onder, Rezza, and Brusaferro 2020), levels of poverty (Lou, Shen, and Niemeier 2020), and inequality (Banik et al. 2020; Elgar, Stefaniak, and Wohl 2020). Among these variables, only globalization affects both health and economic outcomes in the same direction. But globalization in itself cannot explain the variation in performance across countries—for example, European countries are more or less equally globalized, but they ended up with different performances.
Equally, other country characteristics, such as a country’s level of income (and relatedly, the fiscal and/or health capacity of a country), and the type of its political regime, do not seem to explain the variation in performance. Some lower middle-income countries did well, while others did not; and some high-income countries did well, while others did not. One may think that authoritarian countries may find it easier to sanction non-cooperative behavior (The Economist 2020). But while some autocracies did well (for example, China), others did not (for example, Iran), and some but not all democracies did well.

In the rest of the article, we explore the possibility that trust variables can help explain these differences. It is reasonable to expect the role of trust in government (TG) to be important for controlling the spread of Covid-19 because it is necessary both for incentivizing governments to implement strict social distancing policies (lockdowns, mandatory mask wearing, the shutdown of certain economic activities, and so on), and for these measures to be efficient (in the precise sense of improving health outcomes at low economic cost). We also explore the roles of interpersonal trust (IT) and trust in science (TS).

To get a feel for the statistical relation between various trust variables and performance, we start by looking at country-level measurement from the World Value Survey (WVS), the Arab Barometer (AB), and the Wellcome Global Monitor. We use the latest surveys just before the Covid-19 crisis hit since this is what is relevant to understanding the first wave of the pandemic.

Table 1 shows the results of a simple linear regression that seeks to measure the statistical relation between economic and health performances with various measures of trust across countries. Admittedly, our sample is relatively small, as we only managed to assemble the required data for 44–46 countries, and so the empirical results that emerge from this analysis are indicative rather than definitive. The data shows that, on the one hand, countries with high TG have done well in both controlling the epidemic and saving their economies—the coefficients of TG are highly significant at the 1% confidence level. On the other hand, the other measures of trust, namely IT and TS, do not seem to affect either economic or health performance in a significant way.\(^2\) The statistical power

\(^2\) The three trust variables we consider in this article seem to convey different information about individual values. In particular, TG is not correlated with TS (the correlation coefficient is not significantly different from 0) and only moderately correlated with IT (correlation coefficient 0.3 and significant at the 5% confidence level). However, IT and TS are highly correlated (correlation coefficient 0.7 and significant at the 1% confidence
of TG in explaining variations in economic and health performance is, moreover, quite remarkable, accounting for more than a third of the variation of economic and health performance (this is not due to multicollinearity: as seen in the right panel, the $R^2$-squared remain high when IT and TS are omitted from the model).

We develop, in the next sections, a model that can explain the possible mechanisms through which TG may predict both health *and* economic performances. In the concluding section, we also speculate on why the impact of IT and TS may not be as important as that of TG.

### III. An Analytical Framework

Rather than a formal model, we will content ourselves with a stylized framework, based on rational behavior and equilibrium considerations, which will be sufficient to shed light on the interactions involved between the key variables of interest. In our set-up, TG is a conditioning variable that affects government and individual behaviors: a high level of TG incentivizes both governments to select more stringent social distancing

---

Table 1: The impact of trust on economic growth and mortality. Ordinary least squares; standard deviation in parenthesis; $** = 1%$; $* = 5%$.

<table>
<thead>
<tr>
<th></th>
<th>Change in Growth</th>
<th>Mortality</th>
<th>Change in Growth</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG</td>
<td>.0810696**</td>
<td>−13.17787**</td>
<td>.0962007**</td>
<td>−12.94422**</td>
</tr>
<tr>
<td></td>
<td>(.0244774)</td>
<td>(3.37944)</td>
<td>(.0191073)</td>
<td>(2.522612)</td>
</tr>
<tr>
<td>IT</td>
<td>.0636367</td>
<td>−.6219912</td>
<td>(.0374541)</td>
<td>(5.17105)</td>
</tr>
<tr>
<td>TS</td>
<td>−1.225948</td>
<td>147.6503</td>
<td>(1.328985)</td>
<td>(183.4845)</td>
</tr>
<tr>
<td>Constant</td>
<td>−12.41318**</td>
<td>979.0141**</td>
<td>−12.14206**</td>
<td>1056.559**</td>
</tr>
<tr>
<td></td>
<td>(1.067636)</td>
<td>(147.4018)</td>
<td>(.9020019)</td>
<td>(119.0857)</td>
</tr>
<tr>
<td>$R^2$-squared</td>
<td>.4383</td>
<td>.3694</td>
<td>.3655</td>
<td>.3744</td>
</tr>
<tr>
<td>Number of observations</td>
<td>44</td>
<td>44</td>
<td>46</td>
<td>46</td>
</tr>
</tbody>
</table>

Notes: Change in growth rates between 2019 and 2020 from IMF; Covid-19 mortality during March–January 2020 from Our World in Data; TG: Trust in Government from WVS and AB (answers to “How much do you trust the government?”; we add the percentage that answered “A great deal” or “Quite a lot”); IT: Interpersonal Trust from WVS and AB (answers to “How much do you trust others?”; we used the percentage that answered “Trust completely” in WVS, or “Most people can be trusted” in AB); TS: Trust in Science index based on the Wellcome Global Monitor adapted from Borgonovi and Pokropek (2020).
measures and individuals to comply more with these measures (and vice versa).

Our starting point is that a basic determinant of how governments and individuals decide whether to impose or respect social distancing rules is a comparison of costs and benefits. The type of social distancing rule \( SD \) people have been faced with is of the type: ‘individual \( x \) should leave home for no more than \( y \) hours a day’. We assume that an \( SD \) rule will be adopted by a government when the rule’s overall health benefit is perceived to be larger than the overall economic cost it involves, and that people act rationally by comparing the costs and benefits of their actions.

For a government, \( SD \)-type rules can be complex. A choice of an ‘optimal’ level \( SD \) set by a government, call it \( SD^g \), would aim at containing the circulation of the virus at a certain level, in ways that minimize economic costs. Since social distancing reduces mobility and slows down the economy (consumers spend less time in shops, producers who cannot telecommute spend less time producing, and so forth), the choice of how tight social distancing is should balance health and economic concerns. Moreover, for a certain level of virus circulation, the precise nature of \( SD^g \)—who is allowed to circulate more—will aim at minimizing the economic slow-down (for example, by affording more mobility to essential workers). \( SD^g \) thus depends on a government’s welfare function (‘cost of life’)—that is, how it values economic lives over income lost—but also on how it places weights on different individuals in society, on economic structure, and on government capacity.\(^3\)

This type of cost/benefit comparison also applies at the individual level. Even with no \( SD \) policies in place, people will impose restrictions on their own mobility to reduce their chances of getting infected, as long as the cost of the implied economic hardship is below the expected gain on the health side. Those less at risk (the youth) will circulate more and continue to work, while those most at risk will reduce their interactions more, accepting a higher economic loss. Peoples’ behavior will also be affected by that of others: higher virus circulation increases health risks, and will tend to push people to stay at home more. In an ‘equilibrium’ with no government-imposed restrictions, one would expect to see a relatively high level of virus circulation, as rational actors do not take into account

\(^3\) It seems reasonable to presume that effective \( SD \) rules lead to lower circulation of the virus and hence to lower infection rates. We assume here for simplicity that this relation is strictly monotonic.
the risk that their behavior poses to others. Call $SD^n$ the ‘natural equilibrium with no policies’.

By choosing to implement social distancing as a means to reduce the spread of the novel coronavirus, government action thus aims at producing a ‘socially better equilibrium’ than $SD^n$. There are various ways in which this can happen. For example, $SD^g$ can be very constraining, with tight lockdowns and contact tracing eliminating the virus circulation rapidly, at sharp but short-lived economic costs. Alternatively, governments could aim at keeping at home the least productive members of society, while allowing, and even encouraging, so-called frontline and essential workers to keep working, so as to minimize economic costs while keeping the level of virus circulation low. Thus, $SD^g$ should be more socially desirable than $SD^n$, as long as restrictions are reasonably effective, and a government represents the broad interests of society. There might be winners and losers, but in an ideal world, the losers could be compensated (especially the poor who are required to stay at home and stop working), and thus everyone could benefit.

The potential social gains driving governments to optimize $SD$ rules are directly related to the benefit of coordination in an environment replete with agency costs. First, a government would tend to be more risk-averse than self-interested individuals, as it also values the benefit of people not infecting others. In addition, in countries where the cost of healthcare is socialized, the government can be more risk-averse than individuals in relation to their own health, as it bears more of the cost related to the care of sick people. Second, there is a difference between who is optimally allowed to circulate, and who has incentives to circulate. While individuals care about not being infected, in the absence of rules, those at low risk would evaluate the trade-off less tightly than those at a higher risk—with little consideration for the impact of their circulation; and equally, those who fear getting infected most would tend to skip work, even if it is socially desirable that they do not (for example, healthcare workers). Third, information may be asymmetrically distributed. Some governments may hide the real risks for various reasons, and people may not believe the information communicated by governments about the risks posed by the pandemic.

In the simple analytical framework we have just outlined, we now propose three ways in which TG could play an important role in shaping and enforcing $SD$ policies, which we discuss in the subsequent sections: first, a high level of TG could allow for the enforcement of $SD^g$ rules more
effectively by reducing free-riding incentives (an enforcement issue); second, it could help convince citizens more rapidly that the health threat is serious, before the spread of the virus starts exacting visible health and economic costs (an informational issue); and third, it could allow governments to engage in early action, even as the health situation is still replete with uncertainties (an attitude to risk issue). In all these cases, governments with high TG are thus able to curb the spread of the virus more efficiently (that is, at lower economic cost).

IV. ENFORCEMENT OF SOCIAL DISTANCING

A first direct role that TG could play is related to the enforcement of sanctions, which is necessary to deter free-riding, since the socially optimal rule set by a government, \( SD^g \), will be for many individuals above their own preferred \( SD \). For \( SD^g \) to hold, sanctions against those who do not respect rules must be high enough so that individuals with an \( SD \) lower than \( SD^g \) respect the established guidelines. In the end, the actual \( SD \) that is observed, call it \( SD^a \), will be somewhere between \( SD^g \) and \( SD^n \), closer to the first when \( SD \) rules are more effective, and closer to the second when they are not respected.

While governments may threaten costly sanctions, they may fail to enforce them, which is likely to be especially the case when TG is low, as governments may then fear stoking social unrest. This can be the case if there are large and powerful groups for which the distance between \( SD^g \) and \( SD^n \) is high, for example among the poor in unequal countries with no social safety nets.

If people discover that actual penalties are low, some will tend to cheat more. But there are also secondary effects. Since those who cheat expect others to cheat too, they now realize that \( SD^g \) is likely to have to last longer in order to achieve its health targets. For those poor enough, this increases exponentially their economic costs (for example, they could have afforded, with difficulty, two weeks without income, but not a month). Similarly, some among those whose work is deemed valuable and who are not constrained by \( SD \) would stop working if they were from a high-risk health group, hence hurting the economy more. At the limit, faced with many cheaters, the sanctions regime collapses, and the equilibrium approaches the natural situation \( SD^n \).

If we allow IT and TS to vary, the needed level of sanctions to enforce compliance will vary as well (see the last section for a discussion).
Conversely, if TG were high, people would tend to believe that the regime \( SD^g \) would hold and that the virus circulation would be controlled, and infection risks would rapidly fall in the future. This encourages essential workers to work, and low productivity/younger workers to be more willing to restrain their free-riding incentives. As a result, health performance is higher (lower circulation of the virus means lower mortality). Hence the effectiveness of \( SD \) depends on high TG.

These considerations can help explain the large variability in the observed effectiveness of social distancing interventions in curbing the spread of the virus. Several prominent studies (Dehning et al. 2020; Hsiang et al. 2020; Flaxman et al. 2020) have studied empirically the role of government-mandated non-pharmaceutical interventions (NPIs) in reducing the transmission of Covid-19, showing that these policies had a large impact on the transmission rate of the disease in the early phase of the pandemic. However, other studies have questioned these results and have argued that NPIs have at best marginal impact (Atkeson, Kopecky, and Zha 2020; Bendavid et al. 2021; Bjørnskov, forthcoming; Lin and Meissner 2020). These mixed results concerning the effectiveness of NPIs are illustrated by Bjørnskov (forthcoming) who explores the association between the severity of lockdown policies in the first half of 2020 and mortality rates in 24 European countries. The “hard lockdown group” (Bjørnskov, forthcoming, 5) registered 372 additional deaths per million while the other group registered 123 deaths per million. This either means that social distancing policies make things worse, which is unlikely, or that there is enormous endogeneity. Our framework offers one plausible explanation of this endogeneity: countries with low TG can end up imposing longer, but less effective, lockdowns and still end up with higher mortality than high TG countries.

Because of effective \( SD \) policies in curbing the spread of the virus, high TG countries are able to put in place more efficient \( SD \) policies (ceteris paribus) allowing more people to go back to work faster, which allows for a faster economic recovery after a short period of contraction. By efficiency we mean that the stringency of these measures does not have to remain high for a long time or can be implemented less often; in other words, they are less costly in terms of economic loss while still leading to better health outcomes. In contrast, \( SD \) measures are less efficient in low TG countries with low compliance in the population, which requires
recourse to more frequent and longer lockdowns, obliging the government to close the economy more frequently, causing both more infections and sharper economic downturns.

What we observe is consistent with these considerations. Generally, in countries with similar income levels and economic structure, higher TG levels allowed shorter public health interventions to be implemented while at the same time keeping infection rates lower (than in countries with lower TG). We can illustrate this by comparing Germany (TG 44%) and France (TG 31%). During the first wave of the pandemic, the former managed to implement more efficient measures. While Germany kept the stringency of its social distancing high for a shorter period of time (Figure 2a), it managed to achieve better health outcome (Figure 2b) at lower economic costs: Germany’s GDP ended up shrinking by 6.6%, better than France where GDP fell by 11.3%.

V. ASYMMETRIC INFORMATION
People may have trouble evaluating the risk of infection of a new virus, and may think that the government is not honest in its evaluation of risks because of political motives. This is not unreasonable; governments face a number of political trade-offs when devising an SD strategy, which includes slowing down international trade and travel, supporting the survival of businesses and the welfare of individuals, accelerating medical responses, and deciding how to manage schools (Migone 2020). This leads to a lot of lobbying activity, and thus to political trade-offs between the needs of diverse groups, which are hard to manage early in a pandemic when containment measures appear to be out of scale with the perceived level of threat (Tyshenko and Paterson 2010; Kushner Gadarian, Goodman, and Pepinsky 2020). Especially when TG is low, governments may find it difficult to elicit cooperation across sectors and lobby groups. As
a result, individuals may have more reasons to shirk. For example, members of the opposition are likely to suspect that a restrictive SD strategy is imposed to weaken them (for example, demonstrations are forbidden, elections are delayed). Equally, some may think that to gain votes, the government is pandering to the elderly (by exaggerating risks), or to business lobbies (by understating risks), or to populist voters (by exaggerating the necessity of border closures).

The point here is that if people do not trust the government, they are more likely to interpret its actions as politically driven than scientifically motivated. An important implication of this asymmetric information is that governments with low TG may need to signal credibly the severity of the pandemic by allowing the healthcare system to become overburdened before implementing harsher measures, reducing the efficiency of their intervention.

This is consistent with what happened in Italy (TG 24%). During the first wave of the pandemic, the 'Italian scenario' became the benchmark worst-case scenario (Pinedo and Carreño 2020). The same can be said of Spain (TG 20%) which implemented a lockdown about a month after the first case of local transmission was detected. The Spanish healthcare system was quickly overwhelmed with countless lost lives (Organisation for Economic Co-Operation and Development 2020; Working Group for the Surveillance and Control of COVID-19 in Spain 2020). Another example is France (TG 31%) that lost two crucial weeks before it took the decision to act decisively. Arguably, the French hesitancy had to do with its historically low TG level (Galland and Grunberg 2020), which makes decision-makers risk-averse. Indeed, fearing accusations of exploiting the public health emergency for its own political gains, the French government decided to maintain the municipal elections (Briatte, Neihouser, and Kelbel 2020).

**VI. ATTITUDE TO RISK**

A third aspect of TG has to do with governments’ attitudes towards risk. So-called zero-Covid or suppression (rather than ‘mitigation’) strategies have so far only been successful in high TG environments. Taiwan (TG 52%), South Korea (TG 51%), or Vietnam (TG 92%), managed to either suppress or keep infections at near zero levels rapidly resulting in very low economic losses.

A zero-Covid strategy requires two conditions. First, preparedness. But a high degree of preparedness is costly to maintain and thus requires
a public that trusts that its government is acting in the public interest, as opposed to pandering to particular constituencies, or worse, overspending because of corruption. Many Asian countries had anticipated the Covid-19 pandemic precisely because they had learned from previous pandemics. They had put in place early and robust surveillance systems and expanded their healthcare capacity to sustain such shocks (Pollack et al. 2020). As a number of social scientists have argued (Jalan and Sen 2020; Chathukulam and Tharamangalam 2021) in the context of Kerala in India (which had performed exceptionally well during the first wave of the pandemic), preparedness and pro-active actions by the state government in turn helped build trust, leading to greater compliance with public health measures in a kind of “virtuous” synergy (Chathukulam and Tharamangalam 2021, 10). An argument has been made for countries to “mobilize and transition” to a zero-Covid strategy, even in the case of under-preparation, by building on efforts spent on the first wave of the pandemic (Allen et al. 2020, 17). New Zealand, a country with high TG (50%), followed precisely this approach given its unpreparedness to face the pandemic after years of low spending on healthcare (Daadler 2020).

The second condition is quick, decisive, and early action. Such an approach requires tight contact tracing and drastic measures to quickly extinguish clusters when they appear—often at high localized costs (for example, cordonning off an entire neighborhood or a city when only one case is signaled). In this case, some citizens are subject to extreme pain, and enforcement may require the threat of highly punitive sanctions. Governments with low TG are unlikely to bet on such a risky strategy. There is always the possibility that these costly efforts would not be rewarded with public recognition: the virus itself may dissipate rapidly for exogenous reasons, or—if government efforts manage to stop the virus—people may believe that this success was due to exogenous causes and not to government intervention. In all such situations, early movers are at risk of being overturned for constraining individual freedoms for no credible reason. That is why, we believe, early movers act only if the public is perceived to be on their side.

VII. IN LIEU OF A CONCLUSION
Many governments around the world have chosen to implement costly lockdowns in order to save lives. The countries that managed to slow down the pandemic more efficiently—better health outcome at lower economic cost—seem to have one characteristic in common: high TG.
While our article has focused on the role of a particular form of trust, namely trust in government, other forms of trust—notably interpersonal trust (IT) and trust in science (TS)—have been discussed in the literature. Both put more emphasis on the effect of individuals’ own actions to protect themselves, rather than on public policy, in curbing the pandemic. Nevertheless, it is possible to argue that unlike TG, these variables do not have an unambiguously positive effect on health and economic performances. Our cross-country empirical results in Table 1 also suggest that high levels of IT and TS do not seem to have been sufficient to mobilize communities to make costly behavioral changes.

In the context of our analytical framework, we can imagine that both low and high IT push individuals to respect $SD$ more; in the first case, in order to protect themselves from infections (they suspect others may be infected), and, in the second case, to protect others from infections (they fear that they may themselves be infected). On this view, mid-level IT individuals are those least likely to reduce their mobility voluntarily. Another view is that individuals with high levels of IT may have a harder time accepting the mandate to isolate themselves from their close peers. Indeed, while Bicchieri et al. (2021) find a positive but small relation between IT and compliance using experimental methods (in nine countries), Elgar, Stefaniak, and Wohl (2020) find that actual mortality was positively related to IT (in a large set of 84 countries). These considerations suggest that in general relying on IT is unlikely to be sufficient to enforce social distancing, and that a sanction regime will remain necessary.

Similarly, a high-level TS is also unlikely to unambiguously improve health and economic performances by itself. On the one hand, people who trust science more can agree to costly behavioral changes because they believe science provides a better guide. But, on the other hand, they may also estimate (correctly) that their personal risks are low, pushing them to shirk $SD$ rules more often in the absence of a sanction regime. Indeed, while some studies claim to show that TS improves compliance with social distancing (Bicchieri et al. 2021), others actually find a negative effect (Borgonovi and Pokropek 2020).

---

5 A similar logic is found in the economic literature. In their article on the ‘twin peaks curve’, Algan, Cahuc, and Sangnier (2016) show the existence of a ‘twin peaks’ relation between IT and the size of the welfare state: uncivic people with low IT favor large welfare states because they expect to benefit from them without bearing their costs, while civic individuals support generous benefits and high taxes only when they are surrounded by trustworthy individuals.

6 This result may be due to their constraining the data to fit along a line rather than a curve.
In this exploratory article, we have offered an explanation for these stark differences in country performance by presenting a theory linking trust in government with the ability of governments to manage the Covid-19 crisis. The theory allows us to derive several implications about how governments intervene, and about the effectiveness and efficiency of their intervention. The initial empirical evidence we presented is broadly consistent with the implications of the model.

In contrast with other aspects of trust, we have argued that TG can only improve government regulations. In the context of Covid-19, there is little downside to high levels of TG. This does not mean that all high-TG countries decided to implement social distancing policies. In contrast to its Nordic neighbors, Sweden (a country with high TG) chose not to impose tight lockdowns, preferring to rely more on individual decisions and ending up with much higher lost lives (Conyon, He, and Thomsen 2020). But for governments implementing social distancing policies, higher TG can, conceptually speaking, only help make regulation more efficient, a result consistent with our empirical exercise.

REFERENCES


Joelle M. Abi-Rached is associate researcher at the Médialab, Sciences Po Paris. She holds an MD from the American University of Beirut, an MSc in philosophy and public policy from the London School of Economics and Political Science and a PhD in history of science from Harvard University. She has a number of publications on the social, political, historical, and economic dimensions of healthcare and medicine. Her recent book is ‘Asfuriyyeh: A History of Madness, Modernity, and War in the Middle East (MIT Press, 2020).
Contact e-mail: <joelle.abirached@sciencespo.fr>

Ishac Diwan is professor of economics at the Ecole Normale Supérieure and Paris Sciences et Lettres where he holds the Chair on the Socio-Economy of the Arab World. He holds a PhD in economics from the University of California at Berkeley. He was the World Bank director in East and then West Africa during 2002–2011 and has held recent teaching positions at Columbia University and Harvard University. He has published widely on international finance, development strategies, and the political economy of the Middle East. His recent book—co-authored with Adeel Malik and Izak Atiyas—is Crony Capitalism in the Middle East (Oxford University Press, 2019).
Contact e-mail: <ishac.diwan@ens.psl.eu>
How Economists Ignored the Spanish Flu Pandemic in 1918–1920

MAURO BOIANOVSKY
Universidade de Brasilia

GUIDO ERREYGERS
University of Antwerp and University of Melbourne

Abstract: The current Covid-19 pandemic has attracted significant attention from epidemiologists and economists alike. This differs from the 1918–1920 Spanish influenza pandemic, when academic economists hardly paid attention to its economic features, despite its very high death toll. We examine the reasons for that by contrasting the ways epidemiologists and economists reacted to the Spanish flu at the time and shortly after the pandemic. We also explore, but less extensively, some economic and epidemiologic writings during the twenty-five years that followed.

Keywords: Spanish influenza, economists, business cycles, epidemiologists, labour supply

JEL Classification: B19, B49, I19

I. INTRODUCTION

The Covid-19 pandemic is not only a severe health crisis—as of March 2021, at least 100 million cases have been recorded and more than 2.5 million people have died globally, according to the data of Johns Hopkins University (2021)—but it is also a major economic shock. The World Bank estimates that the world economy has shrunk by 4.3% in 2020,
while government debt levels have increased sharply, and extreme poverty is on the rise again (World Bank 2021).

The scientific community has reacted by investing massively in Covid-19 related research. Several vaccines have been developed and approved in record times, and many more are in the pipeline. Epidemiologists have become media stars in many countries, instructing us to practice social distancing and to wear facemasks, and informing us about the meaning of the $R_0$ statistic and herd immunity. As soon as it was clear that the coronavirus epidemic had turned into a pandemic and one country after another went into some form of lockdown, economists began examining what would be the immediate and long-term effects of the pandemic. An impressive and fast increasing flow of papers has been the result; in the US, for instance, the National Bureau of Economic Research (NBER) has published more than 350 pandemic-related working papers, and in Europe, the Centre for Economic Policy Research (CEPR) has created the journal *Covid Economics: Vetted and Real-Time Papers* in order to rapidly disseminate the results of Covid-19 related economic research (see Coyle 2020).

One way to get a better understanding of the effects of the current pandemic is to look at previous epidemics, and this is what epidemiologists as well as economists have done. The devastating influenza pandemic of 1918–1920 has been a focal point of attention for both. Economic historians have studied the Spanish flu (also called the Great Influenza Pandemic) since the 1990s; what is different is that macroeconomists are now turning to the data of that period to learn the dynamics of an epidemic and to assess the effects of public health interventions (for example, Correia, Luck, and Verner 2020). There seems to be growing awareness that epidemiologists and economists will mutually benefit from enhanced exchange and cooperation, as illustrated by the contributions to the recent symposium on “Economics and Epidemiology” in the *Journal of Economic Perspectives* (Avery et al. 2020; Murray 2020).

In this paper, we also focus on the Spanish flu pandemic of 1918–1920, but with a somewhat different perspective. Our aim is to explore how economists at the time reacted to the pandemic. In view of the prompt and massive reaction of economists today, our expectation was that something similar must have happened a century ago. As a matter of fact, the Spanish flu pandemic was much more severe than the coronavirus pandemic in terms of its mortality rate (so far), and therefore the economic shock must have been significant. To our surprise, this
was not the case: it turns out that economists neither paid much attention to the economic effects of the pandemic as it developed, nor discussed it retrospectively in any detail in the following decades. Indeed, the Spanish flu has been regarded as the “forgotten pandemic”, since the only people who studied it at the time and afterwards were actuaries (employed by insurance companies), epidemiologists, and medical historians (see Crosby [1989] 2003; Spinney 2017). We examine the reasons for that by contrasting the ways epidemiologists and economists reacted to the Spanish flu at the time and shortly after the pandemic. We also explore, but less extensively, some economic and epidemiologic writings during the twenty-five years that followed. Our main goal is to highlight the silence of economists regarding the Spanish flu; additionally, we speculate about possible causes for this neglect.

II. A FEW FACTS ABOUT THE SPANISH FLU PANDEMIC
Medical data about the 1918–1920 influenza pandemic are not precise, but it is beyond doubt that it was one of the deadliest pandemics ever. It spread in three waves: in March 1918, the first wave begun in Midwestern US and spread to Europe, Australia, China, and North Africa; the second and more deadly wave started in France in August and quickly circulated around the world; the last wave was not as strong and hit some countries at the beginning of 1919 (Johnson 2006, 37–63). The last cases occurred in mid-1920 in Japan, Chile, and Peru (Spinney 2017, 45). Most of the deaths took place between mid-September and mid-December 1918, which coincided with the last phase of World War I. According to estimates by Patterson and Pyle (1991), the world death toll was in the range between 24.7 and 39.3 million people. India (between 12 and 20 million) and China (between 4 and 9.5 million) had the highest absolute numbers. About 550,000 died in the US, with a mortality rate of 5.2 deaths per thousand. Numbers for Europe were approximately 2.3 and 4.8 million, for mortality rate and total death toll respectively. The name *Spanish Influenza* derives from the fact that Spain was the first country to have its civilian population deeply hit by the flu (in May 1918); accordingly, the first reports about it appeared in the Spanish press, which was not censored as the country was neutral during the war (Spinney 2017, 63). Whereas the pandemic death toll in the US was around 5 times bigger than the number of war casualties (as the US entered the war as late as April 1917), the proportions were distinct in the main European belligerent countries, where more people died from the
war than from the flu: 6 times more in France, 4 times in Germany, 3 times in Great Britain, and 2 times in Italy (Spinney 2017, 63). The Spanish flu (like the war) killed mainly people aged 15–44, resulting in a W-shaped mortality distribution, rather than the customary U-shape for previous pandemics. It represented an unparalleled negative labour supply shock (see, for example, Velde 2020, 3–4).

III. HOW EPIDEMIOLOGISTS AND STATISTICIANS REACTED TO THE PANDEMIC

Epidemics of influenza were nothing new in the beginning of the twentieth century. In the nineteenth century alone, “four great pandemics of influenza” had occurred, the last one around 1890 (Oliver 1918, 356). The new pandemic was widely discussed in both medical journals (such as The Lancet and the British Medical Journal) and general science journals (such as Science and Scientific American). Among the topics of debate were the severity of the pandemic in comparison to previous ones (for example, Soper 1918) and the origin of the disease, with some suspecting a virus rather than Pfeiffer's bacillus was the cause (see, for example, Oliver 1919).

By the beginning of 1919, epidemiologists were alarmed by the scale and the seriousness of the pandemic.

The pandemic of influenza which swept over the world in 1918 was the most severe outbreak of this disease which has ever been known, and it takes an unpleasantly high rank in the roster of epidemics generally. (Pearl 1919, 1743)

With these words Raymond Pearl, the biostatistician, opened the first instalment of his “Influenza Studies”, published in August 1919. For Pearl, this was an urgent call for scientific research:

If every epidemiologist does not take advantage of the present opportunity to investigate with all possible thoroughness epidemic influenza, to the end of making a better defense next time, he will have been derelict in his plain duty. (Pearl 1919, 1744)

Following his own advice, he studied influenza mortality in forty major cities in the US. Using data on excess mortality, he constructed five “epidemicity indices” for measuring “the force of the epidemic explosion in any particular place” (Pearl 1919, 1767). He then proceeded to a multiple
correlation analysis in which he explored the connections between one of these indices, which he called the “peak-time ratio”, and variables such as population density, geographical location, and age distribution of the population. Since not much came out of the analysis, he considered mortality rates just prior to the pandemic. In this case, he found significantly positive correlations between the peak-time ratio on the one hand, and mortality from pulmonary tuberculosis, heart diseases, and kidney failures on the other. In the second, third, and fourth instalments of his “Influenza Studies” (Pearl 1921a), he further refined the analysis, partly in response to criticism he had received. A survey of the results obtained by Pearl and others can be found in the monograph by Warren Vaughan (1921).

Among those who were critical of Pearl’s approach were Wade Frost and Edgar Sydenstricker, respectively surgeon and statistician at the United States Public Health Service. Their primary concern was to get good data. Aware of the fact that the available influenza statistics were of poor quality (Sydenstricker 1918), they organized special surveys in order to obtain more accurate data. In March 1919, they reported preliminary results from surveys made in Maryland (Frost and Sydenstricker 1919a) and, in June, they reviewed the evidence from other countries (Frost and Sydenstricker 1919b). Later that year, Frost used the data of their surveys to compare the 1918 pandemic to previous epidemics of influenza and pneumonia, and arrived at the conclusion:

In general, this epidemic has been quite similar to that of 1889–1890 in its early development, first in mild, scattered outbreaks, later in a severe world-wide epidemic; in the rapidity of its spread, and in its high case incidence. It has been notably different in a much higher frequency of pneumonia and consequently much higher mortality, especially among young adults. (Frost 1919, 318)

Frost concluded his research on influenza by publishing a summary of the main results of the surveys (Frost 1920), while Sydenstricker used the data to estimate the trend of case fatality during the epidemic (Sydenstricker 1921).

After that, both Frost and Sydenstricker moved on to other topics. Remarkably, ten years later, Sydenstricker revisited the data they had collected to examine whether there was any truth in the popular belief that “the flu hit the rich and the poor alike” (Sydenstricker 1931, 154). By means of a meticulous analysis, he arrived at the conclusion that this
was not entirely correct. The empirical evidence pointed in the direction of a clear social gradient, with incidence and mortality higher among the poorer classes of society, even after correcting for differences in age, sex, and “colour”. This paper stands out as an early example of a careful study of socioeconomic inequality of health.

Physicians did not know what caused the Spanish flu and, therefore, did not know how to treat it properly. As pointed out by Tognotti (2003), over-confidence and Pasteur’s revolution (the idea that every infectious disease was caused by a bacterium), led the international scientific community to mistakenly accept the German bacteriologist Rudolf Pfeiffer’s 1892 claim that he had identified the pathogenic influenza agent in a bacterium. It took some time for scientists to admit that the Spanish flu originated from a virus, not a bacterium. The collapse of the ‘Pfeiffer doctrine’ was accompanied by a crisis suffered by bacteriology in the autumn of 1918, around the same time the disease raged worldwide.

IV. THE SILENCE OF THE ECONOMISTS
In contrast to epidemiologists and statisticians, economists remained virtually silent. None of the major economics journals published an article on the pandemic in the period 1918–1921. Indeed, whereas the modern literature on the economic impacts of the 1918–1920 flu refers often to classic works about the flu by epidemiologists and statisticians, such as Pearl and Sydenstricker (see, for example, Basco, Domènech, and Rosés 2021; Beach, Clay, and Saavedra, forthcoming; Velde 2020), references to contemporary economists are conspicuous by their absence.

A striking example of the lack of attention by economists to the flu pandemic at the time is provided by the American Economic Association and its journal, The American Economic Review. From the 31st Annual Meeting (held in December 1918) to the 34th Annual Meeting (held in December 1921), not a single paper was devoted to the pandemic or to health issues, according to the programmes in the ‘Papers and Proceedings’ supplements of The American Economic Review published in the month of March of the following year. In that period, just one article of The American Economic Review referred to the flu pandemic—even so, only metaphorically. The context was the taxation of war profits. Carl C. Plehn noted that the practice of taxing war profits had spread rapidly to many countries after it had been introduced in Denmark and Sweden in 1915: “Like the Spanish influenza it speedily infected all the belligerent
countries on both sides of the fighting lines and also most neutral countries” (Plehn 1920, 285). Such metaphorical usage of epidemic terms by economists may be traced to the nineteenth-century literature on financial and economic crises, which often deployed “epidemic” and “contagion” to describe them—see, for example, Longfield’s (1840, 222) description that “the demand for gold for hoarding […] is like an epidemic […] like the plague or any other infectious disease which may cease of itself” (see also Besomi 2009, 44, where other instances may be found). It was only much later that economists went beyond the metaphorical stage and started applying epidemiological models to the study of economic movements (see Shiller and Pound 1989).

A similar pattern can be observed in the Journal of Political Economy. The pandemic is mentioned, but only sporadically and incidentally; there is no in-depth analysis of the immediate impact of the shock and its effects. In a section on ‘Health and sanitation’ of an article on the shipbuilding industry it is reported that:

Epidemics of smallpox and typhoid fever were successfully handled in seven localities […]. Special aid was rendered during the influenza epidemic, and, where the scourge threatened serious curtailment of shipbuilding, temporary hospitals were erected. (Douglas and Wolfe 1919, 380)

Likewise, in an article on industrial training in the war period, the discrepancy between the actual and planned numbers of trained men “is attributed largely to a shortage of men in October [1918], and to delays occasioned by the epidemic of influenza” (Wolfe 1919, 741). There is one further brief mention of the epidemic in an article on the street-railway system in Seattle (Douglas 1921, 461). In The Review of Economics and Statistics, the situation is more or less the same (with the partial exception of Persons’ 1923 work on trade indexes discussed below). Like Douglas and Wolfe (1919), Berridge (1920, 185, 188) briefly mentioned the influenza epidemic’s effects on labour supply, this time in connection with gold mining industries in South Africa and Rhodesia.

In American journals with a wider economic focus, such as the thematic issues of The Annals of the American Academy of Political and Social Science, one finds a few passing references to the pandemic and its deadly effects. In a fierce attack on communism, Thomas R. Marshall (1919, 199–200) referred to what he saw as the strong but temporary effects of the flu pandemic: “Bolshevism may come the world over, but
it will be like the influenza—it may kill its millions, but sooner or later it will pass away”. Preston Clark (1919, 46) mentioned the fight against typhus and influenza as examples of fruitful cooperation between Americans and Mexicans. In the same issue of the journal, John Caskie (1919, 189) praised the work that had been done in the city of Philadelphia in order to stop the influenza epidemic, “that dreadful scourge”. L. Wallace (1921, 41–42) drew attention to the substantial economic loss (“industrial waste”) entailed by “subnormal standards of health and vigour”, with influenza being one of the prominent causes. Wilhelm Winkler (1921, 5) presented calculations of the effect of the influenza pandemic on the civilian population of Austria, noting that it “had easy play with the population which had been weakened through lack of proper nourishment”.

In European economic journals, the situation was largely similar. The Revue d’Économie Politique reported on Jean Bourdon’s research on the population of France in 1918, which highlighted the brutal increase of the mortality rate in the second half of the year as a result of the “grippe” epidemic. According to Bourdon, mortality was higher than it had ever been in the forty years before (Revue d’Économie Politique 1919b, 815–816). The journal also noted that the epidemic had put a heavy burden on the population of Germany (Revue d’Économie Politique 1919a, 127). In the Jahrbücher für Nationalökonomie und Statistik, Ludwig Elster looked more deeply into the available statistics on the evolution of the German population during the war period. He observed that the high death toll of the epidemic in the second half of 1918 among the civilian population had been largely disregarded, and that it could be attributed, to a certain extent, to the weakened health of the population as a result of food deficiencies, brought about by the blockade during the war (Elster 1919, 155–156). Using data for the city of Berlin, Hans Guradze (1921, 531) found that women were hit harder than men by the food deficiencies and economic difficulties, which explained why they were more susceptible to get influenza and tuberculosis. In another article, H. Fehlinger (1921, 534) saw influenza as the main cause for South Africa’s exceptionally high mortality rate in 1918. Still in the same journal, E. Mittermüller (1921, 7) pointed out that increased mortality and morbidity due to the war and the influenza epidemic had unfavourable effects on the German life insurance sector.

The connection between the influenza epidemic and insurance issues was also made in British and American actuarial journals. Arthur
Hunter (1919, 264) drew attention to the exceptionally high mortality rates in the US in the last three months of 1918. As far as Britain was concerned, while Lewis Orr (1921–1924, 53–54) acknowledged that the mortality rate had been high in 1918, he observed it had been much lower in 1919 and 1920; therefore, it seemed that the secular decline of mortality rates continued. Actuaries Frankel and Dublin (1919a, 1919b) discussed in detail the impact of the flu pandemic on workers’ mortality rates and the American insurance business. The periodical Economic World also featured an article about the flu pandemic and the insurance market (Marsh 1918).

If we extend our scope beyond economic journals, the picture remains unchanged. In the period 1918–1921, no economist published a book dealing with the pandemic, or mentioned the topic prominently as part of economic books. Startling as it may seem, John Maynard Keynes did not mention influenza in The Economic Consequences of the Peace (1919), even though he had contracted the disease while he was in Paris in January 1919 (Harrod 1951, 234). Keynes (1919, 250n1) did, however, refer to fragile health conditions caused by malnutrition and the spread of tuberculosis in Central Europe during the war, which was part of his argument about the inability of Germany, Austria, and other countries to pay for war reparations. Another Cambridge economist, Arthur Cecil Pigou, did not refer to the Spanish flu either in his classic The Economics of Welfare. Pigou (1920, 872, 943) deployed the term “epidemic”, but only as a metaphor (“epidemics of optimism”) or as shock of “misfortune” which insurance businesses and other companies should be protected against.

A few years later the well-known Italian economist and demographer Giorgio Mortara (1925) would refer extensively to the effects of the flu in his detailed account of mortality patterns in Italy during and shortly after the war. The links between epidemics and demography went back to Thomas Robert Malthus’ famous An Essay on the Principle of Population. Malthus ([1798] 1826, Book II, Chapter XII) listed epidemics, together with wars and famines, as main examples of “positive checks” on population growth. In his careful historical-statistical investigation, Malthus treated recurring epidemic episodes as endogenous to the conditions of living. The history of epidemics, he claimed, showed that the “lower classes of people, whose food was poor and insufficient, and who lived crowded together in small and dirty houses” (Book IV, 258–259), were the main victims. He asks, in “what other manner can Nature point
out to us that”—if population increases too fast in relation to means of subsistence—“we have offended against one of her laws?” (Book IV, 259). The bigger impact of the Spanish flu on the poor sections of society seemed to confirm some of Malthus’ claims. However, Pearl and others denied both the war and the flu had permanent effects on population trends. According to Pearl (1921b, 121), those who saw in war and pestilence any solution to the population problem, “as postulated by Malthus”, were “optimistic indeed”.

The Swedish neo-Malthusian economist, Knut Wicksell ([1910] 1926, 12), noticed how the 1918–1920 flu caused an increase in mortality in Sweden, but regarded it as a detour from the overall trend of reduction in both mortality and fertility numbers in that country. Moreover, there is no evidence that Wicksell (or other prominent Swedish economists) ever discussed the economic impact of the 1918 flu, which brought about a marked contraction of economic activity and poverty increase in neutral Sweden (see The Economist 1918c; Karlsson, Nilsson, and Pichler 2014). Wicksell’s ([1919] 1978) economic view of the World War discussed demographic and monetary factors, with no mention of the flu. Economists remained largely silent about pandemics in the next couple of decades, as illustrated by the influential International Encyclopaedia of the Social Sciences, published between 1930 and 1935, which featured many contributions by economists. It included an entry on “Epidemics”, written by epidemiologist Clifford A. Gill (1934), with some discussion of the Spanish influenza pandemic. Gill did not mention any economic references, though. Similarly, in the entries of the Handwörterbuch der Staatswissenschaften on “Public hygiene” and “Infectious diseases”, both written by public health specialist Alfons Fischer (1923a, 1923b), no mention is made of publications by economists. The Palgrave’s Dictionary of Political Economy, edited by Henry Higgs in 1923–1926 as a new edition of the 1894–1899 original version put together by Inglis Palgrave, did not include any entries on epidemics or the Spanish flu.

V. Business Cycle Research
While the Spanish flu pandemic was by and large ignored by economists, surely it cannot be that those who were monitoring business cycles did not notice the impact it had on the economy? As the pandemic unfolded, contemporary reports from agencies such as the Federal Reserve Bank drew attention to its adverse effects. Especially the November 1918 reports of the twelve Federal Reserve districts highlighted the eco-
conomic turmoil caused by the pandemic. In the district of Boston, “[t]he epidemic of influenza which has prevailed during the past month has seriously interfered with business” (Federal Reserve Board 1918, 1126). Likewise, in the districts of New York, Philadelphia, Richmond, Alabama, St. Louis, and Dallas, the pandemic, through its negative impact on labour supply, significantly troubled business and trade, while in the districts of Cleveland and Atlanta the pandemic was just “a slight disturbing element” (Federal Reserve Board 1918, 1131). However, by the end of the year, the district reports referred to the resumption of “normality” in both health and economic conditions, combined with the armistice in November that year.

Hence, the pandemic-induced recession was sharp but short-lived. It could only be captured by high-frequency data, not by annual statistics (Velde 2020; Beach, Clay, and Saavedra, forthcoming). As a result, business cycle experts did not generally refer to the economic effects of the flu. Indeed, Wesley C. Mitchell (1927) did not mention the Autumn 1918 recession of the pandemic in his detailed narrative of economic fluctuations in the US and European countries. In his later book with Arthur F. Burns, there is mention of the “contraction of 1918–19”, but, in view of its “exceptional brevity and moderate amplitude”, its “failure to register in annual summaries is not surprising” (Burns and Mitchell 1946, 109).

Shortly after the flu pandemic, Warren Persons provided a rare mention by a business cycle expert of the fact that “trade was adversely affected in the autumn of 1918 by a severe influenza epidemic” (1923, 72). Persons mentioned that as an example of “numerous irregular fluctuations” explainable by exogenous “contemporaneous events” that did not belong to the theory or measurement of business cycles conceived as the recurrence of regular economic fluctuations. Epidemics and other “minor irregular fluctuations” were not supposed to interfere with “major movements” formed by the “ebb and flow of industrial activity” (Persons 1923, 71). In particular, Persons’ monthly “trade index”, he claimed, indicated that the period 1916–1918 featured high economic activity, despite the downfall in the last quarter of 1918 associated to the flu as captured by the index. Mitchell (1927, 249ff.) further elaborated on Persons’ notion of “irregular fluctuations”, by treating them as “random events”, or shocks (a term Mitchell did not use), which he distinguished from seasonal, cyclical, or trend movements. Mitchell listed “epidemics”, together with strikes, transport congestions, inventions, and natural phenomena in general, as instances of irregular fluctuations.
that posed a problem for the statistical treatment of economic fluctuations.

One might say that the pandemic lurks in the background of the path-breaking study on the social effects of business cycles published by William F. Ogburn and Dorothy S. Thomas in 1922. Using data for the US, they found that 1918 was a bit of an outlier; for instance, the mortality rate of that year was exceptionally high. They attributed this to “the extraordinary conditions of war time” (Ogburn and Thomas 1922, 331), and decided to calculate coefficients of correlation both with and without the year 1918. They did not mention influenza in their paper, but admitted that “climate, health education campaigns, developments of preventive medicine, and epidemics” (Ogburn and Thomas 1922, 338) could affect the relation between mortality and the business cycle. The study was included in the book published a few years later by Thomas ([1925] 1927) and is now considered a seminal contribution to the literature on “economic epidemiology” (Tapia Granados 2015, 1488).

VI. Why the Silence?

If it is true that economists more or less neglected the Spanish flu pandemic, the question is, why? In this section, we speculate about possible explanations.

Could it be that economists were forced or strongly advised not to pay attention to the influenza epidemic? It is well known that during (and also after) the Great War, authorities introduced a wide variety of censorship measures (Demm 2017). Newspapers and personal correspondence were controlled in order to avoid that too many details about military operations (for example, heavy losses on the battlefield) become public knowledge and to keep the spirits of the population high. The American President Woodrow Wilson never once publicly acknowledged the outbreak, fearing it would harm morale. Accordingly, the federal government’s response to the pandemic was essentially non-existent (see Barry 2004). The same applied to other belligerent countries. It is therefore possible that censorship or self-censorship caused economists to be reluctant to attend to the devastating pandemic (apart from the fact that it was not an object of central government policy). Yet medical researchers and epidemiologists apparently did not have the same scruples. Magazines like The Economist did refer on occasion to the pandemic, but that was “hidden” in parts not subject to censorship, with passing references to the pandemic’s effects on business conditions (see, for
example, *The Economist* 1918a). Numbers about the pandemic were not disclosed by that magazine until March 1919, when it was reported that, for the first time since the start of the records in the 1830s, the total civilian population in England and Wales had come down in 1918, largely due to the nearly 100,000 deaths caused by the pandemic in the last quarter of that year (*The Economist* 1919).

Perhaps a more convincing explanation can be found by looking into the topics that economists did pay attention to, for instance by considering the topics discussed at the annual meetings of the *American Economic Association*. Economists certainly did not shy away from studying urgent contemporary problems: ‘War and reconstruction’ was the general theme of the 30th Annual Meeting held in December 1917. It must be said, however, that the focus was on traditional economic issues, such as taxation, international trade, and agriculture. Government policy was also discussed, but health issues received no special mention and very little suggests that the pandemic changed economists’ attitude towards health. There is at least one notable exception: Irving Fisher’s Presidential Address “Economists in Public Service”, delivered in December 1918. Fisher pointed out that since the foundation of the association a tension existed in its ranks “between those economists who were conservative and those who were radical in regard to applying academic study” (Fisher 1919, 6). He wanted economists to do their “bit” and argued they should serve “all humanity throughout the world and throughout future generations” (Fisher 1919, 21). He was convinced that important lessons could be drawn from the war experience. Without mentioning the pandemic explicitly, one of the reforms he advocated related to health:

> The great field of social insurance for workingmen and especially the next step—Health Insurance—should also engage our attention. Here, likewise, we must steer clear of the bias of the employer, the trade union, the insurance company, or the medical profession. (Fisher 1919, 19)

In a certain sense, Fisher called upon his fellow economists to venture beyond the traditional boundaries of their discipline. However, at the time, there was no such thing as ‘health economics’, and economists had no specific expertise with regard to epidemics or health issues in general. Interestingly, at the height of the pandemic, *The Economist* drew attention to the enormous economic costs of bad health in general (in-
cluding a brief mention of the Spanish flu) in an article aptly called “The economics of health”. It claimed that it was in the interest of Great Britain to change its health policy: “It is incontestable that a thorough reform of national health will require a large expenditure; but expenditure upon it, if well and judiciously made, will pay as handsomely as a business proposition” (The Economist 1918b, 833). That marked a significant change in that magazine’s previous editorial line against government intervention (see The Economist 2020a). In many countries the pandemic effectively led to the creation or restructuration of national public health systems (Johnson 2006, 196). In Great Britain, the Ministry of Health replaced the local government boards. It commissioned a Report on the Pandemic of Influenza, 1918–19 (Ministry of Health 1920), which provided a first comprehensive overview of the pandemic in Britain and the rest of the world. What is striking in that report, however, is the lack of input by economists and the absence of an in-depth discussion of the economic effects of the Spanish flu. It is as if economics had nothing to contribute to the debate.

It was not that economists were not generally aware of the Spanish flu pandemic. However, apparently it was regarded as something outside the field of economics properly. Again, Irving Fisher provides a perfect illustration of that. Fisher had been concerned with health issues and campaigned for the improvement of health conditions for a long time. He criticized economists for focusing “exclusively on physical phenomena” and overlooking that the “true ‘wealth of nations’ is the health of its individuals” (Fisher 1906, 176; see also Nordhaus 2005). Shortly before the most acute phase of the pandemic, Fisher (1918) wrote an article about “Health and War”. His main contribution to the study of health improvement was his 1915 book, jointly written with Dr Eugene Fisk, which stressed rules of individual hygiene and life style from a broad hygiene-medical perspective. The 1919 edition included a section on “Cause and Treatment of Spanish Influenza” (Fisher and Fisk 1919, 375–376). Nevertheless, Fisher refrained from discussing the economic dimension and implications of the flu pandemic in his economic articles (including his 1918 Presidential Address) and books.

Moreover, in contrast with the Covid-19 outbreak, business lockdowns were not implemented during the 1918–1920 flu. The current trade-off between mandated closures and economic activity, which has loomed in the economic debates, was not a topic of discussion back then. Economic activity did come down at the time, but it was caused
mainly by flu-related illness and excess mortality that affected labour supply (see Bodenhorn 2020). Hence, unlike the current crisis, the downfall in economic activity was not reflected in unemployment data. In Sweden, for instance, the unemployment rate of trade union members did not go up in the last two quarters of 1918, despite the contraction of economic activity caused by the flu (see Boianovsky 1998, 231). Neither did economists (especially British) pay any attention to the extensive famine and economic effects of the flu in India (briefly mentioned in the 1920 Report of the Ministry of Health). Then a British colony, India’s death toll was higher than all European nations put together (see Arnold 2019). Other countries, such as Brazil, also suffered from the flu (180,000 deaths). Brazil was far from and not directly involved in the war—although it did declare war to Germany in 1917. The flu outbreak featured prominently in the Brazilian press as well as in debates about workers’ rights and cost of living stabilization, with no economic analytical contributions though (Boianovsky 2021).

Modern economists have managed to distinguish statistically the economic effects of the First War and the flu in some individual country cases and in large samples of countries. The Spanish flu was one of the most important negative global economic shocks in the period 1870–2008, next to the two World Wars and the Great Depression. Increasing mortality associated with World War I and the Spanish flu were estimated to reduce income per capita in the typical country by 8.4% and 6.2%, respectively (Barro, Ursúa, and Weng 2020). However, contemporary economists did not and could not grasp the full extent of the economic impact of flu-induced illness and mortality in 1918–1920.

VII. CONCLUDING REMARKS
As discussed above, up to and including the 1918 Spanish flu outbreak and its aftermath, economists had used the notion of ‘epidemics’ in three ways: (i) as a metaphor, (ii) as an element of demographic analysis, and (iii) as a cause of irregular fluctuations. Although epidemiological models found their way into financial economics (see, for example, Shiller and Pound 1989) and the AIDS epidemic led to an increased awareness of infectious diseases among economists (see, for example, Philipson and Posner 1993), the full integration of pandemic phenomena into economics in general would have to wait until the Covid-19 crisis. This reflects as well sharp differences in the structure of the economic
profession nowadays, with its much larger participation in the political discourse.

Our paper has addressed the differences in economists’ reactions now versus then, with some speculative answers. We did not aim to cover the full range of primary sources; but, even if some sources have been overlooked, it is unlikely that they had left significant traces in contemporary discussions. Our investigation indicates that economists did not pay close attention to the Spanish flu outbreak and its economic impact at the time. This may be explained in part by factors related to the organization of the economic profession and of economists’ self-perception of the scope of their activities (as illustrated by Fisher). Moreover, it has to do with the degree of visibility of the economic features of the pandemic around the end of the war (as illustrated by Persons) and with the absence of nation-wide government policies to fight the pandemic, such as lockdowns. Because of the war censorship and timing, the episode was poorly covered by newspapers, especially in Great Britain and other European belligerent countries, which contributed to the fact that the Spanish Influenza was “largely forgotten” (The Economist 2020a). The decision by governments to “bury the human toil of the disease in the collective memory of World War I” was another contributing factor (The Economist 2020a). This is well illustrated by the Carnegie Endowment’s 208 volumes on the Economic and Social History of the World War, published in the mid-1920s, which devoted only a few pages to the “grippe” and then primarily as a medical or statistical phenomenon (for example, by Mortara 1925).

As put by Spinney (2017, 8), the Spanish flu existed for a long time as little more than a footnote to the massive event represented by World War I. For epidemiologists, however, the Spanish flu served as a call to arms, and it was in the aftermath of the flu pandemic that Kermack and McKendrick (1927) put forward the SIR epidemiological model, extensively deployed by economists and other professionals in the current Covid-19 pandemic. Even though the relation between economists and epidemiologists remains “testy” (The Economist 2020b), it is clear that studying the effects of a pandemic has become a priority for both today.
REFERENCES


**Mauro Boianovsky** is a professor of economics at the Department of Economics at Universidade de Brasilia, Brazil. He holds a PhD in economics from Cambridge University, and has published a number of articles, book chapters, and books in the field of history of economics. He served as president of the History of Economics Society in 2016–2017. Contact e-mail: <mboianovsky@gmail.com>

**Guido Erreygers** is a professor of economics at the Department of Economics at the University of Antwerp, Belgium, and visiting researcher at the Centre for Health Policy at the University of Melbourne, Australia. He holds a PhD in economics from Université Paris X – Nanterre, and has published in both history of economic thought journals and health economics journals. He has also edited and co-edited several books. Contact e-mail: <guido.erreygers@uantwerpen.be>
Three Ways in Which Pandemic Models May Perform a Pandemic

PHILIPPE VAN BASSHUYSEN
Leibniz University Hannover

LUCIE WHITE
Leibniz University Hannover

DONAL KHOSROWI
Leibniz University Hannover

MATHIAS FRISCH
Leibniz University Hannover

Abstract: Models not only represent but may also influence their targets in important ways. While models’ abilities to influence outcomes has been studied in the context of economic models, often under the label ‘performativity’, we argue that this phenomenon also pertains to epidemiological models, such as those used for forecasting the trajectory of the Covid-19 pandemic. After identifying three ways in which a model by the Covid-19 Response Team at Imperial College London (Ferguson et al. 2020) may have influenced scientific advice, policy, and individual responses, we consider the implications of epidemiological models' performatory capacities. We argue, first, that performativity may impair models' ability to successfully predict the course of an epidemic; but second, that it may provide an additional sense in which these models can be successful, namely by changing the course of an epidemic.

Keywords: Covid-19, epidemiological models, performativity, prediction, success, model evaluation

JEL Classification: B40, C63, I12, I18, Z18

AUTHORS’ NOTE: The authors thank three anonymous reviewers for their comprehensive and insightful feedback. Philippe van Basshuysen, Lucie White, and Mathias Frisch gratefully acknowledge the support of the Volkswagen Foundation.
I. INTRODUCTION

Philosophers are increasingly scrutinising epidemiological models that have been used to forecast the trajectory of the Covid-19 pandemic (Horner and Symons 2020; Nguyen and Frigg 2021; Winsberg, Brennan, and Surprenant 2020; Schroeder 2021; van Basshuysen and White 2021). The success of these models is often evaluated in terms of their predictive accuracy (Friedman et al. 2021). We argue that there is an additional dimension along which epidemiological models can be successful, namely in influencing the course of an epidemic. Our argument draws on the concept of ‘performativity’, which economic sociologists and philosophers of economics have used to highlight the fact that economic models may not only represent parts of the economy, but can also influence the behaviours of the economic systems modelled (Callon 1998; MacKenzie 2006; Guala 2007; Mäki 2013; Boldyrev and Ushakov 2016). This notion has a long history in philosophy of science (see Buck 1963), economics, and the social sciences (see Henshel 1993), where it is often referred to as ‘reflexive prediction’ (Buck 1963; Henshel 1993). Predictions (or ‘prophecies’) can be ‘self-fulfilling’ or ‘self-defeating’, understood broadly to mean that the dissemination of the prediction is causally relevant in either bringing about the event predicted or in preventing it from occurring (Buck 1963). We similarly understand the notion of performativity in a broad way; models perform by changing important properties of the phenomena they are representing (see Mäki 2013). Whereas earlier accounts tended to focus on the negative aspects of this phenomenon (particularly the difficulties it poses for accurate prediction), contemporary work on performativity also emphasises potential positive aspects, particularly how an understanding of performativity can allow us to intervene constructively in the economy (Guala 2007).

While there has been sustained discussion of the performativity of economic models, the question of whether epidemiological models might be performative has thus far been neglected. We aim to make a first step

---

1 Ian Hacking (1999) describes a related phenomenon involving social kinds.
2 Mäki himself is critical of attempts to characterize this type of phenomenon in the context of economic modelling in terms of Austin’s notion of performativity, arguing that Austin’s notion is concerned with acts being constituted by certain utterances. By contrast, according to Mäki, economic theorizing may have causal consequences for economic activity without being constitutive of it. We set aside the question of whether Austin’s notion is apt in this context here; rather, we wish simply to draw attention to the fact that, like economic models, epidemiological models can have causal effects on the systems modelled.
3 Some social scientists have described epidemiological models as performative (Rhodes
towards closing this gap by arguing, first, that epidemiological models can be performative in several ways, and second, that this has important implications for how these models should be regarded and evaluated. We proceed as follows: after sketching some important features of epidemiological models, we focus on a model that was used by researchers at Imperial College London (ICL) to forecast the pandemic in the UK and the US (Ferguson et al. 2020). On the basis of this example, and focused particularly on its use in the UK, we identify three ways in which models may be performative: (i) models may influence policy advising; (ii) models may affect policy-makers’ responses to an epidemic; and (iii) models may directly influence individual responses to an epidemic. While (i) and (ii) are well-documented in the case of the ICL model, (iii) is more speculative. Having identified the ways in which epidemiological models can be performative, we elaborate upon the ways in which their performative aspects may complicate the task of successfully forecasting the course of an epidemic. At the same time, however, we contend, their performative abilities provide an additional way in which these models can be successful, namely by changing the course of an epidemic.

II. EPIDEMIOLOGICAL MODELS
Epidemiological models are among the instruments that policy-makers have used (and continue to use) when making policy decisions during the Covid-19 pandemic. These models provide highly idealized representations of how the virus will spread through a population, under a variety of assumptions.

The simplest models are SIR or SEIR models, which assign members of a population to different subgroups, ‘the Susceptible’, ‘Infectious’, ‘Removed/Recovered’, and, in the case of SEIR models, also the ‘Exposed’. SIR and SEIR models consist of several deterministic equations that describe how the proportional sizes of the different subgroups change with time based on transition rates between these groups. These rates are represented by a small number of parameters: the transmission rate, the recovery rate, and the latency period. SIR or SEIR models are macroscopic models that treat populations as a whole and abstract from interactions et al. 2020), but their focus differs from ours, as their main concern is that we should adapt these models to the changing circumstances.

The first two types of ‘performativity’ we outline could apply to any model that is used to inform policy, whereas the third sense is more distinctive and methodologically interesting, and it is on this notion that most theorists focus (see, for example, the articles in MacKenzie, Muniesa, and Siu 2007).
among individuals and transmission events. The model used by the Imperial College London to model the global evolution of the pandemic in different regions of the world is a SEIR model (Walker et al. 2020).

By contrast, the ‘CovidSim’ model discussed in the highly influential Report 9 by the Imperial College Covid-19 Response Team, upon which we will chiefly focus here, is an individual-based model (Ferguson et al. 2020), in which individuals are assigned to different types of location where contacts occur—within the household, at school, in the workplace, and in the wider community. CovidSim models transmission events through contacts among individuals at these locations and movements between locations to derive the evolution of infection numbers. CovidSim is a stochastic model containing a very large number of parameters, many of which are only poorly constrained by empirical data. But a sensitivity analysis (Edeling et al. 2021) has shown that the model is particularly sensitive to variations in three parameters: the length of the latent period in which a patient has no symptoms and is not infectious; the delay to start case isolation; and, the ‘relative spatial contact rate given social distancing’ parameter, which captures the effectiveness of social distancing (this last parameter will be particularly important in what follows).

III. THREE PERFORMATIVE ASPECTS OF THE ICL MODEL

In the following, we will outline three salient ways in which epidemiological models (with a particular focus on the above-mentioned, much-discussed ICL model presented in Report 9) might be thought to have influenced the course of the Covid-19 pandemic.

III.I. The ICL Model Performed Scientific Advising

In a recent working paper, Jonathan Birch (2020) documents how the provision of scientific advice in the UK evolved in the lead up to the introduction of the ICL model used in Report 9 (Ferguson et al. 2020). He suggests that the Scientific Advisory Group for Emergencies (SAGE)—the group chiefly responsible for guiding the UK’s initial pandemic response—dispensed advice in three distinctive forms during March 2020. They began by providing advice in the form of what Birch (2020, 6) refers to as “no unconditional recommendations”: that is, they presented a list of potentially effective means of pursuing a variety of goals that policymakers might have, with no indication of which goals might be desirable or worthy of pursuit. As the situation worsened, by March 9, SAGE shifted to providing advice in the form of “disjunctive unconditional
recommendations” (Birch 2020, 6); endorsing a specific end but leaving open a variety of means for achieving this end. For example, during this period, they outlined a range of non-pharmaceutical interventions that they believed could, in some combination, achieve epidemic mitigation, and recommended that some subset of these measures be implemented (SAGE 2020a). On March 18, shortly after SAGE had begun to discuss the ICL modelling paper, the form of advice shifted again, to a “single unconditional recommendation” (Birch 2020, 6), namely, that school closures were immediately required in order to prevent the NHS from becoming overwhelmed (SAGE 2020b).

Birch traces this trajectory in order to argue that the latter form of advice, while perhaps not appropriate in normal situations, is appropriate in fast-moving crisis situations. But he also comments in passing on a possible reason for this final shift to advice in the form of a single unconditional recommendation. Namely, two days before issuing their single unconditional recommendation (on March 16), the SAGE team discussed the ICL modelling paper for the first time. This paper displayed data about critical care capacity in the same graph as projections about demand for critical care in the absence of strict suppression measures (Figure 1), revealing, in stark form, the extent to which demand was predicted to outstrip supply, even with existing mitigation measures in place. Birch suggests that this might have been the first time that some SAGE members were made aware of the dramatic size of the mismatch. The ICL modellers also included a single unconditional recommendation in the report on the basis of their projections: “Epidemic suppression is the only viable strategy at the current time” (Ferguson et al. 2020, 16). It is possible, then, that this model—particularly through the graphical representation of its results—was instrumental in shaping the form of the scientific advice provided by SAGE.

III.II. The Model Performed Policy-Makers’ Responses to the Pandemic
It is clear that the ICL model had a significant impact on public policy in the UK; on March 16, immediately following the release of Report 9, the UK drastically altered (Bosely 2020) its previous, mitigation-based strategy (DHSC 2020), and almost all of the ICL team’s recommended measures were implemented in the following days (see van Basshuysen

---

5 The ways in which graphical representations of models and their results may influence how these models are adopted is an important topic in the performativity literature; see, for instance, Pahl and Sparsam (2016) in the context of the IS-LM model.
This may have further led, however, to some noteworthy interactions between the epidemiological models that provided a basis for policy advice, and the responses of policy-makers. The ICL model and models like it, which project the impact of various public policy measures, are designed to provide policy guidance. Their usefulness depends on their predictive performance. But the implementation of the modellers’ recommendations steers the course of the epidemic away from a model’s reference forecasts. This can lead to claims that the model is ‘too pessimistic’; that is, that its projections about what would happen in the absence of interventions or behaviour changes (but where mitigation measures are implemented) do not match the observed numbers of infected, hospitalized, and dead (see, for example, St. Onge and Campan 2020). We might regard this as a distinctive instance of counterperformativity (see Mackenzie 2006)—the implementation of the recommended policies steers the real-world outcome away from the reference forecast, but notably, in the case of the ICL model, this was the exact effect that the modellers were hoping to achieve.  

---

6 Although this phenomenon has led to criticism of such models for generating inaccurate predictions, and comparison of reference forecasts to actual outcomes is used to evaluate the predictive accuracy of these models (see, for example, Friedman et al. 2021), it should be emphasised that the projection here is contingent on no measures being implemented (that is, this is a conditional prediction). It is helpful here to distinguish...
Joseph Friedman et al. (2021) draw attention to another problem in their evaluation of the predictive performance of various international Covid-19 mortality forecasting models. They suggest that transmission-based epidemiological models tend to overestimate transmission in their reference forecasts, because prolonged exponential growth in transmission rarely continues unabated—governments (and/or individuals, see below) will change their behaviour in response to deteriorating circumstances—and, we contend, in response to updated modelling projections. The accuracy of modelling predictions and projections could be improved, they argue, by taking such endogenous factors into account (which, as they note, are often explicitly included in economic analyses). In addition, they suggest that models (such as the ICL models) that aim to predict the impact of various public policy measures would be better evaluated by looking at the error in the difference between the reference forecast and the forecast for the implemented policy. Adequately modelling endogenous responses may be complicated, however, by the third performative aspect of epidemiological models discussed below. Moreover, as Friedman et al. (2021) concede, the relevant counterfactuals can be difficult to assess. Even when comparing the projection concerning the implemented policy to its actual effects, the effects of the policy must be somehow distinguished from other factors. It should, however, be noted that economists have made progress in recognising and incorporating endogeneity into Covid-19 modelling, which, among other things, can assist in clarifying the role of government policy in disease dynamics (see Avery et al. 2020).

The belief that models were inaccurate and overly pessimistic might have also had potential further follow-on effects for the implementation of the policy advice on the basis of further models. It is possible that the scepticism and controversy surrounding the ICL projections was one of the factors that contributed to a reluctance to implement a further, short term, ‘circuit-breaker’ lockdown when a second wave was unfolding in the predictions, which generate statements about future states of the actual world, from projections, which generate statements about counterfactual outcomes, conditional on some set of counterfactual assumptions. The performative effect of the model consists in influencing which member of a set of projections ends up most closely resembling the trajectory of the epidemic in the actual world and does not compromise its predictive accuracy. While the model steers the trajectory of the epidemic further away from the trajectory in the reference projection, this does not undermine the predictive accuracy of the model. Thus, this does not constitute the type of ‘self-defeating prediction’ that is the focus of much attention in the performativity literature, where, for example, the prediction that an imminent disaster will happen prevents this disaster from happening and thus undermines the accuracy of the prediction (see, for example, Henshel 1993).
UK in autumn 2020 (see Sayers 2020), which SAGE called for with increasing urgency (Stewart and Sample 2020) on the basis of mathematical modelling projections (SAGE 2020c). The refusal to implement this advice constituted a significant break between scientific advisors and policy-makers in the UK (Stewart and Sample 2020).

### III.III. The Models May Have Performed Individual Responses to the Pandemic

While epidemiological models have influenced individual behaviour through scientific advising and policy-making, they may have also been performative in a more direct way, by changing individual behaviour through information conveyed to the public. Individual behaviour during the pandemic was and is not simply determined by social-distancing and other public policy measures. Rather, how and whether social-distancing measures are taken up by individuals depends, to a significant extent, both on their willingness to comply with these measures and on how they interpret these measures.

Obviously, the effectiveness of public policy measures depends on the public's degree of compliance. For example, whether a policy restricting all social gatherings to the members of one household plus one additional person (a restriction in place in Germany through most of the pandemic) can contribute to a reduction in the reproduction rate $R_0$ of the virus depends on the extent to which people abide by this restriction. But the success of many public policy measures goes beyond the question of mere compliance. Many measures, such as many types of restrictions on social interactions, are compatible with a wide range of different social networks, including some quite extensive networks, and with a wide range of different types of spatial contacts among individuals. How such measures contribute to the slowing of the spread of the virus thus partially depends on how individuals interpret a given policy and how they implement it in adjusting their interactions. There are many different ways, for example, of following the ‘one-household-plus-one-additional-person’ rule: I might meet with a large number of different friends individually night after night, who all have their own largely non-overlapping group of friends, with whom they also meet; or I might radically cut down on my social contacts and form an isolated ‘pod’ with just one other person. Both behaviours comply with the rule, but the resulting social networks end up looking quite different. This is also the case in situations where explicit policy guidance is lacking—for example, the number of
infections that result from interactions in the workplace (which is one of the types of interaction explicitly modelled in the ICL model) will depend to some extent on how individuals adjust their behaviour in response to the model’s predictions.

But to what extent individuals are willing to comply with policy measures, and how measures are interpreted and implemented in individual behaviour, depends on how serious individuals take the threat presented by the pandemic to be. This in turn depends, at least partly, on the information the public receives concerning the pandemic—information that often prominently includes model projections in infection, reproduction, and fatality rates (for example, Boseley 2020). For instance, when people are provided with this information through some kind of graphical representation, such as the graph in Figure 1, depicting how the demand for critical care beds could greatly exceed capacity, this might make the urgency of the situation particularly salient, and it might lead to different reactions than if this information had been conveyed through other means.

Thus, the predictions of epidemiological models and how these are understood by the public can play an important causal role in changing properties of the pandemic by influencing individual responses to the pandemic—a connection that is further underscored by the empirically well-supported assumption that people in many places reduced their contacts in advance of the implementation of social distancing policies (for example, Friedson et al. 2020; Sears et al. 2020). That is, our modelling efforts may interact with the very quantities and properties we are trying to model and can thereby change them: epidemiological models and the quantities modelled by them exhibit what Ian Hacking has called a “looping effect” (1999, 105).

For example, CovidSim contains a parameter representing the ‘relative spatial contact rate given social distancing’—a parameter representing the effectiveness of social distancing measures, which, as Edeling et al. (2021) have shown, is particularly influential in determining the model’s projections. To what extent the parameter value chosen provides an adequate representation of the effect of social distancing measures depends on how the measures are implemented in individuals’ behaviour, which may be influenced by the model’s projections.

Simpler models might not distinguish the institution of a policy from compliance with it, and simply model how universal compliance would affect infection rates. Yet in this case, too, the extent to which such a
model can provide adequate predictions concerning the effect of policy measures will depend on the degree to which the public abides with these measures. And the effectiveness of social distancing measures depends on how the model’s results are communicated to and understood by the public.

IV. PERFORMATIVITY IN MODEL EVALUATION AND APPRAISAL

With these three potentially performative aspects of epidemiological models established, we are now in a position to draw out two ways in which recognizing performativity can assist us in the evaluation of these models. We consider, first, how performative aspects may compound the difficult task of successfully forecasting the course of an epidemic and, second, how they might be assessed in light of their ability to change the course of an epidemic.

Epidemiological models serve dual purposes: apart from their epistemic purpose of forecasting the course of an epidemic, they also serve the practical purpose of informing and guiding policy-making. These epistemic and practical purposes go hand in hand: on the basis of forecasts, policy-makers can choose policies that are likely to prevent unwanted outcomes. However, as we have seen in the previous section, the performativity of models complicates this task, because the implementation of these policies can steer the course of the epidemic away from the models’ reference forecasts and can influence the adequacy of assumptions concerning the effectiveness of policy measures. If a modelling study does not include a range of scenarios reflecting different combinations of policies (some of which resemble those that will actually be implemented), some of the estimated parameters may turn out to be inaccurate due to the effects of the implemented policies (for example, levels of mobility), and the forecasts derived from the model may thus be inaccurate, too.

To be sure, many of the models that have been used to forecast the course of the Covid-19 pandemic, including the ones we have focused on here (Ferguson et al. 2020; Walker et al. 2020), do include a range of scenarios which project the impact of different (combinations of) policies. This can generate projections that better resemble the actual policy trajectories followed in a country, and models’ predictive accuracy can accordingly be assessed by comparing these projections with observed outcomes (van Basshuysen and White 2021, forthcoming). However, gauging the accuracy of even these more pertinent scenario forecasts can face significant limitations, due to the fact that most models abstract from
endogenous behavioural responses, such as when individuals engage in social distancing even before government-issued directives take effect, adopt more rigorous hygiene and personal protection protocols, cancel travel plans, and so on. Such behavioural responses are poorly understood and difficult to anticipate (Friedman et al. 2021) and are thus difficult to incorporate into models (which is perhaps why extant models make no attempts to include them). Report 9, for instance, forecasts epidemic trajectories conditional on combinations of policies, but not on these types of spontaneous behavioural responses, potentially impairing the accuracy of important parameter values. Furthermore, even sincere efforts at estimating these parameters may be limited, as behavioural responses can change over time (such as when individuals become tired of social distancing directives and depart from recommended behaviours, or when the public perception of the trustworthiness of expert and model projections changes). Even where modellers endeavour to conscientiously estimate crucial empirical parameters, and to re-parameterize models on an iterative basis (as is done for some of the models we discuss), it might not be possible to completely overcome these epistemic difficulties. The performativity of models thus presents a serious challenge to their predictive capacities.

However, while the performative aspects of epidemiological models can hamper their ability to successfully forecast the course of an epidemic, recognizing the performative dimension of these models can allow us to judge whether they have been successful in a second way, namely, by changing the course of an epidemic. That is, performativity may figure as an important criterion for appraisal beyond predictive accuracy.

When models are performative, their dual roles as epistemic instruments (that predict the evolution of a pandemic) and as practical instruments (that inform policy-making through reference and scenario forecasts) can make it difficult to evaluate them with respect to each purpose individually. As we have discussed, the performative effects of these

---

7 It should be noted that not all distortions due to spontaneous behavioural changes are the effect of performativity being in play. A distinction might be drawn between individuals adapting their behaviour in response to model predictions, or in response to other kinds of information, such as current prevalence, which might be gathered and conveyed independently. It can be difficult, however, to determine which mechanisms are in play in a given case, as it might not be clear whether individuals’ sources of information include model results. For our point that performativity may exacerbate the task of forecasting, it suffices that individual behaviour is sometimes a response to model results, which should be uncontroversial, given that these results and their graphical representations have figured prominently in public perception (as outlined above).
models on policy and individual behaviour can drive outcomes away from the models’ predictions. When the predictions of models are undermined, moreover, this seems to undermine their epistemic credentials, making them unsuitable as a basis for public policy or individual behavioural response. When we focus only on the predictive capacities of such models, we seem to be in a bind here—the more they steer policy and behaviour, the less suitable they seem as a basis for steering policy and behaviour (unless they endogenize behavioural response—although, as we have just pointed out, this can be difficult).

One way out of this bind is to move model appraisal beyond focusing on models’ predictive performance alone. We thus suggest that both predictive and performative capabilities should be considered side-by-side when appraising epidemiological models. This could proceed by evaluating some forms of performativity as a desirable attribute (and others as undesirable) in all-things-considered assessments of what models have done for us, epistemically and practically. Following Philip Kitcher (for example, 2011) in understanding and evaluating science through the roles it plays in, and partly for, society (where those roles often extend beyond purely epistemic ones), we suggest that a more comprehensive assessment of what epidemiological models have done for us (and whether they did a good job of it) should consider their performative aspects, too.

To use an analogy, consider a medical doctor telling her patient that he is likely to die of a heart attack within the next decade if he continues smoking, drinking, and doing little exercise. Such a prediction can be performative in helping the patient change his habits, allowing him to escape the unwanted prediction. Likewise, epidemiological models can have the performative ability to change important aspects of the phenomena they seek to capture. While this can adversely affect their predictive abilities, this does not necessarily imply that a model’s suitability, adequacy, or usefulness is diminished, if we consider its performative impact to be a potential virtue. Just as we should not think that a doctor is unqualified because she cannot tell whether a patient would indeed have died of a heart attack counterfactually, or because she cannot accurately forecast how much longer, exactly, a patient would live under a changed exercise regime, we suggest that it is unhelpful to assess the utility of epidemiological models based on their predictive abilities alone. Epidemiological models have been a crucial resource for informing and justifying policy interventions, and may have contributed to shaping both the public’s understanding of the pandemic and their behavioural response to it. The
ability to make such performative contributions can be understood as a desirable feature of these models, even if models' performativity can diminish their predictive prowess.

It is important to stress, however, that not all forms of model performativity are desirable. We might think, for instance, that some performative successes of epidemiological models can also have unwanted side-effects, such as when some members of the public begin to take the difference between observed outcomes (such as cumulative mortality) and widely publicized pessimistic reference scenarios as reason to call the severity or reality of the phenomena being modelled into question. One might also worry that by choosing particular outcome variables as key modelling targets, such as critical care demand or cumulative mortality, models have been performative in making these outcomes relatively more salient than various economic, societal, or psychological endpoints, and that models might thereby have contributed to steering policy and public response in ways that led to suboptimal societal outcomes. We do not take a stance on which performative aspects are desirable or undesirable here, but rather highlight that it is important to take both kinds of factors into account.

So how, exactly, are we to take models' performative aspects into account? Unfortunately, numerous complications arise when detailing the potential role of performativity in model appraisal. Should it play an evaluative role, providing a criterion by which the merits or deficiencies of models can be judged? Or should it be prescriptive, recommending the construction, selection, or use of models with certain performative capacities? The latter view seems highly problematic: we do not think that models can (or should) be made better by being made more performative, even if it were possible to agree that particular forms of performativity are socially desirable. Deceitfully meddling with a model in order to steer policy or behavioural response would raise significant concerns about illegitimate value-influences, and could severely undermine the epistemic credentials of models. Models are, and should remain, epistemic instruments in the first instance, and while some value-influences are rarely avoidable, model construction and use should proceed in epistemically responsible ways (see Elliott and McKaughan 2014; Parker and Winsberg 2018 for discussions of the roles of non-epistemic values in model construction and use).

To allay such concerns, we need principles that allow us to acknowledge models' performative aspects, good and bad, while
maintaining a commitment to the idea that models’ epistemic functioning should remain untouched by considerations about their performativity. Unfortunately, formulating such principles, and providing a detailed account of how exactly we should weigh or integrate models’ epistemic and performative aspects, involves various non-trivial questions that cannot be addressed here. These include: On which outcomes should we focus when gauging models’ performative effects? How should we weigh considerations about models’ epistemic and performative aspects? Should epistemic considerations enjoy lexical priority, or are there cases where performative virtues may compensate for epistemic shortcomings? Should all performative aspects be considered in model appraisal, or are there cases where we should distinguish between the model-as-such, the practice of model-use, and issues of science communication about model outputs (where some evaluations of performativity should perhaps pertain only to the latter two but not the former)? We suggest, for now, only the following provisional safeguard: performativity should only play an evaluative, but not a prescriptive, role in model appraisal. It may figure in judgments concerning whether the downstream performative aspects of models have been desirable or undesirable, but such judgments should not bear on decisions made at the stage of model construction, selection, or deployment—there should be no wishful modelling. Navigating the additional intricacies involved in formulating fine-grained principles is beyond the scope of this paper and must be addressed in future work.

Finally, we would like to highlight a rather different (and less problematic) route to taking models’ performativity into account at the stage of model construction: instead of aiming to build models with specific performative goals in mind, we can aim to build models that endogenize their own performativity. Social scientists, as early as the 1950s, have made efforts to develop analytic machinery that can help anticipate the performative effects of public predictions (Grunberg and Modigliani 1954; Simon 1954). Despite significant difficulties, there has been progress with attempts to incorporate endogeneity into epidemiological models (see, for example, Avery et al 2020; Eksin, Paarporn, and Weitz 2019). Continuing these efforts could allow us not only to improve models’ predictive capabilities, but also to answer criticisms that social scientists have been insufficiently involved in scientific advisory systems that inform policy. Social scientists could make an important contribution to the predictive endeavours undertaken by epidemiologists by helping
account for and anticipate the intricacies of the social and behavioural underpinnings that govern models’ performativity.

V. CONCLUSIONS
We have argued that epidemiological models may have altered the course of the Covid-19 pandemic in three important respects: (i) through shaping the form of scientific policy advice, (ii) through shaping policy formation, and (iii) through influencing individual behavioural responses to the pandemic. Existing philosophical literature on epidemiological models and extant model evaluation studies have, however, focused primarily on predictive accuracy, and these performative aspects—and their implications—have been largely neglected. We have proposed that it may be important to take the performativity of models into consideration. This is, first, because it can exacerbate the difficulties involved in accurately predicting the course of a pandemic, and second, because certain types of performativity might be regarded as a desirable feature of epidemiological models. It thus might be fruitful to consider, in addition to predictive accuracy, the ways in which we can assess the success of models based partially on their performative aspects, and the role these might play in our fight against the pandemic.

REFERENCES


**Philippe van Basshuysen** is a postdoctoral researcher at the Leibniz University Hannover and research associate at the Centre for the Philosophy of Natural and Social Science at the London School of Economics. His research is in the philosophy of economics and public policy, with a focus on public health.
Contact e-mail: <philippe.van.basshuysen@philos.uni-hannover.de>

**Lucie White** is a postdoctoral researcher at the Leibniz University Hannover, specializing in bioethics and the ethics of technology. She is currently working on a couple of projects which focus, in various ways, on policy responses to Covid-19.
Contact e-mail: <lucie.white@philos.uni-hannover.de>

**Donal Khosrowi** is a postdoctoral researcher at Leibniz University Hannover. His research interests and recent publications focus on causal inference in social science (especially in economics and evidence-based policy), scientific representation with models, and values in science.
Contact e-mail: <donal.khosrowi@philos.uni-hannover.de>

**Mathias Frisch** is a professor of philosophy at Leibniz University Hannover. His research interests lie in causation, climate modelling, and decision making under uncertainty.
Contact e-mail: <mathias.frisch@philos.uni-hannover.de>
Uncertain Policy Decisions During the Covid-19 Pandemic

MALVINA ONGARO
Università del Piemonte Orientale

Abstract: The Covid-19 pandemic has shaken the world. It has presented us with a series of new challenges, but the policy response may be difficult due to the severe uncertainty of our circumstances. While pressure to take timely action may push towards less inclusive decision procedures, in this paper I argue that precisely our current uncertainty provides reasons to include stakeholders in collective decision-making. Decision-making during the pandemic faces uncertainty that goes beyond the standard, probabilistic one of Bayesian decision theory. Agents may be uncertain not just about factual properties of the world, but also about how to model their decision problems and about the values of the possible consequences of their options. As different stakeholders may have irreconcilable disagreement about how to resolve these uncertainties, decision-making procedures should take everybody’s perspectives into account. Moreover, those communities that are hit harder by the pandemic are also those that are typically excluded from knowledge production. Thus, in the face of Covid-19 uncertainty, both democratic and epistemic considerations highlight the importance of stakeholders’ inclusion in policy decision-making.

Keywords: uncertainty, policy making, decision-making, inclusion

JEL Classification: D81, D70, A13, B40

I. INTRODUCTION

The Covid-19 pandemic has taken the world aback. It has shaken almost every aspect of our individual and social life, presenting us with severe, novel challenges. These unprecedented circumstances are so fraught with uncertainty that making good decisions may be difficult. And yet, the pandemic requires timely and apt responses. In the face of this uncertainty, how should policy makers make decisions? While the pressure for timely responses may push towards swifter and less inclusive decision procedures, I argue that our current uncertainty provides both democratic and
epistemic reasons for the importance of stakeholders’ inclusion in collective decision-making.

Bayesian decision theory provides important tools to deal with decision-making under uncertainty, but it assumes that the uncertainty faced by the agent can be captured entirely in probabilistic terms. However, our current uncertainty goes beyond this standard view: it concerns also the construction of decision models and the evaluation of the possible consequences of a policy. In both model and evaluative uncertainty, people may have irreconcilable disagreements due to their values and priorities. For this reason, these uncertainties call for the inclusion of all stakeholders in the decision-making processes. In the case of standard uncertainty, considerations of accuracy may point against this conclusion, privileging expertise over inclusion. However, the exclusionary nature of expertise keeps removed from our collective efforts of knowledge production precisely those communities that are more severely hit by the pandemic.

II. STANDARD UNCERTAINTY

When making decisions, policy makers face situations of practical uncertainty, that is, uncertainty concerning what to do (Peter 2021). Bayesian decision theory suggests that the way to resolve this uncertainty is to choose the option from which the agent expects the best outcome, given their beliefs about the world. The value of the different outcomes is represented by the agent’s utility function, while their beliefs about the possibility of the contingencies on which these outcomes depend are captured by a probability function. Then, the agent should pick the available option leading to the highest utility given the probability of different scenarios.

This picture assumes that the agent’s uncertainty can be entirely captured by the probability function representing their beliefs. While very elegant, this assumption is rather strict. By this, I do not mean to criticise the widespread Bayesian tenet that degrees of belief can be represented by probabilities. Rather, I will focus on the uncertainty that is left out of the picture even on the assumption that it is possible to represent degrees of belief with probabilities. Uncertainty can vary in severity and in nature: if it is too severe, or if it is not of the right kind, then it cannot be captured by probabilities.

Representing degrees of belief with probabilities requires, of course, that the agent be able to form degrees of belief. Yet, the severity of the uncertainty may be such that this is not possible. For instance, it may be
natural to represent the degree of belief in the effectiveness of a certain vaccine with the probability rate of its success. But we do not have probability measures of the likelihood that the government will fail, due to its inadequacy, in organising the vaccination campaign: the agent may have some degree of belief in that eventuality, but be little confident of its accuracy. In some more abstract cases, the agent may be unable to form degrees of belief at all, for example with respect to the possibility of the insurgence of a new pandemic or to the impact of restrictive measures on nationalist sentiments. And finally, there are also all the contingencies the agent is unaware of, that is, those things that the agent does not even know that they do not know.

These illustrative cases show that decision-making in response to the pandemic involves uncertainty that is more severe than mere probabilistic risk. While classical expected utility theory may be inadequate to tackle decisions under more severe degrees of uncertainty, numerous sophistications have been developed to tackle these cases while still holding on to the Bayesian tenet (for example, Ghirardato 2001; Bradley 2017; Karni and Vierø 2017). While going beyond probabilistic risk, severe uncertainty can still be addressed within some expanded version of Bayesian decision theory.

However, this is possible only under two conditions: first, that the decision problem is already modelled; second, that the utilities are given. In the next two sections, I will explore each of these conditions in turn, showing how they lead to further uncertainty, and how they point to stakeholder inclusion as a requirement for decision-making during the pandemic.

III. MODEL UNCERTAINTY

Formal theories require that decision problems be structured in models that they can take as input. As Bayesian decision theory calculates the best option on the basis of the utility of the outcomes of the alternative options and of the probability of the contingencies on which these outcomes depend, then it can only be applied to problems that have been structured in a set of available alternative courses of action, a set of outcomes, and a set of contingencies (or states). But real-life problems do not come with pre-formed sets. Thus, the theory can only be applied once the problem has been modelled—yet the theory itself does not provide any instruction on how the model should be constructed. Insofar as agents
can be uncertain about the right way to model their decision, then practical uncertainty cannot be entirely resolved within the boundaries of Bayesian decision theory. And indeed, there are many reasons to expect uncertainty over decision modelling in policy responses to the pandemic.

As the current circumstances are unprecedented in our lifetimes, policy makers are exploring options that are outside of their usual protocol: extraordinary times require out-of-the-ordinary solutions. Restrictive measures that were unheard of in democratic societies are now commonly debated and implemented; creative solutions have been devised to host events without compromising safety; many activities have had to change shape and move online. This means that policy makers work with an open set of acts: rather than sticking to the old guidebook, they may have to come up with novel solutions to novel problems. For this reason, they may be uncertain as to whether they are really considering all the alternatives at their disposal.

Even in the identification of the sets of outcomes and contingencies there is an uncertain aspect. Any given action leads to an indefinite number of consequences: my decision to go to a certain café will have the consequence of me drinking a certain blend of coffee—but also of sitting on a stool, crossing the street, and entering the café with my left foot, among countless other actions. Not all these consequences will be relevant for my choice. The exact number of steps to get to the café is probably not relevant, while the overall distance may be. Some people may care about the comfort of their seat, while others may only care about the quality of the coffee. Clearly the agent should not consider all the minute consequences of an action, but only those relevant to the decision. The same applies to the description of the different possible scenarios: each state of the world is constituted by a myriad of factors, the relevance of which depends on the grain of specification one adopts and on the selection of consequences one cares about. This means that there is no unique way to model a certain decision problem. If this is so, then agents may well be uncertain about their model, and wonder whether they have included the relevant aspects. I will call this model uncertainty, to distinguish it from the standard uncertainty over beliefs that we have seen addressed in Bayesian decision theory.

Model uncertainty is particularly acute under the pandemic, as policies tend to intervene on complex situations made of several interconnected aspects. A policy may impact both public health and the economy, while also having environmental and political consequences. Of course,
one could strive to construct models that are as comprehensive as possible, but at some point a selection is inevitable—and the more complex the situation, the more controversial the selection may be. Moreover, given the severity of our standard uncertainty, it is possible to end up discarding some factors that may turn out to be relevant after all.

**IV. EVALUATIVE UNCERTAINTY**

Once the decision is modelled, the resolution of the agent’s practical uncertainty requires more than just probabilities. As decisions are the result of both the beliefs and the desires of the agent, Bayesian decision theory requires a utility function to represent the agent’s preferences as well. While probabilities (or lack thereof) are themselves representations of uncertainty, utilities are taken to represent value with certainty, as if desires were not subject to uncertainty the way beliefs are. Instead, I will refer to the predicament of the agent who is not certain about the value of consequences as *evaluative uncertainty*, to distinguish it from both standard and model uncertainty.

Evaluative uncertainty is not usually included in Bayesian decision theory. Some authors have proposed extensions with families of utility functions, but their efforts aim at tackling incompleteness of preferences rather than uncertainty (for example, Bradley 2009; Ok, Ortoleva, and Riella 2012; Galaabaatar and Karni 2013). This is not surprising: it is somewhat counter-intuitive to think of uncertain desires, as these seem to be either trivial or reducible to standard uncertainty. Either the agent is unsure about their own tastes, in which case the uncertainty should be easily remedied with some more introspection; or they are uncertain about some factual properties about themselves or about some normative facts, in which case evaluative uncertainty can ultimately be reduced to standard uncertainty over the agent’s beliefs (Bradley and Drechsler 2014). In either case, there is nothing particularly worrying about evaluative uncertainty.

However plausible this picture may be for individual decision-making, the situation is less straightforward for policy making. Even on the assumption that each stakeholder has definite preferences, these are not usually the same for everyone. Thus, there will be a variety of alternative preferences available to the decision-maker, who will be in a situation of evaluative uncertainty.

Once again, the complexity of the situation brought about by the pandemic makes this uncertainty particularly severe. As we have seen, policy
actions tend to impact many and various dimensions of the social world, and their consequences are not necessarily positive in all of them. This means that policy implementation will lead to important trade-offs, and stakeholders that are positioned differently with respect to these trade-offs will evaluate the outcomes of the policy differently. For the sake of illustration, we can imagine that policies protecting public health while halting the economy may be judged more favourably by the elderly, who are more exposed to the risks of the virus while potentially enjoying more secure financial situations, than by the youth, who tend to be less exposed to Covid-19 complications than to professional insecurity.

V. TWO ARGUMENTS FOR STAKEHOLDERS INCLUSION

We have seen three different types of uncertainty that play a role in decision-making. All three types are made particularly severe by the circumstances of the pandemic. This severity is due not only to our ignorance of crucial information regarding the virus, but also to the complexity of our current situation—which is constituted of many interconnected dimensions—and to the presence of a plurality of stakeholders. In this section, I will argue that all three types of uncertainty provide reasons for increasing the inclusion of stakeholders in decision-making. First, I suggest a democratic argument based on model and evaluative uncertainty; then, I propose an epistemic argument based on standard uncertainty.

V.I. A Democratic Argument

Under model uncertainty, the agent is uncertain about the way they have modelled their decision problem, for instance because they are not sure whether they have included all the options at their disposal or because they fear they have overlooked some relevant aspect of the situation. Under evaluative uncertainty, the agent is uncertain about the value to be assigned to the consequences they expect from the performance of the different actions. These states of uncertainty do not concern the agent’s beliefs over some matters of fact and cannot be captured by probability distributions.

This means that standard uncertainty differs from model and evaluative uncertainty at least in one important respect. Insofar as something either is or is not the case, then there is a sense in which there is a correct solution to standard uncertainty. On the other hand, there is no unique solution to model or evaluative uncertainty, as there is no unique way of modeling a decision or of evaluating some consequences.
As for model uncertainty, different agents may disagree on whether something is relevant to the decision or whether some action is really available. Within some limits, this disagreement could be irreconcilable, in the sense that it ultimately boils down to disagreement over values and priorities. The same goes for evaluative uncertainty: people may disagree on how to evaluate some outcome, even more so given the amount of trade-offs required by the complex policies responding to the pandemic. There is no system of values that is the uniquely correct one for evaluating such consequences.

While there is no reason why, on matters of values and priorities, some viewpoints should hold more influence than others, lack of inclusion will typically favour the more powerful. In February 2020, the Italian city of Bergamo was severely hit by the pandemic. The size of the tragedy was partly due to the late adoption of restrictive measures. In those crucial weeks, Bergamo’s influential Confederation of Industry (Confindustria) started a campaign pressuring the city not to stop its activities, and lobbied politicians to avoid lockdowns (Horowitz 2020). In a situation of uncertainty over the value of their alternatives, decision-makers under time pressure adopted the evaluative stance of the most influential stakeholder.

Of course, Confindustria’s position was not wrong: it reflected a system of values. Indeed, evaluative differences may be due to different relations to the matter at stake. As each stakeholder has a different perspective on the decision and on its results, the inclusion of more stakeholders in the decision-making process implies a more comprehensive and diversified perspective on the matter—one that better represents its complexity. For this reason, model and evaluative uncertainty point to the importance of including all relevant stakeholders in policy decision-making.

V.II. An Epistemic Argument

While democratic considerations in the face of model and evaluative uncertainty call for inclusion, in the case of standard uncertainty epistemic considerations seem to push in the opposite direction. If there is a correct answer to standard uncertainty, then we should not treat all possible solutions equally. Our priority should be to find the correct one, and therefore the guiding principle in assessing the alternative ways to solve the uncertainty should be one of accuracy. For this reason, addressing standard uncertainty is usually the role of experts and scientific advisors in
policy making. However, the exclusionary nature of expertise provides an epistemic argument for the inclusion of stakeholders in decision-making.

The attainment of epistemic authority requires long years of education and access to epistemic institutions. However, these two conditions are not open to all individuals equally. Education costs time and money, and people from low-income backgrounds may not have the possibility to devote resources to it. Academic institutions are by no means immune to structural inequalities (Gillborn and Mirza 2000; Davies and Zarifa 2012), and access to quality education is more difficult for minorities and underrepresented groups (Carnevale and Strohl 2013). Thus, epistemic authority is not equally distributed among society, and expertise tends to be exclusionary.

However, this is not just democratically problematic—it is also epistemically problematic. The Covid-19 pandemic does not affect everybody equally: it hits disproportionally ethnic (Millett et al. 2020; Kirby 2020) and sexual (Banerjee and Nair 2020) minorities. These communities are more exposed not only to the disease, but also to the psychological (Pedrosa et al. 2020; Suen, Chan, and Wong 2020) and economic (Hu 2020; Kantamneni 2020; Mo et al. 2020) consequences of the pandemic. If this is so, then their members hold an important epistemic vantage point on the pandemic and its impacts.

By itself, diversity in science and in scientific institutions is an epistemic value (Intemann 2009). Moreover, feminist epistemology has highlighted that knowers are socially located, standing in particular relations to what is known (Anderson 2020). The under-representation of minorities in science leads to the exclusion of people with first-person access to the experience of Covid-19 and of its effects: our collective enterprise of knowledge production with respect to the pandemic lacks the perspective of some of the most hit communities. For this reason, in the face of standard uncertainty, decision-makers should strive to include a variety of stakeholders in their deliberations, and especially those whose epistemic perspective on the pandemic is particularly relevant for policy making, while also being traditionally excluded from institutional knowledge production.

VI. CONCLUSION
The Covid-19 pandemic has shaken our lives. It presents us with a continuous series of novel problems, requiring often drastic solutions. Policy
making in response to this situation may be particularly difficult, especially given the uncertainty under which we have to make decisions. Even though the pressure to produce timely responses to the crisis may lead to swifter and less inclusive decision procedures, I have argued that the inclusion of stakeholders in decision-making processes is important for both democratic and epistemic reasons. This is so because the uncertainty we face is of at least three different types, all of which are made particularly severe by our current circumstances. Agents can be uncertain not only about some factual issues, but also about the best way to model their decision problems, as well as about the value of the possible outcomes of different policy actions. Different stakeholders may have irreconcilable disagreements about these last two types of uncertainty, which underline the importance of including all perspectives in our collective decision-making. But different stakeholders may also hold different standpoints in relation to the pandemic, and their inclusion is therefore also epistemically necessary to resolve our standard uncertainty—especially since the communities that are affected most by the pandemic are also those that are traditionally excluded by our knowledge production institutions.

REFERENCES


Malvina Ongaro is a PhD student in Philosophy at the University of Eastern Piedmont, Italy. She holds a MSc in Linguistics and Cognitive Studies from the University of Siena and one in Philosophy of Science from the London School of Economics. She is currently working on uncertainty, with a focus on decisions and on the role of values in policy making.

Contact e-mail: <malvina.ongaro@uniupo.it>
How to Handle Trade-Offs in Pandemics

KRISTER BYKVIST
Stockholm University and Institute for Futures Studies

Abstract: Pandemics and other similar crises force us to make difficult moral trade-offs. It is tempting to think that this challenge should be met by invoking fundamental moral principles. This is a mistake. Instead, we need to work hard at designing institutions that enable the officeholders to make reasonable decisions under both fundamental ethical disagreement and empirical/evaluative uncertainty. It is argued that this is best done by supplementing the ethical-cum-legal platforms already in use with an ethical framework inspired by social welfare theory.

Keywords: trade-offs, pandemic, Covid-19, ethical disagreement, evaluative uncertainty, social choice framework, domain restrictions

JEL Classification: D63, D71, I38

I. INTRODUCTION
‘Follow the science’ has been a recurring catchphrase during the pandemic. It has been used by both politicians and scientists to justify radical measures to stop the spread of Covid-19. The obvious problem is that science on its own cannot justify measures—it can only tell you what is, will, or might be the case, and not what ought to be the case. It is tempting for us moral philosophers to think that we have exactly the right theoretical background to fill this gap between scientific results and the adoption of measures or strategies. After all, many of us work full time at constructing well thought-out and philosophically sophisticated fundamental ethical principles that will survive the criticisms of our peers in heated seminar discussions. It is therefore easy to think that the main focus should be on finding the most reasonable, universal fundamental principles. Once these have been identified and made public, the decision-makers can then use them to justifiably move from scientific results to decisions in pandemics and other crises.

AUTHOR’S NOTE: For very helpful comments on earlier drafts of this paper, I would like to thank Andrea Asker Svedberg, Orri Stefansson, Tim Campbell, and four anonymous reviewers for this journal.
I shall argue that this ‘filling the gap with a fundamental principle’ approach to pandemics and other similar global crises does not work, especially if the goal is to make us better prepared to deal with the next pandemic or global crisis. Rather, what we need to do is to work hard at designing institutions that enable the officeholders to make reasonable decisions under both fundamental ethical disagreement and empirical/evaluative uncertainty. This work must be a joint endeavour between moral philosophers, legal experts, politicians, and officeholders in agencies, and, crucially, it needs to gain political legitimacy through democratic processes. There is work to do for moral philosophers, however, especially when it comes to suggesting revisions of the various legally binding ethical platforms that are already in place for healthcare interventions in ordinary circumstances. But here we face a dilemma: either the platform rules are general and vague enough to gain broad societal approval, but this will come at the cost of not providing any clear guidance when it comes to difficult trade-offs, or they do provide such guidance but at the cost of lacking broad societal approval. I shall argue that we can avoid the sharpest horns of this dilemma if we supplement the ethical-cum-legal platforms with an ethical framework inspired by social welfare theory. An ethical framework is a system of rules that are in compliance with ethically binding platforms but provide more fine-tuned and precise trade-off principles, often by providing interpretations of the general and vague principles set by the platforms.

II. EXAMPLES OF TRADE-OFFS
It is clear that the Covid-19 pandemic has forced us to make many difficult trade-offs. Strict measures, such as lockdowns with bans on social gatherings and unnecessary shopping, can save many lives if they are put in place at the right time; but this comes at the cost of severe restrictions of freedom for many people whose wellbeing will also be negatively affected, including deaths caused by economic recession. Indeed, the measures will include restrictions of fundamental democratic rights, such as the freedom of assembly and freedom of movement. Since many of the lives saved will be vulnerable, older people, this trade-off is to a large extent one between the wellbeing and freedom of the young and the lives of the very old. But it is too simplistic to think it is just a matter of the young versus the very old. If we decide to impose minor restrictions with very little social distancing (or impose major restrictions at the last minute), the spread of the virus will cause much illness among younger
adults, and this is likely to overwhelm the health care system. This will have effects for the whole society because normal health care will not function properly. Furthermore, since it is easier for wealthy people than it is for poor people to cushion the negative effects of lockdowns and the widespread circulation of the virus, the trade-offs will also be between the wealthy and the poor, especially if there are no targeted efforts to help the poor. Finally, when designing vaccination plans, agencies will have to decide on a priority list. Difficult choices have to be made about whom to prioritize: those who have greater medical needs, those who are more likely to spread the virus further, or those who are needed to take care of the ill in elderly homes and hospitals. To make matters even worse, there is often grave uncertainty about the outcomes of the measures we take. So, in many cases, it is more accurate to say that we need to trade the risk of harm to one group against the risk of harm to another.

III. HOW TO UNDERSTAND TRADE-OFF QUESTIONS
The questions of how a certain trade-off should be resolved and how certain risks should be managed can be understood in more than one way. Moral philosophers tend to interpret them as asking for the trade-offs and risk managements that are morally right, and then they quickly turn to high-level moral theory for answers. More specifically, they tend to look for answers in some of the alternative foundational moral theories; for example, utilitarianism (maximize overall wellbeing!), Kantian ethics (act on maxims that you could will that everyone acts on!), and virtue ethics (do what the virtuous person would do!). There is nothing wrong with this, of course. High-level theorizing about the foundations of morality is important for the progress of moral philosophy, and it can also, if suitably popularized, provide important contributions to the public debate. But the questions need not be read as asking for the right trade-offs and risk managements. Instead, they can be understood as asking for how agencies, governments, and other institutions should be set up to identify and resolve trade-offs and manage risks. This is still a moral question but it is one about institutional design and it should not be conflated with the former, more abstract ethical questions. Now, the ‘filling the gap with a fundamental principle’ approach need not conflate these questions, but it nevertheless assumes that once we have found a philosophically sound,

---

1 See, for example, Public Ethics, a blog published by the Stockholm Centre for the Ethics of War and Peace (https://www.publicethics.org/).
fundamental ethical principle, it should be wheeled into the institutions and guide the decisions made by the relevant officeholders.

IV. Why is the ‘Filling the Gap Approach’ Problematic?
One obvious problem is that this approach often cannot even get off the ground since there is no agreement among philosophers on fundamental ethical issues. There is thus no unique set of principles that can be offered to the decision-makers. It will not comfort the decision-makers, who are in urgent need of ethical guidance, to be told: ‘Be right with you, as soon as we have resolved the disagreements between utilitarians and virtue ethicists’ (slightly modified from Arras 2020, section 3.1). Indeed, there are even deep disagreements within the different moral traditions (for example, act utilitarians disagree with rule utilitarians, Neo-Kantians with Kantians, and Aristotelian virtue ethicists with non-Aristotelian ones, and so on). Hence, there is simply no hope that the community of moral philosophers will come to agree on a set of fundamental ethical principles that could be then presented to decision-makers before the next pandemic or global crisis.

Another problem is that no set of fundamental ethical principles is likely to gain widespread societal approval. There are two main reasons for this. First, fully worked out fundamental principles tend to be quite complex or presuppose a thorough understanding of certain philosophical concepts (for example, ‘final value’, ‘maxims’, and ‘practical virtue’). Many ordinary decision-makers will therefore have difficulties grasping and distinguishing between these principles (Kymlicka 1993, 4).

Second, what characterizes many fundamental principles is that they take a stand on difficult moral trade-offs and provide a specific explanation of why this is the right one to make. But this is exactly why any such principle will be deeply controversial in a pluralist society. There is no general agreement about difficult trade-offs, and, even when there is agreement, different explanations of why the trade-off is right are often provided. But if there is no widespread support for a fundamental principle, it will not gain political legitimacy through democratic processes.

Suppose we have somehow managed to jump these hurdles and have come to agree on a fundamental principle that has also gained societal approval and it is now to be presented to the decision-makers at an institution. Are we in the clear now? No, and one reason is that fundamental principles are difficult to apply to concrete cases (Kymlicka 1993, 6–8). To see this, take utilitarianism as an example. Suppose the decision-
makers want to apply it when deciding on a suppression or mitigation strategy for Covid-19. They then need to know whether the strategy maximises overall wellbeing. But how do they establish this? They need to know all the options they have and the consequences of each. In order to assess the consequences of an option in terms of overall wellbeing, they need to aggregate the individual wellbeing of all affected individuals by summing them up (or averaging them). It is no surprise that the decision-makers do not have all the relevant knowledge to make such a decision. It does not help much to go for a version of utilitarianism that tells them to maximize the expected overall wellbeing, for this still requires them to assess the aggregated overall wellbeing of all possible outcomes of each option. Furthermore, they need to decide on an account of wellbeing: Is it your pleasure that is good for you, or your preference satisfaction, or freedom, or your friendships, knowledge, and achievements, or something else? Here the decision-makers are likely to be undecided or in disagreement and thus face evaluative uncertainty or disagreement.

V. Ethical Platforms

Instead of filling the gap between science and policy with a fundamental ethical principle, we can fill it with ethical-cum-legal platforms that have been approved through democratic processes. These platforms lay out in general terms certain ethical rules that have widespread support in society. These rules are supposed to be acceptable from very different ethical perspectives. From some perspectives, they capture (in part) what matters fundamentally, but from others they capture what is important instrumentally—that is, they are seen as rules we need to follow in order to respect or promote other more fundamental values. It is this ‘all-purpose’ feature of the rules that explains why they have gained wide societal support. Furthermore, since they are constructed with an eye to application in institutional settings (for example, hospitals), they have to be intelligible to and usable by the relevant officeholders and decision-makers.

Many countries have such platforms in place already, but they have not always been very useful during the pandemic. For example, the ethical platform for prioritizations in healthcare in Sweden is focused on individual-level medical interventions, such as giving one person priority for surgery over another (The Swedish Ministry of Health and Social Affairs 1996–1997). The overarching value is here the medical needs of the patient. One patient can be given priority over another only if the former has a greater medical need. It is not at all straightforward, however, to
apply this platform to measures in a pandemic, since many of these measures are applied at the population level, not the individual level, and it is not clear that medical need can always be the primary deciding factor. For example, in setting up a vaccination campaign for Covid-19, one may want to give extra priority to both vulnerable, older people with urgent medical needs and to certain members of the hospital staff, even though it is difficult to say that the latter group has urgent medical needs. The reason one may want to give priority to the hospital staff is not their urgent medical needs but rather the crucial role they play in supporting the healthcare system.

It is clear then that there is room for improving existing ethical platforms so that they are better poised to handle the kinds of trade-off that a pandemic brings to the fore. Good proposals are found in SMER (2020) and Munthe, Heilinger, and Wild (2021). These proposals do not only provide a set of rules applicable to the specific problems facing us in pandemics, but they also identify the uncertainties and conceptual distinctions that are crucial to take into account when managing a pandemic crisis.

However, revising the existing platforms in light of these proposals would not avoid the dilemma presented in the introduction: they would be general and vague enough to gain broad societal approval but this would come at the cost of not providing any clear guidance when it comes to difficult trade-offs. For example, in Munthe, Heilinger, and Wild (2021, 4), we are told that “what to consider in that balancing [that is, trade-off] must be supported by a valid ethical principle”, but nothing is said about what that valid principle would be. Similarly, in SMER (2020, 11) we are told that we need to “minimize harm and save lives”, but nothing is said about how we should weigh harm against lives. How can we give the decision-makers more guidance?

VI. ETHICAL FRAMEWORKS
One way forward is to let the ethical platforms remain general and quite vague in their recommendations, since they need widespread approval through democratic processes, and delegate the more fine-tuned and precise ethical guidance to ethical frameworks that can be adjusted to specific institutional settings. But what should such frameworks look like?

One popular idea is to rely on a cost-benefit framework to make the necessary trade-offs. Here all moral values are transformed into monetary gains and losses and the aggregation procedure is basically one of
balancing these gains and losses. This is done by asking people how much they are willing to pay for various values, such as freedom, health, and happiness. The problems with such an account in its unadorned form are obvious. It masks ethical disagreements about moral trade-offs by transforming them into an economic trade-off problem where it may look obvious that we should not accrue more financial losses than the financial benefits can compensate for. What is more, since rich people are more willing to pay for these values than poor, values of the rich will be given more weight than those of the poor, when in fact these values are as important (if not more) for the poor. Finally, since all individual costs and benefits are supposed to be summed up, the framework does not take into account inequality and fairness, and so is far from uncontroversial.

A different framework looks at the aggregated loss of life years adjusted by their health quality (so-called QALYs). This framework seems more relevant for moral choice, especially if the quality-of-life years is not defined in terms of willingness to pay for a certain health status. But it will still be too insensitive to equality if the aggregation is simple summation. It will also give less weight to helping the worse-off, since the life years gained by people who would, if saved, be in good health generate a higher health-related quality-of-life than the life years gained by people who would live with disabilities or in poor health. One could also wonder how this framework is supposed to capture other important values such as freedom and achievements. Not all values that are at stake in pandemics can be reduced to health factors.

Common to these two frameworks is that they give very limited guidance on controversial trade-offs. Many moral factors are simply ruled out. This makes them ill-suited for decision-making where many different ethical perspectives must be taken into account. Of course, there is nothing wrong in using the frameworks as one of many inputs in deliberation, but they cannot be the only input.

Is there a better framework that is still action-guiding but more inclusive of different ethical views? Recently, Adler et al. (2020) have proposed a social welfare analysis that could be used as an ethical framework for moral trade-offs in pandemics. Individual wellbeing is measured by a bundle of individual goods that can be compared across people, and moral choice is represented as a function of these bundles. More precisely, the input for this function are $n$-tuples of individual wellbeing values $(v_1, v_2, ..., v_n)$, where $v_i$ is the numerical representation of the wellbeing

---

2 Adler (2019) argues that this framework should be used for all policy-decisions.
of individual \(i\). The output is a ranking of bundles that can then be used for moral choice—it is permissible to bring about a top-ranked bundle. (Alternatively, the ranking step is bypassed, and the output is simply a set of permissible options.) Traditionally, social welfare analysis has been focused on preference-based wellbeing, but it can be generalized to take into account other, more objective values such as freedom and achievements. Furthermore, the choice function need not be one of simple summation. One can, for example, give more weight to worse-off people, and one can represent the disvalue of inequality in different ways (either as an individual harm or as an impersonal disvalue of the distribution).

One problem with this framework is that it is quite formal and mathematical and therefore requires training to be used effectively. Decision-makers will not have time for this, and if they defer to experts, they will have a hard time understanding why a certain option comes out as best in the framework. Another problem is that we might still be far away from any clear guidance on how to resolve trade-offs. Even if it is possible to represent diverse views by listing different values for each person and construct different choice functions, this is not enough to provide any clear action-guidance. It may clarify what the disagreements are all about, but the decision-maker needs more guidance.

Things are not hopeless, however. The problem with the high level of formalism can be solved to some extent if we focus on less mathematical and more informal versions of the social choice framework, where diagrams and other pictorial methods can facilitate understanding. There is still need of guidance on this, of course, but that could be given by philosophers and social scientists serving on ethical boards within the relevant institutions. So far, I have been sticking closely to the framework and ideas presented in Adler et al. (2020). In the following, I will introduce some ideas that go beyond those discussed by them.

The problem of not providing any action-guidance can be lessened somewhat by using a methodology that is part and parcel of social choice theory, namely, to identify a set of *conditions of adequacy* for any acceptable moral trade-off principle. Examples of such conditions are Pareto conditions (for example, making everyone better off is an improvement), non-dictatorship (no single individual’s interests are decisive), independence (what is an acceptable trade-off between two individuals regarding the choice between two options should not be affected by how the individuals fare in other options), and acyclicity (betterness cannot be cyclical: it cannot be that \(A\) is better than \(B\), which is better than \(C\), which is better than \(A\)).
The benefits of focusing on adequacy conditions is that it is likely that there is more agreement on and less uncertainty about them, for they provide only minimal constraints on acceptable trade-offs and are silent on why a certain trade-off is acceptable or not. For example, both utilitarians and virtue ethicists can accept the Pareto condition, but for different reasons. These conditions can then be used to weed out certain principles that violate all conditions. Among the remaining principles, a partial ranking can be generated: if a principle $P_1$ satisfies all the conditions that another principle $P_2$ satisfies, but in addition satisfies some other conditions, $P_1$ is to be preferred to $P_2$. Of course, this will not generate a complete ranking, since one principle may do better under one set of conditions and another principle will do better under a different set of conditions.

This can be remedied somewhat if the decision-makers agree on a ranking of the conditions in terms of their importance, for—other things being equal—violating a more important condition is worse than violating a less important condition. Perhaps acyclicity is seen as more important than independence, for example, since violating it threatens the very possibility of consistent choice. Note also that legally binding ethical platforms may rule in favour of certain conditions of adequacy but rule out others, for instance, those that together imply some verdict that goes against the principles in the platform. The ones that are ruled out are thus not just less important, but they cannot be invoked at all.

We still cannot expect a complete ordering, but we may not always need one for action-guidance. If a principle is ranked lower than all others, it can simply be discarded. If two principles, which cannot be ranked against each other, are ranked higher than all others, following either may seem to be a sensible choice at least in some circumstances. When a unique decision is required, the decision-makers with ultimate accountability, for example, the government, will have to make a choice between the top-ranked ones.

Those who are familiar with the literature on social welfare will now demur: Is not this reliance on conditions of adequacy a non-starter since it is well-known that there is no principle that satisfies them all? After all, so-called impossibility theorems are ubiquitous in the social welfare tradition. Even if we allow for interpersonal comparisons of welfare gains and losses and thus avoid Arrow’s original theorem, we are not out of the

---

3 For more on the role of conditions of adequacy in social welfare theory, see Gaertner (2009).
woods. As Fleurbaey, Tungodden, and Vallentyne (2009) and Fleurbaey and Tungodden (2010) show, even with this rich wellbeing information, there is no moral trade-off principle that satisfies a set of very plausible conditions, including the very plausible requirement that a huge welfare loss to a few worse-off people cannot be compensated for by small welfare gains for many very well-off people. Similarly, Arrhenius (forthcoming) and Blackorby, Bossert, and Donaldson (2005) show impossibility theorems for principles that take into account the wellbeing of future generations and yet allow for rich welfare information.

So, it is true that one tends to face impossibility theorems when constructing universal trade-off principles, but for these theorems to work they need to assume a rich domain of bundles of individual goods. There are possibility results that show that by restricting the domain it becomes possible for a principle to satisfy all conditions. This is good news for the application of the social choice framework to institutional settings, since it is clear that the domain will be severely restricted in these contexts. First of all, there are practical constraints; the institutions are quite restricted in what they can or cannot do (to a large extent set by budgetary constraints), so some bundles of individual goods are simply not feasible. But there are also legal constraints. The government tells the institutions what they have mandate to do, and legally binding ethical platforms will set the legal limits for the trade-offs that can be considered. Some bundles will be ruled out because there is no legally acceptable way to bring them about. One example is trade-offs that would violate some citizens’ legal rights. This means that the domain that is left for the ethical framework is likely to lack the richness that is needed for the impossibility results.

Of course, this means that the universality of the trade-off principles no longer can be maintained. But this is exactly what we want. Recall that the role of these frameworks is to facilitate decisions in contexts where very different ethical perspectives are represented. The focus must therefore not be on finding universal principles, but on finding workable principles for limited domains. Indeed, moral philosophers in the bioethics traditions have already done some useful work here, searching for what

---

4 For more Arrovian impossibility theorems in social choice theory, see Gaertner (2009). Note that these impossibility theorems are relevant even if interpersonal wellbeing comparisons are meaningful, for the decision-maker might lack information about such comparisons.

5 See, for instance, Gaertner (2001) for possibility results when wellbeing cannot be compared across people. See Arrhenius, Budolfsson, and Spears (2021) for a possibility result when wellbeing can be compared across people.
they often call ‘mid-level principles’ (Arras 2020, section 6.4.1). A lot of this work is wedded to a sceptical attitude towards abstract theorizing in moral philosophy. But you do not have to be a theory-sceptic in order to be interested in limited domain questions. Indeed, working out principles for ethical frameworks with limited domains can be seen as a form of modelling of certain parts of the moral landscape akin to the scientific modelling of certain restricted natural phenomena, such as the behaviour of the fish population in a certain lake. Common to both modelling approaches is that: (a) they are local and do not claim to capture phenomena outside the limited domain, but at the same time they do not rule out the possibility of finding more fundamental explanations that have a larger scope; (b) many factors are ignored because they are irrelevant, too complex to handle, or unable to make a significant difference; (c) some formal representations are used, often mathematical in nature but they need not be. If this kind of modelling appeals to theory-friendly scientists, it should also appeal to theory-friendly moral philosophers.

So far, I have suggested ways that evaluative disagreement and uncertainty about trade-off principles can be handled in ethical frameworks. But we have still to tackle the problems of application. Say we have found a set of eligible trade-off principles, perhaps principles about how to weigh wellbeing against freedom. We are still unable to apply these principles since there are many different ways to understand wellbeing and freedom. Does wellbeing consist in pleasure, preference satisfaction, or happiness? Does freedom consist in having options, so that the more options you have, the more freedom you enjoy, or does your degree of freedom also depend on the value of your options? Again, there is no chance that moral philosophers will have anything like an agreed stance on these issues. Even if they did, their views would not gain broad societal approval. The ethical frameworks thus will have to be as neutral as possible on these fundamental matters.

One way they can be more neutral is by focusing on proxies for wellbeing and freedom, factors that to some significant degree co-vary probabilistically with the respective value. Reliance upon proxies is unavoidable, since no one can directly measure every single individual’s wellbeing and freedom in society. This restriction can be turned into an advantage, however, for one can choose proxies for wellbeing or freedom that are

---

6 This analogy between normative and scientific modelling is developed in Roussos (2020).
7 One of the current projects of the author and Joe Roussos is to develop this point.
sufficiently coarse-grained to co-vary with *different* candidates for well-being or freedom. For example, answers to self-reported happiness questionnaires co-vary to a significant degree with happiness seen as life-satisfaction, but also with pleasure and preference-satisfaction, since when you are overall satisfied with your life you tend to feel pleasure and have satisfied preferences. Similarly, some freedom measures, such as answers to self-reported freedom questionnaires, will not be able to clearly distinguish between different notions of freedom.

The final problem with applying trade-off principles is how to deal with *empirical* uncertainty. We often do not know how a certain pandemic measure will affect people’s wellbeing and freedoms (or their proxies thereof). How should we weigh the risk of something bad against the chance of something good? Or, more specifically, when is it permissible to impose a certain risk of a freedom restriction for a certain group of people for the sake of wellbeing benefits for this or some other group of people? Here, again, we can expect disagreement. In fact, the questions of which risks are acceptable to impose on others are moral questions. So, what we have here is another source of a moral disagreement. But one thing people tend to agree on is that the persons the risks are imposed upon should have a say. This is one important aspect of how medical treatments and experiments are regulated in most countries. So, ethical frameworks need to be sensitive to this. Another thing to say is that since risk-imposition is to a large extent a *moral* issue, risk assessments should not be dealt with separately from the trade-off principles. Rather it is often better to formulate the trade-off questions directly in terms of risks: Is it permissible to impose a certain risk of harm to some people for the sake of certain expected gains for others? Disagreements and uncertainty about these risky trade-off principles can then be handled in a way that is similar to the non-risky trade-off principles, namely, by constructing a partial ordering of principles by invoking conditions of adequacy, ranked in terms of importance.

An important caveat: I do not pretend that this ethical framework will provide us with clear guidance for all controversial trade-offs. Disagreements and uncertainty, both about facts and trade-off principles, and even about some conditions of adequacy, will persist in many cases. That is why I can only claim to avoid the sharpest horns of the initial dilemma. The framework can only clarify and reduce these disagreements and uncertainties. We also need to rely on negotiations and compromises, but
they are unavoidable parts of any collective decision-making about contested issues.

**VII. CONCLUDING REMARKS**

Much more needs to be said before we have a detailed proposal of the kind of ethical framework that needs to supplement ethical platforms. For example, more needs to be said about the interplay between legally binding platforms and ethical frameworks, since the platforms to a large extent restrict the domains of the frameworks. But I hope that I have said enough to show the potential advantages of this approach for resolving trade-offs in pandemics and other crises.

**REFERENCES**


**Krister Bykvist** is a Professor of Practical Philosophy at Stockholm University and a Research Fellow at the Institute for Futures Studies, Stockholm. His research concerns questions about our responsibility for future generations, the foundations of consequentialism, evaluative uncertainty, and the relationship between preferences, value, and welfare. His most recent book is *Moral Uncertainty* (co-authored with Will MacAskill and Toby Ord), published in 2020 by Oxford University Press. Contact e-mail: <krister.bykvist@philosophy.su.se>
Collective Responses to Covid-19 and Climate Change

ANDREA S. ASKER  
*Stockholm University and Institute for Futures Studies*

H. ORRI STEFÁNSSON  
*Stockholm University and Swedish Collegium for Advanced Study*

**Abstract:** Both individuals and governments around the world have willingly sacrificed a great deal to meet the collective action problem posed by Covid-19. This has provided some commentators with newfound hope about the possibility that we will be able to solve what is arguably the greatest collective action problem of all time: global climate change. In this paper we argue that this is overly optimistic. We defend two main claims. First, these two collective action problems are so different that the actions that individuals have taken to try to solve the problem posed by Covid-19 unfortunately provide little indication that we will be able to solve the problem posed by climate change. Second, the actions that states have taken in response to Covid-19 might—if anything—even be evidence that they will continue to fail to cooperate towards a solution to the climate crisis.

**Keywords:** climate change, collective action, Covid-19, decision-making, harm

**JEL Classification:** D700, I120, I180

I. INTRODUCTION

The collective response to the Covid-19 pandemic is historically unprecedented (Ljungqvist 2020). In most countries, the public has supported, or at least accepted, measures that have restricted their freedoms and liberties in ways that have no counterpart in modern history. For the greater part of 2020, the majority of people in most European countries supported national lockdowns and restrictions (European Parliament Public Opinion Monitoring Unit 2020). Similarly, governments around the world have imposed measures that, predictably, have led to the

AUTHORS' NOTE: We are grateful to Krister Bykvist, Joe Roussos, the special issue editors of *EJPE*, and three anonymous referees for very helpful comments and suggestions that helped us improve the paper.
greatest economic recession in almost a century. For instance, the unemployment rate soared in both the United States and the European Union, and in many countries it rose even faster than it had done during the Great Depression of the 1930s (Krogstad and Noe-Bustamante 2021).

The fact that both individuals and governments have willingly sacrificed such great interests to meet the collective problem of the Covid-19 pandemic has provided some commentators with newfound hope about the possibility that we will be able to solve what is arguably the greatest collective action problem of all time: global climate change. For instance, The Guardian’s columnist Owen Jones wrote in the spring of 2020 that global responses to Covid-19 show that strong and urgent actions like those needed in response to the climate crisis “can be done” (Jones 2020). Similarly, Australia’s former chief scientist, Professor Ian Chubb, wrote in the summer of 2020 that the pandemic has taught us “that communities of individuals will act cohesively when the stakes are high”, which proved that threats like the climate crisis could similarly be responded to (Chubb 2020).

Unfortunately, we think that these commentators are overly optimistic. Our examination of the collective action problems that Covid-19 and the climate crisis constitute lead us to two conclusions. First, these two collective action problems are so different that the actions that individuals have taken to try to solve one (Covid-19) unfortunately provide little indication that we will be able to solve the other (climate change). Second, the actions that states have taken in response to Covid-19 might—if anything—serve as evidence that they will continue to fail to cooperate towards a solution to the climate crisis. Our main aim in this paper is to illustrate that it is far from clear that the observed responses to Covid-19 should make us optimistic about the chances of appropriate responses to the climate crisis.¹

II. INDIVIDUAL AGENTS
In much of the philosophical discussion of collective action problems there is a tendency to apply the same template to all such problems and treat them as structurally analogous. In doing so, however, there is a risk of missing important differences. We suggest that the inference

¹ Schliesser and Winsberg (2020) point out other important differences between climate change and Covid-19 (in particular, early in the pandemic), for instance, that the scientific understanding and cross-disciplinary discussion of climate change is much greater and more sophisticated than the understanding and discussion of Covid-19.
from the collective response by individuals to the pandemic to their anticipated response to the climate crisis is a case in point. Because the decision problem that individuals face differs in important respects between these two collective action problems, we should be careful in making inferences from one to the other. (By ‘collective action problem’, we here simply mean a decision problem where a group of at least two agents act individually, but, for each individual, the outcome that they get depends at least partly on the actions of others in the collective. An ‘agent’ can, for our purposes, be anything from a single person to a whole nation.)

The collective action problem posed by the pandemic is similar to that of climate change in that there is some morally significant outcome caused by many individuals acting in a certain way, but no individual action fully determines the outcome. In particular, both the pandemic and the climate crisis will be very harmful regardless of the actions of any one individual. Nevertheless, the decision problems faced by individuals differ between these two collective action problems, both in terms of the type of uncertainty people face, and in terms of how individual actions stand in relation to the harm produced by these two problems.

Consider first an individual (private) decision maker in the pandemic situation. To keep things simple we shall assume that, when it comes to Covid-19 decision problems, the individuals we consider have the same and rather simple preferences. In particular, we assume that they prefer the pandemic to end sooner rather than later and that they prefer its health and fatality effects to be less rather than more severe. In addition, we assume that individuals prefer that neither themselves nor their family and friends get infected by the coronavirus.

We shall moreover assume that these are the only preferences that are relevant to the decision problem we describe below. This is not to deny that individuals may have other interests related to the pandemic. For instance, they may prefer the economic effects of the pandemic to be less rather than more severe. Similarly, they may prefer that the measures to contain the pandemic limit their freedom as little as possible. But we assume that they do not take these interests to be relevant

---

2 Of course, limits on freedom and economic suppression may be a function of the severity of the pandemic, which in turn is largely determined by individuals collectively spreading the virus. But this is not the harm that we take individuals to be primarily concerned with when making decisions in this context, although it may of course be part of their total concern.
when making a decision in examples like that below. So, when an individual reasons about such a decision, the relevant outcomes concern primarily *how many* (if any) new infections and fatalities the decision may cause, and, perhaps, *who* might become infected as a result of the decision.

Now consider Jane, who has the above preferences, but wants to meet her friends and family in a neighbouring town. The only means she has of reaching the town is taking public transport. Let us assume that Jane can do so without breaking any law (as was true in most countries even at the height of the pandemic), although it goes against official recommendations.

What does Jane’s decision problem look like? To keep things simple, let us ignore any (for example, social) sanctions of violating Covid-19 recommendations. It would seem, then, that the main uncertainties that Jane has to consider concern, on the one hand, whether she is already infected, and, on the other hand, whether she would become infected if she were to take the trip. Moreover, given that most countries published and regularly updated information about the rate of infection in different areas and, in addition made testing widely available (at least during later waves), it would be plausible to assume that Jane can ascribe some meaningful probabilities to these uncertainties; that is, she has a meaningful way of assessing both the probability that she is already infected and the probability that she will get infected if she takes the trip. (In addition, the widely published $R$ number may give Jane an indication of how many she should expect to infect if she is or becomes infected.) This is one feature that makes individuals’ Covid-19 decisions importantly different from individuals’ climate decisions, as we shall shortly explain. Table 1 summarises Jane’s decision problem.

Jane’s preferences are as follows: if she is infected, then she prefers to stay home. If she would get infected, were she to take the trip, then she also prefers to stay home. However, if she is not infected, nor would get infected were she to take the trip, then she prefers to take the trip.

Since each new infection can be considered a significant harm (to the infected person), and since someone who is infected has a considerable chance of infecting at least one other person, an individual in Jane’s position has a considerable chance of making a significant difference in harm. This is another feature that makes individuals’ Covid-19 decisions importantly different from individual climate decisions, to which we turn next.
Consider an agent—call him Walt—who is concerned by the harm of climate change, but nevertheless enjoys going on Sunday drives in his gas-guzzling SUV. (We assume that most readers will recognise Walt’s predicament, at least if they substitute Sunday drives for their favourite greenhouse gas emitting activity, such as flying, buying new jeans, eating meat, and so on). Opinions differ as to whether Walt’s Sunday drive, and even his lifetime consumption and transportation choices, has any chance of affecting the harm of climate change. But for the analogy between climate and Covid-19 decisions to stand any chance—that is, to give some chance to the hypothesis we want to refute—let us assume that Walt’s drive has some chance of making some difference to the climate problem.

There is a broad consensus in the climate ethics literature that the problem facing Walt is either a triggering case or an imperceptible harm case, or potentially both. On the one hand, there is a possibility that the emission from his drive moves the concentration of greenhouse gas (or the corresponding temperature) closer to a threshold, which is such that if passed, it will trigger some harmful event, such as a storm, flood, or drought. On the other hand, it is possible that the greenhouse gas emission from his drive warms the climate by a tiny degree, which makes living conditions in some part of the world imperceptibly worse. So, either Walt’s actions will greatly increase harm by, say, increasing

---

3 For example, Kingston and Sinnott-Armstrong (2018) argue that an individual action makes no difference with respect to climate change, whereas Morgan-Knapp and Goodman (2015) and Broome (2019) argue that it can make a difference.
5 This simplified explanation might make it seem as if greenhouse gas concentrations build up linearly over time whereupon some individual action hits a threshold that triggers a harm, but this is not the only threshold view. Some recent accounts, such as Broome (2019), suggest instead that because of the atmosphere’s instability, a small disturbance may trigger significant climate harms without assuming that there are fixed thresholds.
the number of harmful climate events—which is (at least subjectively) unlikely—or it will cause imperceptible harm, or possibly both.

Now, it is widely accepted that climate change related decisions are associated with severe, maybe even extreme, uncertainty (Bradley and Steele 2015). In particular, there is no meaningful way in which Walt can estimate the probability that his action will make a perceptible difference when it comes to climate related harms. For instance, he cannot in any meaningful and precise way compare the probability that some harmful climate event (drought, storm, flood, and so on) will occur if he goes for a drive with the probability that such an event will not occur if he goes for the drive. At best, he may have good reason to believe that the drive is more likely to cause such an event than to prevent it; but it would be very implausible to claim that Walt has any meaningful way of assessing how much these probabilities differ. And the same holds for other potential climate related harms from his Sunday drive. So, in this sense, his uncertainty is more severe than (and qualitatively different from) Jane’s.

We are now in a position to see that Walt’s decision problem is in important ways different from Jane’s. First, as explained above, there is a difference in how severe their uncertainties are. Second, there are differences in how individual actions stand in relation to the respective outcomes. For instance, there is no chance that Jane’s decision makes the Covid-19 problem imperceptibly worse. If she makes the problem worse, then that is by getting herself infected, and/or infecting others, which is not imperceptible. Now, by getting herself infected, and/or infecting others, she might move society a tiny (perhaps imperceptible) bit closer to a threshold the passing of which triggers a lockdown or some other severe restriction. So, a part of the harm that her action could cause shares this feature with the harm that Walt’s climate decision could cause. However, our point is that if Jane’s actions cause (Covid-19 related) harm, then there is no chance that the overall (Covid-19 related) harm is imperceptible, since one more person getting infected is not imperceptible. And that is an important difference between her decision problem and Walt’s. In addition, as we have seen, there is, in the midst of the pandemic, a significant chance that Jane causes a significant harm

6 What about asymptomatic infections? If Jane infects a person who has no symptoms, and that person does not infect anyone else (or only causes an asymptomatic infection, and so on) then her action causes no harm, not an imperceptible harm. Unless, that is, the asymptomatic infection physically harms the person despite the lack of symptoms, in which case it would plausibly count as a perceptible harm.
by infecting one or more people, which could even result in them dying. In contrast, Walt will either cause a great increase in harm (which is unlikely) and/or cause an imperceptible harm.

There is an additional difference in terms of how individuals might reason about their respective decision problems. Jane might reason as follows:

If I take the trip, and one of the unfortunate outcomes occurs, then there is at least one *ex post* identifiable person whom I harm in the very near future. Moreover, if I cause someone to be infected, then that infection could potentially be traced back to me. I would not be able to live with that.

In contrast, Walt might reason as follows:

If I go for the drive, there are many people who are exposed to a tiny risk of harm. Even *ex post*, however, no actual harm can be traced back to me. In addition, these harms—if they materialise—are likely to occur far into the future, perhaps even after my death.

Now, one could of course doubt that these last differences between the two decision problems are *morally* relevant. That a harm cannot be *traced back* to my action would seem irrelevant to its moral standing. Similarly, whether a harm is immediate should not matter (much) morally. Nevertheless, these differences are relevant to how much we can infer about people’s behaviour from one problem to the other. The same is true of the differences between these decision problems when it comes to the level of uncertainty and the relationship between actions and outcomes. These differences undermine the plausibility of inferring, from the fact that a large part of the global population has shown restraint in situations analogous to Jane’s, that a large part of the global population can be convinced to show restraint in situations analogous to Walt’s. In other words, that many people seem to have sacrificed their short-term interest for the ‘public good’ in the Covid-19 pandemic is not, by itself, strong evidence that people will sacrifice their short-term interest in response to the climate crisis.

---

7 For instance, Sinnott-Armstrong (2005, 299; emphasis added) says: “No storms or floods, or droughts or heat waves can be *traced to* my individual act of driving”.
III. STATE AGENTS

Let us turn now to the collective action problem faced by states in the Covid-19 pandemic. As previously discussed, most states intentionally sacrificed an unprecedented proportion of their economy at least during parts of 2020 in order to fight the pandemic. In other words, states by and large prioritised containing the pandemic over avoiding negative short-term economic effects. It is not hard to understand why this has given some hope that states will take the necessary actions to limit the harm of the climate crisis. These actions will be costly in the short run, which is a reason why states have been reluctant to take them. But since states have shown willingness to accept such short-term costs in response to the pandemic, it might seem reasonable to believe that states will accept such costs in response to the climate crisis.

Unfortunately, an examination of the decision problem that states have faced and the actions that they have taken suggests that this newfound hope may be too optimistic, as we argue in this section. However, as in the previous section, our aim is not to conclusively show that the actions that have been taken do not bode well for the climate. Instead, our aim is to explain why there is reason to be sceptical of claims (such as those we cited in our introduction) that states’ actions in the face of Covid-19 do bode well for the climate.

To explain this, we need to make some empirical assumptions. But our analysis is not sensitive to any precise empirical assumptions; quantitatively close assumptions would lead to the same qualitative conclu-

---

8 Of course, containing the pandemic also involves avoiding some negative economic effects, for example, by preventing productivity loss from sick leave. In addition, an aggressive early reaction to the pandemic may be good for economic growth in the medium-term, as witnessed by China’s performance. After having shrunk by 6.8% in the first quarter of 2020, China’s GDP grew by 6.5% in the fourth quarter of the year (both in comparison with the same quarters the year before); and overall, China’s GDP grew by 2.8% in 2020, making it the only ‘major world economy’ to grow in 2020 (Cheng 2021). However, for our purposes, the important point is that many states took a gamble, whereby they sacrificed an unprecedented proportion of their economy in the short-term.

9 It is worth noting, however, that the current generations could, in theory, take these actions at no cost to themselves, by ‘borrowing’ from future generations, as Broome (2012) has pointed out.

10 What counts as ‘short-term’ versus ‘long-term’ is of course very different in the Covid-19 problem compared to the climate problem: the ‘long-term’ as far as Covid-19 is concerned is part of the ‘short-term’ as far as the climate crisis is concerned. This point of course only further supports our claim, namely, that strong action against Covid-19 might not provide much evidence that there will be strong action against the climate crisis. Since our aim is to focus on other arguments for our claim, we set aside this disanalogy between the two problems.
sion. Still, one could simply read what follows as claims conditional on the assumptions we make.

The first empirical assumption concerns the so-called *infection-fatality ratio* from Covid-19, which is the ratio (typically presented as a percentage) between the number of people who die from Covid-19 and the estimated number of people who have been infected by the coronavirus. As has been much discussed, the infection-fatality ratio is considerably lower than the rate reported in most countries, which is based on officially confirmed cases and is called the *case-fatality ratio*. A large number of Covid-19 cases are not officially confirmed, and mild cases are, of course, heavily overrepresented amongst these. And unsurprisingly, those countries that did the most testing reported the lowest case-fatality ratio (Ljungqvist 2020).

In Iceland, which is one of few countries that from the start of the pandemic tested even people without symptoms, the case-fatality ratio in July 2021 is 0.52% (The Directorate of Health 2021). Countries with less developed health-care systems than Iceland’s might have a higher infection-fatality ratio. But countries that have a lower proportion of very old people may have an even lower infection-fatality ratio (Lawal 2021). Thus, the average global infection-fatality ratio might not be far from the case-fatality ratio reported in Iceland, which we use below.

A second empirical assumption we need to make concerns how many people would have to be infected to achieve *natural* herd immunity, that is, herd immunity through natural infection as opposed to vaccination. The estimates vary, and differ for different countries, but the estimated figures for rich countries like Sweden, The Netherlands, and the United States are around 70%, but lower for most poorer countries (Kwok et al. 2020). Putting these two figures together, one could estimate that a reasonable upper bound (using the 70% figure) on the total number of fatalities from a world-wide natural herd immunity would be almost 29 million.

The social and economic impact of the measures taken to contain the pandemic have been much discussed. Global GDP is estimated to have fallen by 4.3% in 2020 (the highest annual drop since the Great Depression) (*The Economist* 2021), while unemployment rates soared. However, the impact on vulnerable people in developing countries has been even more severe. For instance, David Beasley, Executive Director of the United Nation’s World Food Programme, warned in December 2020 that the number of people at risk of starvation had doubled, from
135 million to 270 million, due to the global recession caused mainly by states’ response to the pandemic (The World Bank 2020; United Nations 2020).

It is instructive to try to compare the expected number of life years lost from such an extreme increase in the number of people risking starvation with the expected number of life years lost from world-wide natural herd immunity to the coronavirus. For that is one way of determining whether the cost of halting economic activity in order to contain the virus was worth paying. Such a comparison of course ignores the early deaths that we should expect as a result of increased unemployment in richer countries (Laditka and Laditka 2016), but it at least gives some indication of the cost, in terms of life years lost, of halting economic activity to contain the pandemic.11

A recent study has found that the average number of life years lost in a Covid-19 death in the United States is 9.34 (Quast et al. 2020). Assuming that the number would be the same in a global population with natural herd immunity, one could expect that the total number of life years lost through world-wide natural herd immunity would be almost 269 million. However, the average number of life years lost when someone starves is much higher. In fact, currently around half of those living in extreme poverty are under the age of 18. If, say, 10% of these 135 million additional people that risk starvation actually starve, and if, say, the average number of life years lost when someone starves to death is 40, then that adds up to 540 million life years lost. And note that this only includes the additional people that risked starvation as a result of the recession in 2020. As the recession continues, more people will surely risk starvation and experience other harms, such as poverty and missed education, which will eventually translate into additional life years lost.

In light of the above, it could well be that if all states had decided, at the start of the pandemic, to protect the very old and vulnerable while allowing “those who are at minimal risk of death to live their lives normally to build up immunity to the virus through natural infection”, as urged by the Great Barrington Declaration (Kulldorf, Gupta, and...
Bhattacharya 2020), then the world as a whole would have achieved an outcome that is better, in terms of the total number of life years lost, than the outcome that we are heading towards. Now, we should stress that the above is highly speculative and very far from conclusive. Moreover, the Great Barrington Declaration has a number of prominent critics, such as the head of the World Health Organization (The Guardian 2020). But, for now, we simply want to suggest there may be reasons to doubt that the strong actions that states have taken in response to Covid-19 are globally optimal, as for instance judged by life years lost.

Now, the global recession, which by the end of 2020 had resulted in the number of people risking starvation doubling, was not caused by the coronavirus as such, but rather the global response to the virus, such as the intentional halting of economic activity. Hence, a plausible counterfactual assumption, that we will make below, is that the global economic recession would have been considerably lessened (if not avoided) if states had decided, at the start of the pandemic, to protect the very old and vulnerable while allowing others to build up natural immunity. However, no state could have single-handedly avoided the global recession, and thus, given their reliance on the global economy, no state could have avoided a national recession. In addition, it may have been—or at least perceived to have been—in the national (short-term) interest of each state to impose strict restrictions. In particular, there was a widespread belief amongst policy-makers that lockdowns and other strict measures would reduce the number of local fatalities, even without some sort of global cooperation in the fight against the pandemic—as revealed by decisions to almost close borders while failing to properly cooperate on, for example, global vaccine programs.12

Now, such ‘vaccine nationalism’ (Eaton 2021) and increased (unilateral) border controls might in itself be taken as evidence that increased global cooperation in the fight against climate change is unlikely. In what follows we, however, want to make a somewhat different argument against the hypothesis that states’ actions in the pandemic are evidence that they will succeed in cooperating on a response to climate change; an argument that is based on examining more closely the collective action problem in which states have found themselves.

To make this argument we rely on two additional assumptions about the attitudes of the state agents with which we are concerned. First, we assume that each state has two goals that are relevant to the decision

12 See, for example, Eaton (2021).
problem at hand. On the one hand, each state would (other things being equal) like fewer rather than more of its citizens to get infected by Covid-19. On the other hand, each state would (other things being equal) like the measures taken to contain the pandemic to have as little negative economic effect as possible.

In light of the above, one could use Table 2 to formulate the decision problem that each state found itself in when deciding between imposing strict restrictions and aiming for natural herd immunity. To simplify, we assume that the states whose decision is not being modelled all act in the same way (that is, either they all opt for strict measures, or they all opt for herd immunity). Another simplification is that we ignore the many degrees that the different variables can take; for instance, we assume that there is either no local recession, or mild local recession, or local recession, and we assume that there are either more or fewer short-term local fatalities. Finally, by local fatalities, we here mean fatalities directly due to the coronavirus, not including the indirect fatalities from Covid-19 such as those due to poverty and unemployment.

Most states opted for strict restrictions. If the above is the correct framing of the decision problems that states faced, and, given that each state prioritised (in the short-term) fewer local fatalities over increased local economic performance, strict restrictions moreover seem to have been the rational choice. Whatever ‘others’ do in this problem, each state does better by imposing strict restrictions. Therefore, by each state acting individually rationally, they all impose strict restrictions—which is indeed the outcome we have observed.

However, given what we said above, one might reasonably think that the choice that most states made does not bode well for the prospects of solving the climate crisis. Granted, states have accepted an enormous economic cost to try to deal with the corona crisis, which, as previously mentioned, some have taken as evidence that perhaps they will do so too in response to the climate crisis. But recall that it is far from evi-

---

<table>
<thead>
<tr>
<th>Aim for herd immunity</th>
<th>Others aim for herd immunity</th>
<th>Others impose strict restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global recession</td>
<td>Global recession avoided</td>
<td>Global recession</td>
</tr>
<tr>
<td>No local recession</td>
<td>Mild local recession</td>
<td>Mild local recession</td>
</tr>
<tr>
<td>More (short-term) local fatalities</td>
<td>More (short-term) local fatalities</td>
<td></td>
</tr>
<tr>
<td>Global recession</td>
<td>Global recession</td>
<td>Local recession</td>
</tr>
<tr>
<td>Mild local recession</td>
<td>Local recession</td>
<td></td>
</tr>
<tr>
<td>Fewer (short-term) local fatalities</td>
<td>Fewer (short-term) local fatalities</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: States’ decision problem.
dent, as we have argued, that the outcome where each country imposes strict measures is optimal from a global point of view, for instance, when counting global life years lost. In fact, we suggested that there might have been fewer global life years lost had states all aimed for natural herd immunity. In addition, imposing strict measures seems to have been perceived by states to serve their own (short-term) interests.

The economic costs that states have accepted to try to limit the harm of Covid-19 may, therefore, have resulted in an outcome that is suboptimal for the global collective, even though it may have been optimal for each individual state. In contrasts, the economic costs that states will presumably (in the short run) have to accept to limit the harm from climate change are costs of reaching the outcome that is optimal for the global collective, but suboptimal for each individual state, at least as far as their own short-term interests are concerned (which would be best served by free-riding on the actions of other states). So, contrary to what seems to be a widespread view, it is far from evident that states’ actions in response to Covid-19 bring hope for the prospect of responding to the climate crisis.

IV. CONCLUDING REMARKS

In conclusion, we have argued that, contrary to common claims—both by journalists and scientists—it is at least questionable that individuals’ and states’ responses to the Covid-19 pandemic should give us much hope about an appropriate response to the impeding climate catastrophe. The Covid-19 decision problem that individuals face is simply too different from the individual climate decision problem for their actions in one problem to provide evidence of how they will act in the other. And it is far from clear that the Covid-19 decisions that most states have made should give us hope that global cooperation to solve the climate crisis will be reached.

REFERENCES


Andrea S. Asker is a PhD candidate at Stockholm University and research assistant at the Institute for Futures Studies in Stockholm. She has an interdisciplinary background with degrees in philosophy and economics from Stockholm University and Lund University. She works on topics in normative and applied ethics and decision theory. Her PhD thesis is on the topic of individual decision making in collective action problems.

Contact e-mail: <andrea.asker@philosophy.su.se>

H. Orri Stefánsson is an associate professor of practical philosophy at Stockholm University and Pro Futura Scientia fellow at the Swedish Collegium for Advanced Study. He holds a PhD in philosophy from the London School of Economics and Political Science, and has published a number of journal articles, chapters, and a book (together with Katie Steele), mostly on topics in decision theory, normative ethics, and Bayesian epistemology.

Contact e-mail: <orri.stefansson@philosophy.su.se>
Vaccine Refusal Is Not Free Riding

ETHAN BRADLEY
Oakland University

MARK NAVIN
Oakland University

Abstract: Vaccine refusal is not a free rider problem. The claim that vaccine refusers are free riders is inconsistent with the beliefs and motivations of most vaccine refusers. This claim also inaccurately depicts the relationship between an individual’s immunization choice, their ability to enjoy the benefits of community protection, and the costs and benefits that individuals experience from immunization and community protection. Modeling vaccine refusers as free riders also likely distorts the ethical analysis of vaccine refusal and may lead to unsuccessful policy interventions.

Keywords: fairness, free riding, game theory, public health, vaccination

JEL Classification: C72, I18

I. INTRODUCTION

The Covid-19 pandemic has highlighted problems of non-compliance with public health guidelines. Some people refuse to wear masks in public, to engage in social distancing, or even to avoid large gatherings (Breslow 2020; Glanz et al. 2020; van Rooij et al. 2020). Of particular interest to us is that some have indicated an unwillingness to receive Covid-19 vaccines (Freeman et al., forthcoming; Lazarus et al. 2021). Refusal of Covid-19 vaccines would therefore seem to be another instance of the general phenomenon of vaccine refusal, which has recently contributed to outbreaks of previously well-controlled infections (Jacobson, St. Sauver, and Finney Rutten 2015; Phadke et al. 2016).

Researchers have sometimes characterized vaccine refusal as a free rider problem (Bauch, Bhattacharyya, and Ball 2010; Betsch, Böhm, and Korn 2013; Betsch et al. 2017; Buttenheim and Asch 2013; van den Hoven
2012; May and Silverman 2005; Schröder-Bäck et al. 2009; Siegal, Siegal, and Bonnie 2009). That is, they have claimed that vaccine refusers promote their individual interest by benefitting from the community’s protection from infection without also making a contribution to the community’s protection by becoming vaccinated themselves. Notably, this characterization of vaccine refusers applies only in cases in which the community already has protection against infection—that is, when a sufficiently large percentage of the population, having been vaccinated, is immune to the disease so that outbreaks are unlikely. Most countries possess community protection for many vaccine-preventable diseases, including, for example, polio (Global Polio Eradication Initiative n.d.). Though none yet have community protection against Covid-19, we may hope that ongoing vaccine distribution efforts will soon generate it.

The diagnosis of vaccine refusal as a kind of free riding is supposed to inform our understanding of the origins of vaccine refusal (vaccine refusers rationally pursue their own interests), the ethical analysis of vaccine refusal (vaccine refusers make unfair use of community protection), and potential policy responses (vaccine refusers will vaccinate if one makes slight changes to their incentives).

But, as we shall argue, vaccine refusal is not a kind of free riding. First, a free rider model misrepresents the subjective motivations of most vaccine refusers. Vaccine refusers often doubt that vaccines provide benefits to individuals and communities, and often radically overstate the risks associated with vaccination. That is, vaccine refusers do not think they are using a valuable public good (community protection) or that they are refusing to make a reasonable contribution to that good since they think the expected costs of vaccination are very high. Second, a free rider model misstates the objective relationship between individual vaccination choices and one’s ability to benefit from community protection. A free rider benefits from a public good that they could also be contributing to, but it is not possible to both contribute to community protection and to benefit from it. This is because the means by which a person contributes to community protection (individual immunity, usually generated through vaccination) makes it impossible for that same person to benefit

---

1 We follow contemporary usage in referring to ‘community protection’, rather than ‘herd immunity’, because the community is protected rather than immune and because we are speaking of a moral community of human persons rather than a passive animal herd (Anderson et al. 2018).
from community protection. Furthermore, the act of ‘contributing’ to community protection (by becoming vaccinated) is, by itself, individually rational, even when community immunization rates are high, because there is rarely an objective anticipated cost of vaccination for individuals. Free riding is individually rational, by definition, but refusing vaccines rarely is.

II. VACCINE REFUSAL AS FREE RIDING

An individual is a free rider if they enjoy the benefits of a public good (that is, a good that is non-rivalrous and non-excludable) without also contributing to it, even though they could also contribute to the good, and even though the personal benefits they acquire from the public good would outweigh their personal costs of contribution (Olson 1965). Free riding creates a collective action problem: it is individually rational to free ride, but if enough people free ride, then no one can, since insufficient contribution undermines or destroys the good that they were attempting to enjoy. That is a collectively irrational outcome. In this way, free riding problems may take the shape of a Prisoner’s Dilemma, wherein a social good can be created by cooperation, but each individual benefits from defecting (Hardin 1971).

Furthermore, some have argued that free riding is ethically wrong when the public good is sufficiently valuable and the costs of contribution are sufficiently low. On this kind of view, free riding is not only collectively irrational, but also individually unethical, because it is unfair to benefit from the contributions of others without making a reasonable contribution oneself (Cullity 1995).

If vaccine refusal were a kind of free riding, then (1) mass vaccination must create a public good, (2) it must be possible for individuals to contribute to that good and also to benefit from it, though (3) it must be better for individuals not to contribute, but just to benefit from it. Furthermore, vaccine refusers would have an intention to free ride just in case they intended to enjoy the benefits of the public good that mass vaccination creates, but they did not intend to pay the (small) costs associated with contributing to that good.

2 An individual can (1) intend to contribute to community protection (by getting vaccinated) without actually contributing to community protection (if the vaccine does not give them immunity), and (2) continue to benefit from community protection. But such a person does not both contribute to and benefit from community protection.
Mass vaccination creates a public good. ‘Community protection’ (sometimes called ‘herd immunity’) exists when a community has a sufficiently high rate of individual immunity, such that outbreaks are highly unlikely and such that it is highly improbable that individuals who are not immune to infection will become infected (Anderson et al. 2018). The logic of free riding on community protection may seem straightforward: vaccination has expected costs for individuals, including time, money, and risks of side effects and adverse events. Community protection has many expected benefits for individuals, including decreased morbidity and mortality. The benefits an individual enjoys from community protection outweigh the individual costs associated with vaccination. However, someone who free rides on community protection enjoys the benefits of community protection without paying any costs associated with contributing to it.

Scholars writing about empirical issues in vaccine refusal have sometimes modeled vaccine refusal as a form of free riding (Bauch, Bhattacharyya, and Ball 2010; Betsch, Böhm, and Korn 2013; Betsch et al. 2017; Buttenheim and Asch 2013; van den Hoven 2012; Ibuka et al. 2014; May and Silverman 2005; Siegal, Siegal, and Bonnie 2009). The details of their models differ. For example, Bauch and colleagues (2010) show that free riding practices can develop quickly after the introduction of a vaccine that is universally and freely available but not mandatory. For another example, Ibuka and colleagues (2014) found that individuals were less likely to vaccinate in a simulation when the rate of vaccination was high in the previous round of the simulation. But the common thread that ties together ‘free riding’ models of vaccine refusal is that vaccine refusers pursue their rational self-interest in avoiding the costs associated with vaccination while benefiting from the vaccination behaviors of others.

People who have written about the ethical issues involved in vaccine refusal have also sometimes invoked the idea that refusers are free riders who make selfish and unfair use of others’ contributions to community protection (Clarke, Giubilini, and Walker 2017; Dawson 2007; Giubilini, Douglas, and Savulescu 2017; Giubilini 2019; Hendrix et al. 2016; van den Hoven 2012; Navin 2016; Salmon and Siegel 2001). Some have argued that one goal of immunization policy is to fairly distribute the burden of achieving community protection while avoiding unfair free riding. For example, they have proposed that nonmedical exemptions to vaccine mandates only be granted to people who make an alternative contribution to public health—that is, contributions similar to those of some conscripted
pacifists who are asked to provide alternative (for example, non-combatant) services to a war effort (Clarke, Giubilini, and Walker 2017; Giubilini, Douglas, and Savulescu 2017).

III. WHY NOT FREE RIDING?

Vaccine refusal is not a case of free riding. Vaccine refusers do not have the subjective beliefs and attitudes of free riders, and the objective costs and benefits associated with vaccine refusal are not consistent with free riding.

III.I. Beliefs and Attitudes

Vaccine refusers rarely possess the beliefs and attitudes towards vaccination and community protection that are consistent with them being subjective free riders. Someone intends to free ride when they recognize the benefits of a public good that they are enjoying and when they acknowledge that they are benefitting from that good without making a reasonable contribution to it. For example, in Prisoner's Dilemma models, a defector (who wishes to be a free rider) seeks to take advantage of the benefits of other people's cooperative behavior without themselves engaging in cooperative behavior. Vaccine refusers do not fit this profile. First, vaccine refusers often believe that vaccines are ineffective, that vaccines do not cultivate individual immunity to disease, and that community protection does not exist (Harmsen et al. 2013; Sobo 2016). Others are wholly unfamiliar with the concept of community protection (Quadri-Sheriff et al. 2012; Sobo 2016). Such vaccine refusers cannot be motivated by a desire to take advantage of others' cooperative behavior because they do not think other people's vaccination choices create a public good. There are some exceptions. For example, Dr. Bob Sears recommends that parents who refuse vaccines for their children do not “share their fears with their neighbors, because if too many people avoid the MMR [measles, mumps, and rubella vaccine], we'll likely see the disease increase significantly” (Sears 2007, 96–97). The clear implication is that parents of children in Sears' practice should want other parents to continue to vaccinate so that their own children will be safe from disease. But it is rare to see vaccine refusers acknowledge that they are benefitting from other people's decisions to vaccinate. A more common view is that vaccines are ineffective, that they are collectively harmful, and that everyone would be better off if no one were vaccinated (Dubé et al. 2013; Harmsen et al. 2013; Sobo 2016).
Second, part of the idea of free riding is that it better promotes an individual’s interests than contribution does, but that contributing to (and enjoying) a public good is better for the individual than non-contribution and non-enjoyment of the public good would be. This is because an individual’s costs of contribution are outweighed by the benefits they receive from the public good. So, even though free riding on a public good is individually rational, contribution is also beneficial (though less so than free riding) compared to not enjoying the public good at all. In the case of modeling vaccine refusal as free riding, a free rider would better promote their own interests than would someone who were vaccinated, but a vaccinated person enjoying community protection would better promote their own interests than would someone who did not benefit from community protection at all. On such a view, an individual’s expected costs of vaccination are more than compensated for by the benefits they receive from community protection.

But many vaccine refusers do not believe that the expected costs of vaccination are low. They reject the scientific consensus that vaccines usually have only mild side effects and that the risks of serious adverse events (for example, death or life-threatening illness) are very low (Smith 2015). Vaccine refusers commonly believe that vaccines often cause serious disorders (including autism), damage one’s immune system, degrade one’s genetic code, or place one’s body under unacceptable forms of government surveillance or control (Dubé et al. 2013). Someone who intends to free ride aims to avoid paying a reasonable price for a public good. But many vaccine refusers think that vaccination is dangerous and that it would be unreasonable to ask them to vaccinate. Such people conceive of vaccination choices as a matter of avoiding or acquiescing to significant harms, rather than as a matter of whether or not to contribute to a public good from which they benefit.

III.II. Costs and Benefits
Vaccine refusal also usually fails to meet objective criteria for free riding. First, vaccine refusers are not free riders on community protection, because it is not possible to both contribute to and benefit from community protection. A free rider benefits from a public good to which they could also contribute. When free riding is immoral, free riders should also contribute to the public goods from which they benefit. The way to contribute to community protection is to possess individual immunity, either through vaccination or by recovering from a disease. But people who
possess individual immunity do not (and indeed, cannot) rely on community protection to protect themselves from the diseases against which they have individual immunity. They rely on individual immunity. Accordingly, someone who contributes to community protection cannot also benefit from it. Consequently, someone who does not contribute to community protection—for example, by refusing vaccines—is not a free rider.

Second, free riding is individually rational, but vaccine refusal is not. Vaccination almost always promotes the interests of the vaccinated individual, even at very high levels of community protection. Since the costs of vaccination are generally negligible—for example, serious complications are exceedingly rare (Spencer, Trondsen Pawlowski, and Thomas 2017)—it is almost always in a person's interest to vaccinate, even when community protection makes their odds of infection very low. Consider the fact that community protection against some vaccine-preventable infections requires very high levels of population-level immunity, such that even universal vaccination with very effective vaccines may still leave someone vulnerable to exposure (and therefore make it rational to be vaccinated). For example, roughly 95% of the population needs to be individually immune to measles to eliminate outbreaks (Gay 2004; Moss and Strebel 2011), while only 95% of people fully vaccinated against measles have individual immunity (Demicheli et al. 2012), so that communities with less than 100% measles vaccine uptake remain at risk for measles outbreaks, and therefore individual vaccination against measles is almost always rational (Bester 2017). More importantly, it is very difficult to know one's risk of being exposed to a vaccine-preventable infection, even if reliable community-wide immunization rates are available. Outbreaks commonly appear in small geographically clustered groups of under-vaccinated persons (Omer et al. 2008; Phadke et al. 2016), and it can be all but impossible to identify and avoid such groups. In contrast, vaccination reliably generates individual immunity at negligible costs. Notably, attempts to model vaccine refusal as an instance of free riding sometimes presume that individuals can make reliable predictions about the proportion of immune individuals in the groups they interact with (for example, Betsch, Böhm, and Korn 2013). Yet, it seems highly unlikely that most vaccine refusers possess (or even could possess) such knowledge.

IV. ‘FREE RIDING’ AND POLICY INTERVENTIONS
If policy responses to vaccine refusal are to be effective and ethically justified, then they should be based on an accurate understanding of vaccine
refusal. The claim that vaccine refusers are free riders may lead to policy proposals that are ineffective or morally unjustified.

**IV.I. Incentives and Pro-Vaccine Interventions**

There are a set of standard interventions for solving free riding problems (Buttenheim and Asch 2013). These include offering economic incentives for cooperation (or disincentives for defection), restricting defectors’ access to other goods, invoking or creating social norms to pressure individuals to cooperate, or using state coercion. These enforcement mechanisms shift the relevant payoffs in ways that incentivise potential free riders to cooperate, instead of defect, by making it rational for would-be free riders to contribute (Buttenheim and Asch 2013).

If vaccine refusal were a free rider problem, then one way to encourage vaccination would be to slightly shift the perceived payoffs of potential vaccine refusers. Governments should be able to avoid vaccine refusal with small (or even trivial) amounts of incentives or disincentives. Recall that actual free riders acknowledge the value of the public good they enjoy but refuse to pay the small cost of contributing to that good when they can get away with not incurring the cost. A government could therefore overcome ‘free rider vaccine refusal’ if it increased the costs of vaccine refusal so that those costs were somewhat higher than the costs of vaccination. This means that placing small additional burdens on vaccine refusers should lead to overcoming vaccine refusal that results from subjective free riding.

The evidence is not consistent with the hypothesis that the behavior of most vaccine refusers can be changed by making small modifications to their incentives. For example, many countries have recently adopted or revised childhood vaccine mandates as a way of providing additional incentives for vaccination or of disincentivizing vaccine refusal. For example, Australia’s federal government withholds state payments to parents who do not vaccinate their children (‘No Jab, No Pay’), while their state governments prevent unvaccinated children from being enrolled in childcare (‘No Jab, No Play’) (Attwell et al. 2018). All US states require children to be vaccinated to enroll in school, and many states have recently made those enrollment mandates more difficult to escape, for example, by eliminating nonmedical exemptions or imposing burdensome administrative procedures (National Conference of State Legislatures 2021). Other political communities, like Italy, have decided to fine parents who do not vaccinate their children (Vaz et al. 2020).
These kinds of policy changes sometimes somewhat increase vaccination rates or decrease nonmedical exemption rates (Omer et al. 2012; Navin, Largent, and McCright 2020). That is not evidence that the parents who change their minds in the face of such changed payoffs are free riders—we have already presented evidence to motivate skepticism about that conclusion—but the number of parents who do change their behaviors clearly identifies a ceiling for the maximum number of possible free riders among vaccine refusers. Even if we supposed that every parent who changed their vaccination behavior in the face of changes in their incentive structure were a free rider, then the evidence suggests that perhaps around a third of parents could possibly be free riders—that is about the maximum number of vaccine refusers who seem willing to change their behavior in the face of substantial changes to their incentives. Notably, even quite serious penalties seem insufficient to change the behavior of many vaccine refusers. For example, Delamater and colleagues (2019) found that eliminating nonmedical exemptions in California caused few parents to vaccinate their children. In fact, the positive results of that policy change were almost entirely a consequence of better administrative oversight and record-keeping, or were offset by corresponding increases in medical exemptions or the enrollment of children not in compliance with immunization requirements. Many thousands of California parents allowed their children to be denied access to public or private schools, or to be prevented from participating in group homeschooling activities rather than be vaccinated. This is not the profile of a free rider.

Pro-vaccination policies that are likely to overcome free riding behavior, especially subjective free riding behavior, are unlikely to be as effective against someone who refuses vaccines because they believe vaccines are ineffective or dangerous. If someone does not think that community protection is valuable, or if they think vaccination is a grave risk to their health, then small changes to their incentive structure are unlikely to change their minds. Relying on a free rider model of the attitudes and motivations of vaccine refusers may therefore encourage unjustified optimism that vaccine refusal can be overcome with minimally invasive interventions.

IV.II. Ethical Justifications for Pro-Vaccine Interventions
The mere fact that community protection is a public good does not, by itself, justify government coercion (Bernstein and Randall 2020), but the substantial individual and collective benefits of community protection
strongly favor government efforts to promote vaccination (Brennan 2018; Flanigan 2014; Giubilini and Savulescu 2019; Giubilini 2019, 2020; Navin 2016; Pierik 2018). If vaccine refusal were objectively a form of free riding, then one weighty ethical reason for societies to promote vaccination would be to promote fairness in the distributions of benefits and burdens associated with mass immunization. The people who benefit from community protection should be incentivized or even compelled to contribute to community protection, rather than be allowed to unfairly free ride on the socially productive contributions of others, as long as they can do so at a reasonable cost to themselves.

But vaccine refusers are not objectively free riders, even though some of them may think they are free riding, as in the case of Dr. Bob Sears’ patients we discussed above. The goal of pro-vaccination policies, therefore, should not be to punish people for free riding, that is, to make sure that they ‘pay’ for the benefits that they enjoy from community protection. As we argued above, it is not possible both to contribute to and to benefit from community protection. Inasmuch as fairness plays a role in ethical arguments for pro-vaccination policies, it should be to make sure that there is a fair distribution of labor among the members of society who have a duty to protect others from harm (Giubilini, Douglas, and Savulescu 2018). That is, fairness in vaccination behavior is about doing one’s part to help others. It is about being a contributor to community protection, rather than a beneficiary of it, whenever one is reasonably able to do so. But fairness with respect to vaccination is not about ensuring that one is paying for the public goods they are benefiting from: this is not possible.

Aside from the fact that vaccine refusal is not objectively a free rider problem, it also matters for an ethical analysis that vaccine refusers are rarely subjectively free riders. Vaccine refusers usually do not think they are acting unfairly. They do not believe they are taking advantage of a public good that they could also be contributing to. Instead, they think that (almost) everyone would be better off if they refused vaccines. So, inasmuch as vaccination policies aim to punish vaccine refusers for their unethical behavior, it would be incorrect to focus on refusers’ supposed intentions to free ride. Relatedly, inasmuch as vaccine advocates express anger at vaccine refusers for (what they believe to be) unfair free riding (Bernstein 2021), those moral emotions are not entirely apt. Vaccine refusers seem to be motivated by ordinary ideas about avoiding serious risks to their health based on false beliefs about the costs and benefits of
vaccination and about the risks of vaccine-preventable infections. We acknowledge that vaccine refusers may sometimes be morally responsible for these false beliefs and that they may be acting unfairly, but their moral failure is not the straightforward unfairness of free riding.

V. CONCLUSION

Many people have endorsed pro-vaccination interventions or made ethical arguments about vaccine refusal that presume vaccine refusers are free riders. We have argued that vaccine refusal is not an instance of free riding and that few vaccine refusers believe themselves to be free riding. Accordingly, effective policy interventions and apt ethical judgments about vaccine refusal will need to rely on other explanations for why people sometimes refuse vaccines.

If those who study vaccine refusal want to introduce game-theoretic models to explain this phenomenon, then perhaps they should turn away from free rider models and consider embracing information problem models. The core dynamics of information problem game-theoretic models is that participants in the game are not fully aware of the ‘rules’ of the game, for example, who else is playing, what each participant’s interests are, and what actions have been taken previously (Jones 1977). Participants must therefore make educated guesses about the best course of action, rather than calculate a deterministic result. In light of the prominent role that false beliefs play in the decisions of vaccine refusers, such models may present more accurate depictions of the objective and subjective attributes of vaccine refusal than free rider models do.

REFERENCES


Ethan Bradley is an undergraduate student of philosophy and political science at Oakland University (Rochester, MI, USA). Contact e-mail: <embradley@oakland.edu>

Mark Navin is professor and chair of philosophy at Oakland University (Rochester, MI, USA). Contact e-mail: <navin@oakland.edu>
Past and Future of Humanomics: A Conversation with Deirdre Nansen McCloskey

Deirdre Nansen McCloskey  
*University of Illinois at Chicago*

Paolo Silvestri  
*University of Turin*

Deirdre Nansen McCloskey (Michigan, 1942) is Distinguished Professor Emerita of Economics and of History, and Professor Emerita of English and of Communication, at the University of Illinois at Chicago. Trained at Harvard in the 1960s as an economist, she has written twenty-five books and some four hundred academic articles on economic theory, political theory, economic history, philosophy, rhetoric, statistical theory, feminism, ethics, and law.

She taught for twelve years at the University of Chicago in the Economics Department in its glory days, but now describes herself as a “literary, quantitative, postmodern, free-market, progressive-Episcopalian, ex-Marxist, Midwestern woman from Boston who was once a man. Not ‘conservative’! I’m a Christian classical liberal” (McCloskey, n. d.).

Her most recent popular books, for example, are *Why Liberalism Works: How True Liberal Values Produce a Freer, More Equal, Prosperous World for All* (2019a); with Art Carden, *Leave Me Alone and I’ll Make You Rich: How the Bourgeois Deal Enriched the World* (2020); and, with Alberto
Mingardi, *The Myth of the Entrepreneurial State* (2020). Also, in 2019, the University of Chicago Press published a third edition of her classic manual on style, *Economical Writing* (2019b), and a twentieth-anniversary reissue of *Crossing: A Transgender Memoir* ([1999] 2019c), with a new Afterword. She’s technical and quantitative, too. For example, with Stephen Ziliak, in 2008, she wrote *The Cult of Statistical Significance*, widely praised, which shows that null hypothesis tests of ‘significance’ are, in the absence of a substantive loss function, meaningless. The point, made long before McCloskey by a few statisticians, is becoming widely accepted, for example, in the American Statistical Association, though not yet in economics and medicine.

Her latest scholarly book, again from the University of Chicago Press, *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World* (2016b), was the final volume of the Bourgeois Era trilogy (2006, 2010, 2016b). It argues for an ‘ideational’ explanation of the Great Enrichment of 3,000 percent per person from 1800 to the present in places like Britain, and Japan, and Finland. The accidents of Reformation and Revolt in northwestern Europe, during 1517–1789, led to a new liberty and dignity for commoners—ideas called ‘liberalism’ in the proper sense—which led in turn to an explosion of commercially tested betterment, ‘having a go’. The second book in the trilogy, *Bourgeois Dignity: Why Economics Can’t Explain the Modern World* (2010), had shown that materialist explanations such as saving or exploitation, don’t have enough economic oomph or historical relevance to explain the Enrichment. The alleged explanations that do not focus on the new ideology of innovism—her name for the ill-named ‘capitalism’—are mistaken. And the Enrichment did not corrupt our immortal souls. The inaugural book in the trilogy, *The Bourgeois Virtues: Ethics for an Age of Commerce* (2006), had established that, contrary to the clamor since 1848 by the clerisy left and right, the bourgeoisie is pretty good, and that commercially tested betterment is not the worst of ethical schools. In short, the trilogy looks forward, if populism does not spoil the prospect, to a world of universal dignity and prosperity created by liberal innovism.

Paolo Silvestri interviews Deirdre McCloskey on the occasion of her latest book, *Bettering Humanomics: A New, and Old, Approach to Economic Science* (2021a). The interview covers her personal and intellectual life, the main turning points of her journey and her contributions. More specifically, the conversation focuses on McCloskey’s writings on the methodology
and rhetoric of economics, her interdisciplinary ventures into the humanities, the Bourgeois Era trilogy with its history of the 'Great Enrichment', her liberal political commitments, and the value and meaning of liberty, equality, and solidarity. Finally, the conversation returns to McCloskey’s ‘humanomics’ approach: an economics with the humans left in.

**PAOLO SILVESTRI:** You have described yourself as a “literary, quantitative, postmodern, free-market, progressive-Episcopalian, ex-Marxist, Midwestern woman from Boston who was once a man. Not ‘conservative’! I’m a Christian classical liberal” (McCloskey, n. d.). It seems to me that this description may be a pivot around which our conversation can rotate. More exactly, I would like to develop the questions of this interview through a game of ‘cross-references’ between your career, personal and professional, the main turning points of your intellectual life, and what you have achieved with your publications.

Let’s start with your ‘Humanomics project’, and the (just published) book, Bettering Humanomics: A New, and Old, Approach to Economic Science (McCloskey 2021a). Can we say that the Humanomics project has been, without you quite knowing it, your life-long project and achievement?

**DEIRDRE NANSEN MCCLOSKEY:** Yes. I only realized in the past couple of years that humanomics is what I have been struggling to say from class poet in secondary school to old-lady economist. That’s sixty years to get the point. I am not the swiftest of thinkers! Economists like Arjo Klamer (as in Klamer 2017 and earlier writings), and Don Lavoie (1985), and Albert Hirschman (1977), and Kenneth Boulding (1956), and of course the Blessed Adam Smith, did humanomics well before letter. As I say in a forthcoming intellectual autobiography:

It’s an advantage in having an intellectual development to be a little stupid, or at best earnestly naïve, as on both counts I am. Natural economists—natural because their personalities make it a snap for them to grasp how a maximizing person would behave—find it hard to develop. They are too smart at the outset. They get it immediately,

---

1 I (Paolo) had the pleasure and honor of following and previewing this book when it was still in its ‘embryonic’ state, that is, before the embryo split in two, generating its twin brother, Beyond Behaviorism, Positivism, and Neo-Institutionalism in Economics (McCloskey, forthcoming(c)). We held another, and more specific, conversation about this book in McCloskey and Silvestri (2021).
and then keep it, forever.² On the contrary, I am not a natural economist. It was hard for me to learn economics of the Samuelsonian and Friedmanite and Koopmanslijk sort. It didn’t fit my personality or ethical upbringing. But I declare on oath here that at each stage of learning, from child to old lady, I was stupidly, naively sincere, believing earnestly in each of the half-truths grasped along the way. Really.

My little witticisms, to which I am addicted, give people sometimes the opposite impression, of an insincere ironist of the sort the great Joseph Schumpeter was. He was unable to quite let go of his early (and sound) understanding of markets, or then of his youthful (if disproportionate) admiration for a Marxian sociology, or then of his mature (if premature) findings on economic and intellectual history, all of them busily making ironic little jokes about each other throughout his book of 1942, Socialism, Capitalism, and Democracy.³ In my case, no, believe me: I am a serial believer. Contradictory as are the successive economic opinions I have held, I was stupidly loyal to each. (McCloskey, forthcoming(a))

So I had to trudge slowly, slowly through nearly every rhetorical and epistemological position relevant to economics one could imagine, from positivism to postmodernism, until, as I now suppose, I got it approximately right. The purpose of humanomics is to gather all the evidence—counts and categories, both. You can’t count until you know through humanistic, rhetorical, ethical, philosophical inquiry what the justified and humanly interesting categories are. It is one reason for the unique program in the Department of Philosophy at Erasmus University, in which I taught for many years, the Erasmus Institute for Philosophy and Economics (EIPE). Get the categories straight, if you don’t want to look like a fool when you turn to counting. Count national income, to be sure—but which definition of the nation, or of income? French overseas possessions? Diasporas? Housewives’ work? Environment? As grown-up economic scientists, we need to use all the evidence—experimental, simulative, introspective, questionnaire, graphical, categorical, statistical, literary, gossipy, historical, journalistic, psychological, sociological, political, theological, ethical. Get the numbers right and the qualities to be numbered.

Let’s go back, then, to the origins. In the preface to Bettering Humanomics, you trace the origins of that work to turning points in your career, in particular, the books The Rhetoric of Economics ([1985] 1998), If You’re So Smart: The Narrative of Economic Expertise (1990), and

³ See Schumpeter ([1942] 1950) and McCloskey (forthcoming(b)).
Knowledge and Persuasion in Economics (1994). *I think this particular turning point, among many in your life, was a big leap forward—considering your conventional training at Harvard, then your conventional university career getting tenure at Chicago. What caused this ‘linguistic turn’ in your intellectual journey? (I use the expression ‘linguistic turn’ not by chance, given your relationship with Richard Rorty and his first wife, Amélie Oksenberg Rorty, from whom you got your often repeated motto ‘listen, really listen to your friends’ questions and objections’.)*

Our lives depend on tiny incidents, don’t they? Whom you marry, what you do, books you read, incidental encounters. Of course, the incidents have their effect on a person or a people who are prepared in some way. Having learned by age 37 how to be a conventional economist, I was by then thoroughly irritated by the method-talk of both my Harvard teachers and my Chicago colleagues, each claiming the white coat of Science and sneering at the other place. Both the Harvardsian talk about realism and imperfections and the Chicagoan talk about as-if and the evidence ‘consistent with’ a favored hypothesis seemed silly, even childish.

The causal incident, *si non non*, was an invitation, during what became my last year of twelve at Chicago in the fall of 1979, by the literary critic Wayne Booth—because I suppose of a not altogether justified reputation for being less barbarous than the other economists—to speak to his undergraduate program on Politics, Philosophy, and Economics on ‘the rhetoric of economics’. As I narrate this episode in the autobiography, “I said, ‘Sure, Wayne. But what’s that?’ Oh. He gave me a short reading list, which I crammed during a visit to my father-in-law’s house in Vermont over Christmas of 1979, and gave the talk early in 1980” (McCloskey, forthcoming(a)). Chicago in spring that year refused to promote me to full professor—‘Wait’, they said—and I realized with a jolt that they viewed me as one of the Help. The Barons at Chicago had or were going to have Nobels. Becker, Lucas, Mundell, Heckman. The Help did the work. Harry and Gale Johnson, Gregg Lewis, . . . McCloskey. Oh, no, I said, and decamped to the University of Iowa. There and on long visits to the Australian National University in Canberra and at the Institute for Advanced Study at Princeton, I laboriously, with a great deal of reading such as Dick Rorty’s *Philosophy and the Mirror of Nature* (1979), revised and extended the talk to Wayne’s class into the 1983 article on the rhetoric of economics, and then the book ([1985] 1998).
It makes much more sense to view science as rhetoric, that is, as human persuasion, than as an exercise in epistemology. You see it better. I eventually escaped entirely from the ghost of logical positivism which to this day haunts so many philosophers of economics, and drives the actual economists insane. The three books you mention were the fruit of that in the 1980s and early 1990s. *Rhetoric* said that economists were poets, using metaphors, that is, models. *If You’re So Smart* noted that they were novelists, writing stories—out of a continuum in the human mind having metaphors at one end and metonymies at the other. *Knowledge* then defended these obvious assertions from the indignation of economists and their philosophers at being called poets and novelists. They seemed to worry that admitting that they thought and persuaded like other human beings would make them non-Scientists. Yet all scientists use metaphors and stories, such as quantum mechanics and evolution by natural selection. Meanwhile, I kept up my scientific work on historical topics such as the gold standard and English agricultural history. I have always wanted to be chiefly an economic and historical scientist. But I kept being dragged back into methodology by the evident insanity of my beloved colleagues in economics. I felt it was my duty to try out the clinical psychologist’s talking cure on them, to save the poor dears from the results of their insanity, such as tests of statistical ‘significance’ and the proliferation of imagined ‘imperfections in the market’ based on no scientific evidence.

**Whom do you consider among your masters, and who are the scholars who most influenced you in that first decade or so of the post-Chicago period?**

There were a great many, most of them humanists of one sort or another. I was trying to add a serious understanding of the humanities to my understanding of social science, and took, for example, courses in Latin, Greek, and Italian for the purpose, and co-taught courses at Iowa with rigorous professors of communications and literature and with humanistic professors of physics and economics (Klamer, for example). The chairs and deans hate co-teaching, as inefficient, they think. But it is vastly educational for everyone involved. My heroes among economists at the time were from the previous decades, the 1960s and 1970s, not this decade-and-a-half, 1980–1995, learning the humanities. To list some of the humanistic influencers: that Wayne Booth (in his brilliant *Modern Dogma and the Rhetoric of Assent* of 1974, and in his deeply humane personage). Richard Rorty (very shy, though). Another literary critic. Stanley Fish, of
Is There a text in This Class (1980), who later, astonishingly, became, at the University of Illinois Chicago, my dean. A colleague at Iowa, the political theorist John Nelson, we made together the Project on Rhetoric of Inquiry (POROI, the acronym came from my new study of Greek), which prospers yet. The mainly British sociologists of science such as Harry Collins (1985), Michael Mulkay (1985), Bruno Latour and Steve Woolgar (1979). Among philosophers of science, Paul Feyerabend (1975), who comes as close as one can to a rhetoric of science without saying it. Of course, the historian of science Thomas Kuhn, the father of us all. Yet not The Structure of Scientific Revolutions (1962) but his better book, a collection of detailed studies of the rhetoric of physics, The Essential Tension (1977). Tom would never admit he was doing rhetoric of science, but he was. Likewise, not Lakatos’ somewhat silly sociology-that-doesn’t-know-it-is The Methodology of Scientific Research Programmes (in book form, 1978) but his much better D.Phil. thesis issued in book form also only after his death, Proofs and Refutations (1976), which is rhetoric-that-doesn’t-know-it-is. The great chemist (his son did win the Nobel prize in the field) and classical liberal Michael Polanyi (Karl Polanyi’s smarter brother), especially, The Tacit Dimension (1966). The philosopher Stephen Toulmin, with his philosophy of persuasion (1958).

Notice that they are all men. I was still very much a guy, paying attention as guys do to other guys. After 1995, I started to have heroines, such as Anne Hollander (1994), Carol Gilligan (1982), Deborah Tannen (1990), Philippa Foot (1978), Elizabeth Anscombe (1958), Simone Weil ([1940–1941] 1945), Simone de Beauvoir ([1949] 1953; the two Simones both scored higher on the final exams at the École Normale Supérieure than their contemporary Jean-Paul Sartre, who was eventually the [unfaithful] lover of one of them).

Let’s dwell, for a moment, then, on your ‘Crossing’ (a ‘Midwestern woman from Boston who was once a man’). How did that decision influence your intellectual path and your university career?

Calling it a ‘decision’, note, takes it out of the realm of identity and puts it into the realm of rational choice—which is one of the habits in economic science that humanomics corrects. Of course, as born women can testify, being a woman is not an advantage to a career in economics. Ask the shades of Joan Robinson, and Anna Jacobson Schwartz, and Barbara Bergmann, in life all Nobel-worthy (McCloskey 1998). Barbara was a graduate student at Harvard in the mid-1950s, and she told me once of how her
assertive personality grated on the professors. Ten years later, ‘Donald’ McCloskey’s assertive personality evoked cries of delight from the same professors. Anyway, the transition during 1995–1997, chronicled in Crossing (1999), gave intellectual permission to explore queer studies, novels by women, feminist theory, women’s history, which were extensions of range supplementary to the humanistic decade-and-a-half. Vernon Smith gives the advice to young economists to dig deep in one or another field of economics, but then read very widely outside economics, as he does. I dug deep into Chicago-style price theory and into economic history, and then read widely.

We could also speak of another type of ‘crossing’, interdisciplinary. Your work is extremely broad and you have made major contributions to economic theory, history, methodology, and statistics—in this regard, I’m also thinking about your book with Stephen Ziliak, The Cult of Statistical Significance (2008). You have crossed, or broken through, various disciplinary boundaries. Considering the growing hyper-specialization of knowledge, which is reflected both in the ever higher disciplinary walls and in the disciplinary journals, have you had difficulty publishing in certain journals in recent years?

Oh, sure. Science is conservative, and should be. Economic science should not go running after every alleged novelty, such as the current madness for Modern Monetary Theory, or the lesser madness of obsession with equality when the relief of the wretched of the earth depends mainly on raising the mean, not reducing the variance. Some principles—Kuhn called them ‘normal science’, Lakatos called them ‘the core’—are permanent. \[ MC = MU. \ MV = PT. \] So when a mere economic historian comes along claiming, say, that econometrics is bankrupt in its absurd misuse of \( t \) tests and \( R^2 \)—even if the mere economic historian relies in her claim on scores of the best theoretical and applied statisticians of the past century, and now also on an official report from 2016 of the American Statistical Association (Wasserstein and Lazar 2016)—nonetheless the average journal referee in economics, grossly overtrained in that one quantitative method and ignorant of all the others, is entitled to be a little skeptical. But anyway I do not now knock insistently on the door of refereed journals. I have a hundred refereed articles in my CV, which is perhaps sufficient to establish my scientific credibility, if such a number was relevant to assessing science, which often it is not. The well-known idiocies of the referee system should be replaced by actually reading the article.
Furthermore, I have always felt that it is selfish (I can name the worst egotists) for a senior academic to clog up the refereed journals, especially the ‘top’ ones, considering that the juniors depend on a publication in the *Journal of Political Economy* or *The American Economic Review* to survive. For the rest, over the past couple of decades especially, I have been directing the additions to my CV to conference volumes and invited responses and the like. Here I follow my admired colleague (though the Help) at Chicago, Harry Johnson (1923–1977), a great citizen of the profession, spreading his admittedly somewhat repetitive articles around to the lesser outlets, for their benefit, not his.

Increasingly, too, I have ventured into journalism, especially in defense of liberalism against its numerous enemies. Decades ago, I was in Sweden and found in a university library a bibliography of all the publications of the great Swedish economists of a century ago: Knut Wicksell (who well understood $MC = MU$ and $MV = PT$), Eli Heckscher (also an economic historian), Gustav Cassel, Bertil Ohlin. Startlingly, about every fortnight or so throughout their careers all of them dashed off a piece of popular journalism in aid of liberalism. Deirdre can only follow such a public-spirited example.

And what do you think of the current obsession with rankings and citation counting of journals, scholars, and universities?

It’s disgraceful, corrupting of science and scholarship, a lazy method in line with many others that people like. It’s a piece with the naïve positivism that I gradually escaped from. Reading some of her work is the only intelligent way to assess another scholar or scientist (actually, as a humanist and as a serious student of science I don’t believe in such a dichotomy as ‘scholar’ and ‘scientist’, and in the talk of ‘hard’ and ‘soft’ in science; but never mind). You don’t have to read much of her work to know. Colleagues and chairs and deans will whine, ‘Oh, it takes so much work!’ But that’s silly. You know after reading ten pages of *Foundations of Economic Analysis* (Samuelson [1941] 1947) or *A Theory of the Consumption Function* (Friedman 1957), not to speak of anything by Wayne Booth or Stanley Fish or Richard Rorty, that you are in the presence of a first-rate mind, to be hired or promoted, and listened to. Einstein’s CV was short but epoch-making. So was Ronald Coase’s. At the other end of the spectrum of quality, it takes a paragraph or two to come to the opposite conclusion. (A tip, dears: Never start an abstract with ‘This article . . .’; it’s a reliable signal of incompetence.) The assessing depends on one knowing
the field well and having mature tastes and then doing the reading, which after all is the point of having departments and senior academics making such decisions in the first place.

Yet, in the Netherlands, each academic has in effect a number tattooed on her forehead giving the count of publications. In Britain, each department has it, which results in comically cynical relocating of people with big CVs in advance of every research assessment exercise. As to university rankings, the same. No one ranking 100 universities worldwide has the breadth of experience to make such a judgment. True, I can tell you from experience that the University of Chicago has a higher percentage of first-rate minds than does, say, the University of Illinois at Chicago. But so what? I can also tell you that there are idiots at the University of Chicago (no, I will not name them) and geniuses, such as the philosopher Samuel Fleischacker or again for a while that Stanley Fish, at the University of Illinois Chicago, and Stanley’s wife Jane Tompkins. Whether a student or a faculty member, go find them and make them talk to you. If you listen, really listen, you’ll get educated.

And by the way, it should be your own department that does the reading and then discusses seriously the product and promise of the candidate, not outsiders. You are going to live with her. Kenneth Arrow isn’t going to. Ken’s generous letters touted every successive student he had as the best ever, in monotone increasing series. Letters of recommendation, as I have written in *The Chronicle of Higher Education* (2002), are public opinion surveys with wretched statistical properties. In History, at the University of Iowa, we had the Aydelotte Rule, which was that only people who had actually read the book and therefore were in a position to assess the mind were permitted to speak or vote. You couldn’t fake it. Anyone who has been in a faculty meeting in Economics knows that imposing the Aydelotte Rule would revolutionize scientific work in the field.

*Another turning point and a great achievement in your intellectual journey is certainly the Bourgeois Era trilogy (2006, 2010, 2016b), a sort of Opus Magnum about the ‘Great Enrichment’. It is impossible to summarize the trilogy in a few words, though you recently tried to do so in a ‘lighter’ book (written with Art Carden), Leave Me Alone and I’ll Make You Rich: How the Bourgeois Deal Enriched the World (2020), almost a ‘pop’ version of the trilogy itself, to be read together with Why Liberalism Works: How True Liberal Values Produce a Freer, More Equal, Prosperous World for All (2019a).*
Would it be correct to say that the discovery of liberal thought represented, considering your leftwing youth—well, even an ‘ex-Marxist’—a real spiritual ‘conversion’?

Yes, but not a conversion to opposites, like my conversion from agnostic to Christian. It was a realization that I had always been a liberal, by the European definition—that is, a believer that diversity of thought is essential, such as what comes from liberty of the press and from academic freedom, and that no truth is final except that humans should live as adults liberated from masters whether a husband or a Party or an orthodoxy in economics. A real, Party-line, Leninist like Jean-Paul Sartre believes, as Lenin himself wrote, that “from this Marxist philosophy, which is cast from a single piece of steel, you cannot eliminate one essential part, without departing from objective truth, without falling prey to bourgeois-reactory falsehood” (1909, 326). As I learned more about economics and the economy and actual instead of fairy-tale economic history, I realized that I was not such a Marxist, as thrilling as it was to be one. The songs of the left and labor unions are great, and I sang them in the 1960s to guitar accompaniment with gusto. But as the old joke says, someone who is not a socialist at age 16 has no heart. Someone who is still a socialist at 26 has no brain (I adjust the ages). I realized that I actually was a boring, bourgeois nineteenth-century liberal, who sees more than a single piece of steel. And in my later education as a humanist, I realized the foolishness of the positivistic belief that there is a view from nowhere, in which a number or a word comes supplied with its own interpretation independent of human concerns, Lenin’s objective truth. Look, here is a stone. Yes, God’s objectivity. Granted. But its human meaning can be as a projectile to throw at the Tsarist police, or as a geological sample, or as a decoration along the garden path.

Which liberal thinkers have contributed most to your reflection? And how has your reflection further contributed to the development of liberal ideas?

Any American growing up in the 1950s was supplied tacitly with a host of liberal principles, such as disdain for Senator Joseph McCarthy’s Trump-like authoritarianism, but also a host of statist principles, such as that the US role in the world should be as an anti-Communist policeman. At university, in the early 1960s, I thrilled to Mill’s On Liberty, but took it as completed, routine, done. And I thrilled equally to Zola’s socialist-leaning Germinal. Then the civil rights movement and the anti-Vietnam War
movement, both of which I supported but didn’t do much about, shook my belief in the completeness of US liberalism.

But mainly I came to liberalism through economics, learning in the mid-1960s from John Meyer and Alexander Gerschenkron that economic analysis of a liberal cast actually works as science and policy. Then I spent twelve years at Chicago, my first job of three. Yet the big reveal came in 1974, half-way through Chicago, with Nozick’s *Anarchy, State, and Utopia*. I had read Rawls, and found his sweet statism mildly persuasive, though noting his technical deficiencies, such as assuming without evidence that people behind a veil of ignorance would be ‘maximin’ in utility. Nozick converted me. (He himself in later books drifted away from the neither-left-nor-right position of true liberalism, which floats above the usual spectrum. He spent too much time at Harvard, I reckon.) Note that the influence was not the Austrian economics I later learned, from the late 1980s on. Though Hayek was certainly correct that “in a free society the general good consists principally in the facilitation of the pursuit of unknown individual purposes” (1976, 1), I still clung to social welfare functions and the like. You can read about them in my underground-classic textbook of microeconomics ([1982] 1985). It took me a while to get over Harvard and then to get over Chicago, while keeping the good bits of both.

*Since we are talking about liberalism, it seems high time to explain in what sense you think you are a ‘Christian classical liberal’. And I’d like to understand better how you see the relationship between Christianity and liberalism. Certainly it is an important aspect of your reflection on ethics and economics, virtues and economy, but I suspect it is much more than that.*

Yes, it is more. Last year I wrote an article for a conference mainly of theologians, which came out in January 2021 in a new *Journal of Economics, Theology and Religion*, entitled ‘The Liberty of the Will in Theology Permits the Liberated Markets of Liberalism’ (2021c). That about sums it up. (I have gotten into the habit of naming chapters and articles as declarative sentences making the main point. Alexis de Tocqueville’s *L’Ancien Régime et la Révolution* ([1856] 1955) is the inspiration. I recommend the practice, and his book.) The opening sentence of the theology article is, “There is an intimate, and perhaps desirable, connection between liberty

---

of the human will under Abrahamic theology and the liberty of human action under liberal economic ideology” (McCloskey 2021c, 81). The argument will be elaborated, with the considerations on ethics I articulated in The Bourgeois Virtues (2006), in a book I am writing, God in Mammon: Public Theology for a Commercial Age. The book is critical of the quasi-socialist line of Catholic social teaching and certainly of the frankly socialist beliefs of many progressive Christians (progressive such as on other theological and social matters I myself am).

I think one can also say that, in the trilogy of books, there is another ‘trilogy’: liberty, equality, fraternity. Let’s start with liberty. Yes, you describe yourself as a ‘free-market economist’, but your defense of human freedom does not seem to be reducible to economic freedom, especially since you use the broader concept of human dignity in the very title of the second volume of the trilogy. Your idea of freedom often evokes the Smithian idea of independence, also understood as non-domination, if not non-slavery. It would almost seem an idea of freedom linked to the tradition of republicanism. Other notions of freedom have also been elaborated: positive and negative (I refer to the famous distinction of Isaiah Berlin [1958] 2002), or freedom under the law à la Hayek. Again, you often talk about freedom as creativity, which may perhaps be traced back to Christianity. Are these notions of freedoms all equally important to you, or do you think there is some hierarchy or even incompatibility between them?

The tri-logy, Greek for ‘three words’, of the early French Revolution (first articulated in 1790 by, of all people, Robespierre) was, after all, liberal ideals, in the line of Locke, and Voltaire, and Smith. Liberalism is a throughgoing equality of permission, with the loving fraternity of a republic. True, the tri-logy was also in the line of the autocratic possibilities in Rousseau’s volonté générale, as, for example, the dictatorship of the proletariat. And Rousseau and his evil spawn in European statism came from a France that already had disabilities in carrying out the liberal motto of 1790, in its previous lack of experience in liberty and its habit from Richelieu on of centralization—as Tocqueville showed in L’Ancien Régime. Add these together and liberty is doomed. The French tri-logy was immediately betrayed in the Terror, the Directorate, the Consulate, the Empire. But so is under threat now, obviously in Russia and China, but less obviously in Italy and the US.
Liberalism is precisely non-slavery. It overturns hierarchy ruling you and me, whether by aristocrat or bureaucrat. The very word ‘liberalism’ contains the program. ‘Liberal’ is of course from classical Latin liber, understood by the slave-holding Romans as (in the words of the Oxford Latin Dictionary) “possessing the social and legal status of a free man, free (as opp. to slave)” (Glare and Thompson [1982] 2012, 1125), and then libertas as “the civil status of a free man, freedom (as opp. of slavery or captivity)” (Glare and Thompson [1982] 2012, 1128).

As is so often the case in English, however, there are paired words, the Latinate ‘liberty’ and the Germanic ‘freedom’. The two have relatively recently acquired significantly different connotations, and it is essential to distinguish them if we are not to become muddled. ‘Liberty’ retains the political connotation of people being non-slaves to other humans. ‘Freedom’ in English, though, has increasingly come to mean not being subject, happily, to any constraint at all, by physics or, in particular, by wages. Thus Franklin Roosevelt, in his Four Freedoms speech in 1941, numbered as third a ‘freedom from want’, and Amartya Sen wrote in 1999 of economic ‘development as freedom’. The trouble is that we already have words for such lack of want, or for economic development, namely, income, wealth, riches, capabilities, adequacies, economic development, and lack of want itself. To push together, as the modern English usage of ‘freedom’ does, the politics-idea of non-enslavement to others (liberty) with the wage-idea of ability to buy things from others (income, wealth, capabilities) leads only to self-imposed confusion. I advise dropping entirely the corrupted word ‘freedom’ and always using ‘liberty’. The liberal claim, to be sure, is that liberty does result in an increased ability to buy things—and so it has done spectacularly over the past two centuries. But the claim is that liberty leads to dignity, independence, positive liberty (that is, income, capabilities), liberty of religion, even, I would claim, the rule of law—notice about that last one how indignation against the King or President claiming to be above the law is foundational to a rule of law; without such indignation, Trump wins and the rule of law disintegrates: see the Republican Party in 2021. But for such claims to be meaningful, their truth needs to come from the evidence, not from a mis-definition of development as being freedom, simpliciter. An ideology of equality of permission, considered quite absurd by the ancient hierarchy, and under attack since then by special interests seeking protection or simply some graft, made the modern world. Not equality of outcome or opportunity. Permission. No tyranny.
Speaking of equality, the trilogy was written right during the years when everyone was starting to worry about growing inequalities. Piketty was emblematic. Among other things, you wrote a critical review (2014) of his Capital in this journal. What do you think of the concern, shared by many scholars, pundits, and politicians, about growing inequalities?

I think my beloved colleagues are, to use the Southern US expression from hunting squirrels with a pack of hounds, barking up the wrong tree. The full case is in the Piketty review and in further comments in Why Liberalism Works (2019a). But I can briefly sketch it here. For one thing, inequality has not in sober fact increased. We can agree I suppose that a non-fact is a poor basis for panicked policy. In only three of the countries Piketty studied—the USA, the UK, and Canada—did financial wealth, which is all he studied, become less equal. For the rest he is worrying about the future, based on his structural 'model'. He ignored human capital and public capital and social capital. Piketty aside, the inequality statistics anyway never change much, not on the scale of the astounding rise in the average since 1800, continuing today—that 3,000 percent in the Great Enrichment, which any economic historian can tell you about. After all, there is a bottom 10 percent of any distribution whatever, so talk of 'relative inequality' is meaningless, in the old logical positivist sense of being irrefutable and therefore empty. And recently, of course, world inequality, measured in a liberal way individual-by-individual, has dramatically fallen, because China and India and other poor countries have liberalized some in their economies and become richer.

Science, I’ve said, deepens on starting with proper categories, which is a humanistic, philosophical inquiry. Then you can measure with an assurance that you are doing the science correctly. Measuring inequality in financial wealth is largely a category mistake. Further, equality of income is not an ethically decisive category. Why not equality of height, beauty, parentage, intelligence, cheerfulness, memory, creativity, ability to score goals in football? Only a vulgar Marxist or a Chicago-School economist could think that income covers it all. Are we to achieve equality of intelligence by pounding mails into the heads of the embarrassingly large number of people much smarter than me until they are as stupid as I am?

As the American economist John Bates Clark predicted in 1901:

The typical laborer will increase his wages from one dollar a day to two, from two to four and from four to eight [which was accurate in real terms of per-person income down to the present, though the
calculation does not allow for the radically improved quality of goods and services since 1901]. Such gains will mean infinitely more to him than any possible increase of capital can mean to the rich [...] this very change will bring with it a continual approach to equality of genuine comfort. (Clark [1901] 1949, 79)

That is what we should focus on to achieve a better world, not envious and erroneous calculations of one person getting ahead of another. Using envy as a principle of social policy is ethically obnoxious and will stop growth. It’s happened repeatedly, for example in India before 1991.

Modernity has awakened to the cry of ‘Liberté, égalité, fraternité’, but many believe that fraternity, or brotherhood (or sisterhood!), has been forgotten, even if some continue to associate it with the more contemporary ‘solidarity’. In the first book of the trilogy, Bourgeois Virtues, you provide a detailed account of what you call the seven bourgeois virtues: Love, Faith, Hope, Courage, Temperance, Prudence, and Justice. In an attempt to restore a balance between ‘Prudence Only’ and the other virtues, you group the other six virtues together under the broad label of ‘Solidarity’. Can you explain this thesis and its meaning? And to what extent does your idea of solidarity differ from some contemporary interpretations of solidarity (I am thinking, for example, of those who believe that the welfare state is a form of institutionalized solidarity)?

I am not sure I agree that I use ‘solidarity’ this way. True, I use it 58 times in The Bourgeois Virtues. But half the time it is used to criticize the historical fairy tale that we have lost it in the modern world, and the other half of the time it is used as a combined virtue of love and justice and faith that yields people who want a liberal society in the first place. Solidarity of this sort underlies the sweet claims by, say, James Buchanan or Martha Nussbaum that we can make a liberal constitution, though neither Jim nor Martha ever quite got what is missing in their programs—namely, raising up ethical people to begin with (McCloskey 2011). The book analyzes the seven elemental virtues, which are pagan and Christian, and only have bourgeois versions (consult the one-page sermon that concludes The Bourgeois Virtues). My central claim is that virtues, whether elemental like love or molecular like solidarity, are not reducible to arguments in a utility function characterizing that charmless sociopath Max U. So, for example, love of a sort is a virtue enhanced in a commercial society. Doux commerce, said Montesquieu. In the one-page sermon I say:
Beyond the pagan virtues is the Love to take care of one’s own, yes. But it is also a bourgeois love to care for employees and partners and colleagues and customers and fellow citizens, to wish well of human-kind, to seek God, finding human and transcendent connection in the marketplace in 2006, and in a Scottish benevolence c. 1759. (McCloskey 2006, 508)

Sisterhood has not been forgotten, either. It has flourished in bourgeois society, as I have argued at length elsewhere: the bourgeoisie liberated slaves, but the innovism it also caused was crucial to the liberation of women, too (McCloskey 2000).

You may be referring, Paolo, to the very common claim about modernity that it has led to alienation, anomie, urban isolation, bowling alone, and all. The claim again is factually false. I wish people would get their historical facts right before sounding off about how awful the modern world is. I get tired of pointing out again and again that olden times were miserable. Suffice perhaps for the present to mention that wise phrase in the Communist Manifesto, “the idiocy of rural life”. And of course, “The bourgeoisie, during its rule of scarce one hundred years, has created more massive and more colossal productive forces than have all preceding generations together” (Marx and Engels [1848] 1978, 477).

I think I phrased the question badly, because what you say is what I meant (but of course my interpretation of your ‘solidarity’ may be wrong, and of course I should re-read the first volume of the trilogy!). In truth, I meant to emphasize two aspects of your reflection. On the one hand, your having taken into due account ‘solidarity’, while the vast majority of scholars (regardless of their discipline or ideology), from the past century until today, have mainly been focusing on liberty and equality (or liberty versus equality), forgetting about fraternity (or the more secularized solidarity). On the other hand, it seemed to me that your account of ‘solidarity’ has its own virtue, precisely because many (especially on the left) complain about the loss of solidarity (in society), and can only see such solidarity in ‘institutionalized forms’ such as the welfare state (I’m thinking about Richard Titmuss’ The Gift Relationship and other advocates of the welfare state), while, in my view, the various forms of spontaneous solidarity (that is, third/voluntary sector organizations, cooperatives, mutual assistance associations, social enterprises, or, for that matter, the market) are alive and kicking.
They are often neglected because much of this way of thinking is trapped within state/market dichotomy.

But to continue, the last book of the trilogy, Bourgeois Equality, was, among other things, the subject of a stimulating symposium in this journal, with contributions by Gaus (2016), Goldstone (2016), Baker (2016), Amadae (2016), and Mokyr (2016), and your final thoughts (McCloskey 2016c). Do you have any further (or second) thoughts to add to that debate, also in light of the further reflections and criticisms that the trilogy has continued to arouse in recent years?

Not anything urgent. With slight revisions, I reprint my responses, mostly appreciative but sometimes a little sharp, in Bettering Humanomics, by way of defending what I claim is the ‘killer app’ of humanomics, namely, the demonstration that new ideas (not investment or exploitation) and especially the master new idea of no masters, liberalism, made us rich and pretty good.

Let’s go back to our starting point: Bettering Humanomics. First, I would like to understand better what the similarities and differences are between your approach and Vernon Smith and Bart Wilson’s Humanomics (2019). As for the similarities, for example, both you and they insist on a return to the ‘father’ of humanomics, Adam Smith (and with particular reference to his Theory of Moral Sentiments), and on the key role played by language in understanding human behavior.

Bart Wilson coined the very word humanomics before I used it. Bart and I are starting an annual prize to the person who does the best work in humanomics. We’ll have to give the first prize to Adam Smith! Bart, and I, and Vernon have the same message as Smith did, namely, that economists need to attend to human speech, because human speech—as, for example, in the chat rooms of Bart and Vernon’s experiments on groups of students simulating markets and economies—is scientific evidence of the first order. Bart and I go further, to literary evidence, which has figured heavily in all my work in economics and economic history since the 1980s and especially since the 2000s. I write about this in Bettering Humanomics:

For many years Wilson has taught with Jan Osborn (a colleague from the Department of English at Chapman University in California) a freshman course introducing economics through such texts as an English translation of Goethe’s Faust.

---

5 See also Bart Wilson’s most recent book The Property Species (2020).
Yes, you heard that right, *Faust*. Early in the epic, for instance, the misled Doctor Faust articulates a complaint that illuminatingly violates the no-free-lunch postulate of economics or its related twenty-shilling-note theorem. The theorem is that routine learning, picking up twenty-shilling notes that might perhaps have fallen on the roadside, earns merely routine profits [...].

The Doctor whines, ‘I have neither money nor treasures. / Nor worldly honors of earthly pleasures.’ [...] He therefore turns to magic, or chartist financial advisers, or econometricians, ‘That I might see what secret force / Hides in the world.’ And finally in vexation he turns to Mephistopheles. (McCloskey 2021a, 5–6)

Thus humanomics.

*I believe, however, that there are differences between your and their humanomics approach. For example, they insist a lot on the prediction paradigm, while you, since your first articles on the rhetoric of economics (McCloskey 1983), have denied the methodological value of this paradigm. Is this methodological difference due to your proximity to the Austrian school, which prioritizes ‘understanding’ over prediction?*

No. I know you don’t mean to start pointless quarrels, but there’s no reason to start dividing humanomics up into sub-schools. It’s too small to indulge in the sort of infighting that one saw in the old days among Marxists (see Lenin above), or indeed nowadays in some Austrian circles. The alleged difference you mention is not important. No sensible person disagrees that prediction and explanation are both involved in science: forward prediction in some physics and linguistics, backward explanation in all evolutionary biology and human historiography. Prediction versus explanation is a non-issue, which excites economists only because they do not know how physics actually operates and have read no philosophy or history of science and can’t get Samuelson and Friedman out of their heads. It’s not even a useful *philosophical* ‘problem’.

*I actually agree, following the great liberal Italian economist, journalist, and maker of the post-war Italian constitution, Luigi Einaudi, who fought for considering economics among the humanities, in an epoch of strong positivism and scientism. He opposes this Vilfredo Pareto, who believed that he could treat ‘humans like ants’ (I believe Pareto*
coined the expression) and, in the style of Comte, predict and control their behavior.

With reference to the prediction problem, I have the impression that there is another important difference between your and Smith and Wilson’s approach. Consider the title of the first chapter of your Bettering Humanomics: “Humanomics and Liberty Promise Better Economic Science” (emphasis mine). And consider your claim:

Growing up requires an expanded but modest humanistic science that analyzes the creativity of human action in retrospect and accepts in prospect the epistemological limits on ant-like prediction and control. It’s the humanities in humanomics. (McCloskey 2021a, 7).

If your human creativity coincides, at least in part, with liberty, then it seems to me that you say very clearly that human liberty and ‘prediction and control’ are at odds with each other (well, unless we like to be treated as ‘ants’), right?

Well, sure, but Vernon, and Bart, and Peter Boettke, and Don Boudreaux, and David Boaz, and Jeffrey Tucker, and George Will, and David Brooks, and hundreds of others—may their tribe increase—are humane classical liberals. None of us is enthusiastic about top-down mastery of humans in aid of Comte’s hideous ambition of 1830, which charmed me as a youth, ‘savoir pour prévoir, prévoir pour pouvoir’. I do not want to raise Methodological quarrels among us liberals! We have quite enough trouble dealing with the very numerous amiable statists in economics, such as Daron Acemoglu and Jason Furman, not to speak of the non-amiable statists, such as Trump and Putin.

But you raise a deeper question, which I have recently raised with Pete Boettke in an exchange in the Journal of Austrian Economics (McCloskey 2021b). It is true that mechanical models of humans, as useful as they sometimes are for traffic control and even for tax policy, are always subject to the indeterminacy of language or of Christian liberty of the will. A fancy example is Bob Lucas’ point about an inferred change in regime making econometric estimates misleading. Austrian economics is indeterminate-language and liberty-of-the-will economics, concerning always the veiled future—as economics always should be: ‘structural’ economics is a replay of the erroneous, backward-looking labor theory of value and the erroneous, backward-looking fixed coefficients of input-output tables, the double beloved, again, of my youth, before I finally grasped price theory
by teaching it to Chicago graduate students. I’m trying to persuade Pete to abandon the statism of neo-institutionalism, which denies all this.

*I wish now to reflect on a couple of aspects of the methodology and economic philosophy of your work, in general, and of Bettering Humanomics, in particular. I’d call them the pars destruens and pars construens of your approach. First, the two great obstacles to the realization of a new humanomics, to which you dedicate your critical efforts (also in the forthcoming sister book, forthcoming(c)), are behaviorism and positivism.*

That’s right. They are both top-down, infantilizing, as in nudging, and industrial planning, and other anti-liberalisms. And both are indefensible philosophically. And both are poor guides to understanding the economy.

*As for the pars construens, the aspect that seems decisive to me for a new humanomics, as the word itself indicates, is rethinking and rearticulating the connection between the famous ‘two cultures’. A gigantic undertaking! The humanomics project is ‘your life-long project’. You argue that it is necessary to distinguish between hard sciences and social sciences, and, at the same time, you insist on the claim that “the humanities are scientific” (McCloskey 2016a). This, of course, has to do with your being ‘literary and quantitative’. So, what idea of ‘science’ do you think is compatible with the humanities?*

The hard/soft locution in talking about science is so deep in our minds, as in the two-culture sociology that C. P. Snow noted in 1959, that you, Paolo, think I agree with it! I do not. I say explicitly that some of the humanities are ‘hard’ in any sense you care to specify (the Greek aorist, for example), as are some social sciences (the analysis of kinship terms, for example), and some of biology and physics and geology are ‘soft’. Talk to a physicist about dark matter and dark energy, for example: all his masculinist chatter about hardness will dissolve into embarrassment.

I do believe that it is sometimes helpful, though not in some matters (traffic control, the law of demand), to distinguish between method in the physical sciences (not ‘hard’) and method in the social sciences, for two reasons. For one thing, as the Austrians point out, atoms and genes do not talk to each other. Humans do. Therefore, introspection, among other uniquely human techniques such as opinion surveys and historical analogy, is a perfectly legitimate method in the social sciences and the humanities. For another, the human sciences entail ethics in a way that
chemistry and physics and biology normally do not (though consider Fritz Haber making poison gas; consider the atomic bomb; consider Dr. Mengele). The positive/normative distinction that economics inherited from logical positivism is sharply true at some uselessly high level of abstraction. But at the middle level in which human scientists actually live, both fact and value figure. In economics, for example, liberated bargains among adults have a valence.

Oh, sorry, I think I phrased the question badly (again!). You go well beyond Snow's distinction! (And this is also one of the reasons why I believe that your approach is extremely fascinating, and why I emphasized your being 'literary and quantitative'.) By 'hard' sciences, I meant the positivism and scientism that is to be kept far away from social sciences, thus your pars destruens.

Thinking about the future . . . Humanomics is a better economics and it is also 'bettering'. What other directions will your 'humanomics project' take? What is next on your research agenda?

My next project, on which I am furiously working, is The Prudent and Faithful Peasant: An Essay in Humanomics, a big book going back to my research in the 1970s and 1980s on English open fields and enclosures, 1300–1800. It will come out, I hope from Cambridge University Press, probably early in 2023. The book supplements what I did at Chicago and Iowa, early in my struggle with an inchoate positivism. I told you: I am an economic and historical scientist and a methodologist only out of public spirit, which it would have been profitable not to have. But it arouses my indignation that economists keep on with a bankrupt positivism, behaviorism, and now neo-institutionalism, 'supported' by blackboard theorems and statistical 'significance'. A purely Maxine-U form of Deirdre would not have been so aroused, and would have stuck coolly to her scientific last. It would probably have been a better strategy even for the good of economics, because people react so very badly to being told, correctly, that they are grossly misled. They say, 'Oh, Deirdre, you are so critical! Therefore I can close my ears to the truth of what you say'. It probably would have been wiser just to exhibit the better way, and hope young people got it, and then have them put the science back on track. The other of the pair of books published this year by the University of Chicago Press is the more critical one, Beyond Behaviorism, Positivism, and Neo-Institutionalism in Economics. But I put more hope in The Prudent and Faithful Peasant to persuade people about humanomics.
To conclude, from your experience what would you recommend to young scholars who are starting a university career?

Just what Vernon said: dig deep in one place, becoming a world expert on it (which takes about three years after the BA in most fields), so that you know what depth of understanding really means, but then read extremely widely. My father was a professor of political science. A graduate student who admired him and had completed the first year of the Ph.D. program at Harvard in Government asked for suggestions for reading over the summer, preparing for the second year in which one drifts towards a thesis topic. The student expected some suggestions of advanced readings in political science, perhaps in my father’s field of constitutional law. Instead: ‘Read these twenty great novels’. The student was startled, but he was a good graduate student. Harvard unfairly collects lots of them. Bad students substitute their own necessarily defective judgment for the professor’s correct judgment. The good student obeyed, and became a serious scholar/scientist.

And what books would you suggest they read? Or, to put it from your perspective: imagine you get stuck on an island and you can only choose five books to keep you company, what would you choose?

That’s tough. Francis Bacon (whom in other respects I do not admire, an arrogant aristocrat with silly views about how science should be centralized, convicted in 1625, for example, of selling legal judgments to both sides) wrote wisely:

> Some books are to be tasted, others to be swallowed, and some few to be chewed and digested; that is, some books are to be read only in parts; others to be read, but not curiously, and some few to be read wholly, and with diligence and attention. (Bacon [1597] 1911, 151)

That is, your mother’s rule that if you start a book you should always finish it is quite wrong. Scientists or scholars (as I’ve said, I dispute the distinction) have to read and digest an incredible amount, expanding at 2% per month. They had better, an economist notes, equalize marginal benefit to marginal opportunity cost.

Let me see. I, of course, recommend that a young economist read all the works of that path-breaking economist D. N. McCloskey, such as the collections of her shorter reviews and columns in three volumes in course of publication by the American Institute for Economic Research (McCloskey 2020, forthcoming(d), forthcoming(e)). A little more seriously, I do
recommend that every young economist, or philosopher of economics, read and do the problems in *The Applied Theory of Price*, and in the similar price-theory texts in the Chicago tradition, starting with the elementary book by Paul Hayne, Peter Boettke, and David Prychitko, then moving on to Armen Alchian, Steve Landsburg, George Stigler, Milton Friedman, Gary Becker, and Kevin Murphy. No one is an economist until they understand the theory of price. Then they can criticize it intelligently, as humanomics does. If you don’t do the work, you will fall into the superficial criticisms one sees daily, which have no scientific or philosophical depth, and are regularly simply mistaken, such as hard/soft, positive/normative, number/words, male/female.

And every economist should have read critically and appreciatively in the history of her discipline: Smith (*The Theory of Moral Sentiments* as well as *The Wealth of Nations*), Ricardo, Mill, Marx (you may omit volumes two and three), Walras, Menger, Marshall, Keynes, Mises, Hayek, Samuelson, Friedman.

Beyond that, read this summer twenty great novels—never the latest best-seller, always something like the novels of Jane Austen or Leo Tolstoy that have, as the cliche has it, stood the test of time.

Would you like to add something to explain one or more of the qualifying adjectives you used to describe yourself and that I left aside?

No. You’ve covered it! Though I should mention a long forthcoming autobiographical essay, “*Apologia Pro Vita Sua*: A History of My Economic Opinions” (forthcoming(a)).

Thank you very much, Deirdre, for the time you dedicated to this interview. I am deeply honored.

The honor, carissimo, is mine!

REFERENCES


Fish, Stanley E. 1980. *Is There a Text in This Class? The Authority of Interpretive Communities*. Cambridge, MA: Harvard University Press.


Mathematical Psychology

JAN TINBERGEN

Translated and annotated by:
CONRAD HEILMANN
STEFAN WINTEIN
RUTH HINZ
ERWIN DEKKER
Erasmus University Rotterdam

INTRODUCTION
‘Mathematical psychics’ was the name of the approach and the book by Edgeworth for a burgeoning scientific approach, also pioneered by Pareto, for that part of psychology on which economics rests. The nature of the subject of this approach raises the prospect that this approach can also be of interest to practitioners of other sciences related to psychology, which is why an attempt is made here to give an overview of the contents of this approach and some results already achieved with it in economics. In addition, some problems outside economics, narrowly construed, are indicated, for the solution of which one might also make fruitful use of mathematical psychology.

TRANSLATORS’ NOTE: This article appeared originally in 1930, in Dutch, under the title “Mathematiese Psychologie” in Mens en Maatschappij, see Tinbergen (1930). We would like to thank the journal Mens en Maatschappij and the Tinbergen estate for their permission to publish the translation. We gratefully acknowledge funding from the research project “Jan Tinbergen: The Thinker”, supported by Stichting Erasmus Trustfunds, Erasmus Initiative “Dynamics of Inclusive Prosperity”, Erasmus School of Economics, Erasmus School of Philosophy, and Erasmus Institute for Philosophy and Economics (EIPE). An accompanying article by Conrad Heilmann and Stefan Wintein that contextualizes the translation appears in the same issue of this journal, see Heilmann and Wintein (2021).

Footnotes with roman numbering appear in the original 1930 article. Annotations by the translators are preceeded by 'Translators’ Note'.
I. OUTLINE OF THE MATHEMATICAL-PSYCHOLOGICAL APPROACH

The essence of this approach is that the 'psyche' of an individual can be described by 'functions of equal ophelimity'. The meaning or intent of this can best be explained by a simple example. The degree of satisfaction of an individual will depend, among other things, on his possession of certain goods. Suppose, for instance, that only two goods $X$ and $Y$ are of interest to the individual. Every possible state is now characterised by the quantities of $x$ and $y$ the individual possesses of those goods. Each such state can be represented graphically by a point (Figure 1), such that $OP_1 = x$ and $PP_1 = y$ ($x$ and $y$ being coordinates). A priori, all states are possible that are represented by points within $YOX$. The curved lines connecting all those points representing states which give the individual the same satisfaction, called ophelimity curves by Pareto, are important for the behaviour of the individual.

So, when the individual is indifferent whether it is in $P_1$ or $P_2$ (that is, possessing either $x_1$ and $y_1$ or $x_2$ and $y_2$), then $P_1$ and $P_2$ lie on the same ophelimity curve. Such curves are, as one can easily see, manifold: there are infinitely many of them. In case both goods are positively valued, the lines will proceed as shown in Figure 2, and the higher lines will then indicate states of greater satisfaction than the lower ones; for it is obvious that, for example, proceeding along the line $P_1P$, one arrives at states where, besides a constant quantity of $x$, an increasing quantity of $y$ is possessed. Only one curve goes through each point, because the individual cannot experience two degrees of satisfaction at the same time.

For Pareto, the ophelimity curves serve as a substitute for the utility functions used by other authors. The great advantage of these curves is that they can be experimentally determined and are devoid of any metaphysics. Indeed, one can determine through a survey which combinations of two goods are equally valued by a certain individual. Starting from a certain combination, one can ask, what increase of $x$ can compensate the loss of a unit $Y$. If these are the quantities $x_1, x_2, x_3, x_4, x_5$, respectively, then $P_0 \ldots P_5$ is the sought-after curve of constant ophelimity (Figure 3). Of course, these amounts can only be determined within certain limits—thus relatively vaguely—so that one will end up with an area of possibilities instead of lines. But it is clear that the same variation will

---

* Translators’ Note: Tinbergen uses the Paretian term ‘ophelimity’ for utility. Functions of equal utility later became known as indifference curves.

1 Cf. the very fine expositions in Bowley’s The Mathematical Groundwork of Economics (1924); Pareto’s Cours d’Economie Politique (1896–1897); Lenoir’s Études sur la Formation et le Mouvement des Prix (1913).
then also occur in the actual actions of people, so that no objection to the
depiction of reality lies here.

Thus, one does not ask the question, which in Pareto’s opinion cannot
be answered, of how large the value is (the degree of satisfaction, the
utility) of a certain combination of goods for the individual. And where
it is possible to deduce from these curves the exchanges of goods, which
are the foundation of the whole economy of our time, this seems to me a
great methodological advantage.

What has been explained here graphically can also be explained alge-
braically, which is, for cases involving more goods and other parameters
that determine the state of the individuals, the most appropriate way.
Such a system of lines as described above can be approached by a for-
mula of the form

\[ w(x, y) = c \]

where \( w(x, y) \) is some function of \( x \) and \( y \), while \( c \) has a value which is
constant for all points of one line, but different for different lines. Here,
\( w \) is only determined in such a way that any function of \( w^2 \) may be sub-
stituted for it, that is, the above equation may be replaced by

\[ F(w(x, y)) = F(c). \]

When dealing with several goods, \( w \) depends on several variables, for
example, \( w(x, y, z, t) \). This is impossible to represent graphically. For
the benefit of the non-mathematically inclined reader, we shall confine
ourselves to examples with two variables.

---

\[ ^2 \text{Which fulfils the condition to be not only finite and continuous but also monotonous.} \]
\[ \text{For such details, please refer to the literature referenced in footnote 1.} \]
II. RESULTS IN ECONOMICS

From the above diagram of the ‘psyche’ of economic actors, it is possible to form an idea of their actions in certain circumstances. From the many theoretical-economical inferences for which this scheme allows, we choose a few cases to illustrate.

II.I. Exchange at a Given Price

In economics, exchange means the exchange of property with the consent of each of the relevant individuals—the homines oeconomici.

Assume an individual is given the opportunity to exchange, at a given price $p$, as much $Y$ for $X$ as it chooses. The question is, how much will this individual exchange, when we know how much it originally possesses and when his ophelimity curve is given. This problem I call the problem of exchange at a given price.

Figure 4 illustrates this case, in which the initial state is indicated by the point $P_0$ (the individual is endowed with $x_0$, $y_0$). The opportunity to exchange at a given price $p$ can be represented graphically by allowing the individual to move its property combination anywhere along the line $P_0Q$, which is drawn so that $p = RQ/P_0R$. The individual will now choose, among the given possibilities, the one that yields the most satisfaction. And that is the situation, denoted by $P$, where the line $P_0Q$ intersects the highest ophelimity curve: every other point of $P_0Q$ lies on a lower ophelimity curve. The quantity $x$ (see Figure 4) which the individual will want to exchange is called in economics the demand at this price (or, if negative, the supply).

With the help of the above, one can now construct a theory of exchanges between many individuals among themselves, each of which considers the price to be fixed. At any conceivable price, which could prevail in such a ‘market’, the supply and demand of each of the individuals is
fully determined, and the actual (‘equilibrium’) price will now be such that total demand and total supply are equal.

II.II. Isolated Exchange

By isolated exchange we understand an exchange where there is one individual on each side, in contrast to the exchange at fixed price discussed above. In the case of isolated exchange, there is, as a matter of principle, no competition. There is no price which each of the individuals considers given. It now appears that the method of mathematical psychology clearly demonstrates what the most important peculiarity of this exchange is: without a further organising principle this exchange has no fixed outcome; through economic forces alone no equilibrium is established. Figure 5 illustrates this. The possibilities in this case can be represented very clearly in the manner shown in this figure: the rectangle $O_1 AO_2 B$ has as sides, respectively, $O_1 A =$ the sum total of property $X$ owned by both individuals and $O_1 B =$ the sum total of property $Y$ owned by both individuals. All possible distributions of that total property—that is, all possibilities after the exchange—can now be represented by points within this rectangle. An arbitrary point $R$ therein may, by its distance from $O_1 B, O_1 A, O_2 A, \text{ and } O_2 B$, respectively, indicate the possession of $X$ and $Y$ by individual 1 and 2 ($x_1, y_1, x_2, y_2$). In this rectangle, we can now draw the ophelimity curves: for individual 1, relative to the axes $O_1 A$ and $O_1 B$ as $x$- and $y$-axes, and for individual 2, relative to $O_2 B$ and $O_2 A$ as $x$- and $y$-axes (for example, $m$ and $n$). If the rectangle is thus covered with a net of ophelimity curves (Figure 6), such that through each point there pass two, one from each individual, it is possible to deduce for each change what either of the individuals think about it. Moving from $P_1$ to $P_2$ will be gladly accepted by individual 2, as he moves to a higher ophelimity curve. On the other hand, 1 will not desire making this step. Since an exchange
always requires the consent of both parties, from a given starting point $P_0$ (Figure 5), only those points which are situated between the ophelimity curves of both parties, $m$ and $n$ (as they are drawn through $P_0$), are possible as endpoints of the exchange. The endpoint can thus only be in the lens-shaped area $P_0mqn$. We can specify the endpoint a little further. For it shall not be possible that either individual can still gain an advantage. This means, as can easily be seen, that from the endpoint one cannot draw anew a lens-shaped area, as was the case from $P_0$. As a consequence, only points where the ophelimity curves of both individuals meet, as in point $S$, Figure 6, can serve as endpoints. There is a whole series of such points, forming a line, which is usually called a contract curve ($K_1K_2$ in Figure 5 and Figure 6).

Therefore, all points on the line $K_1K_2$, in so far as they lie within the lens $P_0mqn$, can function as end points of the exchange. Without further information, it is not possible to determine which of these points is preferable. Thus, the exchange result is indeterminate within certain limits.\(^3\)

A number of similar issues, in which the consequences of a certain market organisation for the determination and level of the price are examined, find a clear formulation in the mathematical-psychological way of thinking. One thinks here of cases like imperfect competition.

III. MATHEMATICAL PSYCHOLOGY AS A STARTING POINT FOR MORE CONCRETE RESEARCH

I would now like to indicate a few directions in which, to my mind, expansion of the mathematical-psychological approach could be of interest.

In the first place, it is clear that there is much room for experimental research. Among other things, attention will have to be paid not so much to the ophelimity curves for individual goods as to those for groups of goods (food, clothing, housing, recreation, study) against each other. Apart from the actual shape of the ophelimity curves, which can be expressed numerically, it is very interesting to ask whether the ophelimity curves differ greatly between persons, and to what extent these differences are due to characteristics that can be easily identified (illness, profession). Here lies an area of psychology which, if I am well informed, is still open to considerable expansion.

\(^3\) The importance of these questions is especially pointed out by Schumpeter in his study “The Instability of Capitalism” (1928); it is true that Schumpeter arrives at a different conclusion as far as the question raised here is concerned; however, his clear statement of the problem is only possible through the use of the mathematical-psychological method.
Certainly, some such research already exists. I am thinking here of the meritorious budget studies of various municipal statistics offices. However, in their usual form, these do not provide enough data to learn about the ophelimity curves.⁴

Secondly, there is an extensive possibility to further accentuate the considerations given in section II and similar ones, by working with special cases of $w$. One of the most important forms for $w$ is: the profit of an enterprise.

IV. QUESTIONS OF JUSTICE; FORMULATION OF A CRITERION OF ‘JUST DISTRIBUTION’

In addition to these purely economic applications, the mathematical-psychological method seems to be useful for the analysis of the common conception of justice with regard to the rules governing the distribution of economic goods.⁵ This analysis, which concerns one of the oldest societal problems, has recently come to the fore again, often even in a very practical, quantitative form, for example, because of the increasing organisation of society, as a result of which distribution through free competition is no longer possible and the need for a separate distribution principle has come to the fore.

The determination of ‘just distributions’, for example, the distribution of the total production among all individuals, must of course be preceded by a definition of justice. As far as I can see, the common conception of justice⁴ entails the equal treatment of equally situated individuals. This criterion then provides a solution, for example, in the following simple case:

A) Two identical workers perform the same amount of labour of a certain kind. Which distribution of the product can be considered just? The criterion mentioned above gives the solution of equal distribution: equal

---

⁴ Translators’ Note: Tinbergen uses *het rechtsbewustzijn*, which alludes to a (shared) sense of justice; it is not clear from the text whether he believes it is shared by all.

⁵ They give the quantities bought of different commodity groups in relation to income. From some of these investigations it can be concluded that a quadratic form of $w$ gives a sufficient approximation; of the $n^2$ unknown coefficients ($n$ being the number of commodity groups) only $n$ combinations can be calculated.

At this point, it is worth mentioning the interesting idea of Irving Fisher (1927), developed in the Festschrift for Clark, to determine the marginal value of money in relation to income in such a way.

⁵ These problems cannot be considered part of pure economics today; according to Sombart’s terminology in *Die Drei Nationalökonomien* (1930, 295), they belong partly to “economic ethics” (“Wirtschaftsethik”), partly to “the art of economics” (“Wirtschaftskunstlehre”). The following problems and solutions were developed in very animated discussions with Prof. Ehrenfest.
wages. However, it cannot provide a solution in problems involving individuals in unequal circumstances, as, for example, when one extends the above case A in one or more of the following ways:

1) the quantities of work are unequal;
2) the workers have different ophelimity curves;
3) the type of work is different.

As soon as one of these cases arises, in which the circumstances (in the broadest sense) of the individuals are different, the possibility of applying the principle of equal treatment is no longer there. The question is, how to extend the criterion of just distribution to these cases. In my opinion, the requirement that the individuals should not have the desire to occupy each other’s position (in the broadest sense) should be adopted as such. We postulate this answer to the above question without any further justification other than that the first-mentioned ‘simple justice criterion’ must be included as a special case, which is indeed the case. Thus, we leave a discussion of the correctness of our concept of justice to the relevant science (in the broader sense of Lehre†).

The more general criterion we formulated can be represented very conveniently in the language of mathematical psychology. Is the position (in the broadest sense) of individual 1 indicated by the quantities \(x_1, y_1, z_1\) (in our example, we denote with \(x\) the amount of work type 1, \(y\) amount of work type 2, and with \(z\) wages) and by the function \(w_1\), and likewise the position of individual 2 by \(x_2, y_2, z_2\), and \(w_2\), then it reads:

\[
\begin{align*}
  w_1(x_2, y_2, z_2) & \leq w_1(x_1, y_1, z_1) \\
  w_2(x_1, y_1, z_1) & \leq w_2(x_2, y_2, z_2)
\end{align*}
\]

For the first equation expresses that the circumstances of individual 2 do not give greater satisfaction to 1 than his own circumstances; the second that the circumstances of individual 1 do not give greater satisfaction to 2 than his own position.

For cases in which there are only two dimensions, such as when our case A is only expanded according to 1 or 2, one can again use the graphical representation.

† Translators’ Note: This is most likely a reference to the art of economics as referenced in footnote 5.
6 An open question for me here is whether it is sufficient for the psychological ‘identity criterion’ to hold for two people to be equal in terms of their ophelimity curves.
A1) So here $w_1$ and $w_2$ are the same functions, while $x_1$ and $x_2$ are unequal; our criterion becomes:

$$w(x_2, z_2) \leq w(x_1, z_1) \quad w(x_1, z_1) \leq w(x_2, z_2)$$

which can only be satisfied simultaneously by the equality signs; the meaning of this is, that $z_1$ and $z_2$ will be such that the points $x_1z_1$ and $x_2z_2$ lie on one ophelimity curve. Thus $z_2$ is determined by $x_1$, $x_2$, and $z_1$ (Figure 7).

A2) The workers do not have the same ophelimity curves. If, however, they do the same amount of work, the 'just distribution' will still be equal pay; if, however, also the amounts of work are different, our criterion is:

$$w_1(x_2, z_2) \leq w_1(x_1, z_1) \quad w_2(x_1, z_1) \leq w_2(x_2, z_2)$$

The two points $x_1z_1$ and $x_2z_2$ must now be situated such that $x_2z_2$ lies below the ophelimity curve of individual 1 through $x_1z_1$ (the line in Figure 8) and at the same time $x_1z_1$ lies below the ophelimity curve of individual 2 through $x_2z_2$ (the dotted line).

The relevant difference with case A1 is that, for given $x_1$, $z_1$, and, for example, $x_2$, $z_2$ still has a certain degree of freedom, or that, for given $x_1$ and $z_1$, $x_2$ and $z_2$, the point $x_2z_2$ can still lie in the shaded area.\(^8\)

The examples provide sufficient clarity of the criterion of justice posited here.

It seems to me that this train of thought can be extended in two directions. Firstly, under what conditions does a ‘just distribution’ automati-

\(^8\) Translators’ Note: For a more extensive description of Figure 8, see Heilmann and Wintein (2021, 230).
cally come about, and secondly, when should one consciously strive for such a distribution.

V. AUTOMATIC ACHIEVEMENT OF ‘JUST DISTRIBUTIONS’; IMPEDIMENTS AGAINST AUTOMATIC ACHIEVEMENT

According to our criterion, a just distribution exists when any individual can indeed, if so desired, reach a state (in the broadest sense) equal to that of any other individual: when there is ‘free mobility’ in all respects. With respect to wages and labour performance, this is the case, for instance, with free competition, infinitely easy mobility, and ‘retraining’, etc. Amongst workers, an approximation of justice is then achieved.

Where, however, the aforementioned freedom of movement, taken as broadly as possible, is lacking, it is not guaranteed that a ‘just distribution’ is reached. This freedom of movement can be absent in very many ways, due to very different causes. We mention some examples:

1. Levelling the living conditions of people in different geographical areas is still mostly impossible due to the lack of free migration (legal, physical, and psychological barriers). Many and large injustices still exist in this area: for example, the difference in real wages for the same work in the United States, the Netherlands, Germany, Austria, Russia, and China.

2. A person cannot, to a large extent, exchange his own state of health with that of others. Great injustice still exists, though to a much lesser degree than before, between the sick and the healthy. Something similar existed and still exists with regard to the size of the family.

3. Freedom of movement from one profession to another is significantly impeded for all skilled professions by the duration and costs of education. This duration and costs are an absolute impediment for important groups of people to learn certain trades. In this respect, the greater or lesser degree of aptitude, which in many cases has an important influence, does not, in my opinion, justify a higher income, just as poor health cannot justify a lower income.

4. Another impediment to freedom of movement is the inequality of property, as a result of the private ownership of the means of production and inheritance law. In the opinion of many, these injustices lie at the heart of the class contradictions of our time.

As the Liberal Socialists, led by Oppenheimer, demanded in contrast to the other Socialist reformers.

If an individual once studies a trade and is disappointed with it later on (here, however, another element plays a role, that is, not knowing one’s own ophelimity function properly or changing it), this too can lead to an unjust distribution later on.
5. An interesting cause of reduced freedom of movement is also the economic organisation of society, such as on the one hand the formation of trusts etc., and on the other hand trade unions. The peculiar seclusion of certain trades, which occurs more and more and of which the final and most serious consequence is unemployment, also opens the possibility of important injustices in the distribution of the incomes.

Amongst the obstacles to free movement mentioned, there are some that could be eliminated by human effort if desired; there are others, such as state of health and ability, that certainly cannot be eliminated. In such cases, therefore, it makes sense to strive consciously and directly for a just distribution of 'benefits and burdens'. This is done through various social institutions and laws. However, the quantitative yardsticks applied are for the time being rather arbitrary and, consequently, the opinions of different people on these numbers fluctuate widely. Examples of quantitative indeterminacy are sickness benefits, child allowances, accident benefits, unemployment benefits (on which the views vary between full pay and nil), holiday allowances, and widows' and orphans' pensions. In my opinion, the method of mathematical psychology outlined above can contribute to the solution of these contested issues.

The method described above may also be useful in cases where freedom of movement could be restored through human action but where this is not considered desirable for other reasons, such as unjust distribution as a result of the increasing organisation of society.

For other, larger questions, such as the question whether inheritance laws are in accordance with the common conception of justice, one will first have to consider, as is done in legal science, the complications of the questions mentioned, such as in this case, for example, the question to what extent one should consider a person as one whole with his ancestors or as a completely independent individual. Difficulties of this kind are those arising from the changeability of \( w \), whereby a distribution originally in accordance with our criterion of justice may later no longer be in accordance with it. Another, very interesting set of questions, which, however, seems to me for the time being to be only of academic interest, concerns the question: to what extent does exchange make an originally just distribution of goods unjust or vice versa?

However, it would take us too far to go into the issues mentioned in this last paragraph.

**REFERENCES**
Conrad Heilmann is associate professor of philosophy and co-director of the Erasmus Institute for Philosophy and Economics (EIPE) in the Erasmus School of Philosophy and Core Faculty of the Erasmus Initiative “Dynamics of Inclusive Prosperity” at Erasmus University Rotterdam. He is the co-editor (with Julian Reiss) of The Routledge Handbook of the Philosophy of Economics.

Contact e-mail: <heilmann@esphil.eur.nl>

Stefan Wintein is assistant professor of theoretical philosophy at the Erasmus School of Philosophy (ESPhil) and member of the Erasmus Institute for Philosophy and Economics (EIPE), both at Erasmus University Rotterdam in the Netherlands. Together with Conrad Heilmann, he runs the Fairness project: http://fairness-research.org.

Contact e-mail: <wintein@esphil.eur.nl>

Ruth Hinz is a research master’s student in philosophy and economics at EIPE, Erasmus University Rotterdam. Her research interest focuses on fairness; in particular, on fairness in questions related to social choice theory.

Contact e-mail: <hinz@esphil.eur.nl>

Erwin Dekker (1984) is assistant professor of Cultural Economics at the Erasmus University Rotterdam and has been post-doctoral fellow at the Mercatus Center at George Mason University. He has recently published Jan Tinbergen (1903–1994) and the Rise of Economic Expertise (2021) and previously The Viennese Students of Civilization: The Meaning and Context of Austrian Economics Reconsidered (2016), both with Cambridge University Press.

Contact e-mail: <e.dekker@eshcc.eur.nl>
No Envy: Jan Tinbergen on Fairness

CONRAD HEILMANN
Erasmus University Rotterdam

STEFAN WINTEIN
Erasmus University Rotterdam

Abstract: The important ‘no-envy’ fairness criterion has typically been attributed to Foley (1967) and sometimes to Tinbergen (1946, 1953). We reveal that Jan Tinbergen introduced ‘no-envy’ as a fairness criterion in his article “Mathematische Psychologie” published in 1930 in the Dutch journal Mens en Maatschappij and translated as “Mathematical Psychology” in 2021 in the Erasmus Journal for Philosophy and Economics. Our article accompanies the translation: we introduce Tinbergen’s 1930 formulation of the ‘no-envy’ criterion, compare it to other formulations, and comment on its significance for the fairness literature in philosophy and economics.

Keywords: no envy, subjective equity, equity, fair, fairness, fair division

JEL Classification: B21, D63

I. INTRODUCTION

The economist Jan Tinbergen—along with Ragnar Frisch, the recipient of the inaugural 1969 Nobel prize in economics—is not only widely regarded as an econometrician and engaged scientific expert, but also as a theorist of redistribution.\(^1\) In this article, we reveal that Jan Tinbergen already

\(^1\) See Dekker (2021) who develops these themes, in particular the topic of expertise, in the first intellectual biography of Jan Tinbergen.

AUTHORS’ NOTE: We thank Erwin Dekker for bringing Jan Tinbergen’s (1930) article to our attention as well as for his invaluable comments on earlier drafts of this article. We thank Ruth Hinz for her excellent research assistance. We also thank Måns Abrahamson, Roger Backhouse, Annalisa Costella, Kristi Olson, William Peden, anonymous reviewers of EJPE, and the editors of EJPE for many important suggestions and comments. We gratefully acknowledge funding from the research project “Jan Tinbergen: The Thinker”, supported by Stichting Erasmus Trustfunds, the Erasmus Initiative “Dynamics of Inclusive Prosperity”, Erasmus School of Economics, Erasmus School of Philosophy, and Erasmus Institute for Philosophy and Economics (EIPE).
introduced the ‘no-envy’ criterion in print in 1930, compare it to other formulations, and comment on the significance of his article for the fairness literature in philosophy and economics.

The ‘no-envy’ criterion has been a cornerstone of many influential theories of fairness for a few decades now (Thomson 2011). It characterizes a state of affairs as ‘envy-free’ if no individual prefers another individual’s allotment to their own. Arnsperger (1994, 155) characterizes it as a fundamental criterion of distributive justice that is “absolutely central” for economic theory. So far, the ‘no-envy’ criterion has typically been attributed to three sources: firstly, and most commonly, Foley’s (1967) article is cited. Secondly, the two editions of Jan Tinbergen’s book Redelijke Inkomensverdeling (1946, 1953) are sometimes also cited alongside it. Thirdly, there are occasional references to conversations between Jan Tinbergen and his supervisor, Paul Ehrenfest, in the early 1920s (such as in Kolm [1971] 1997; Young 1994; Lambert 2012; Olson 2018). In this article, we demonstrate that an additional attribution of the ‘no-envy’ principle to Jan Tinbergen is in order: the ‘no-envy’ principle appears in his 1930 article “Mathematiese Psychologie” (Mathematical Psychology), published in the Dutch journal Mens en Maatschappij and translated in 2021 in the Erasmus Journal for Philosophy and Economics (Tinbergen [1930] 2021). There are thus not three, but at least four attributions relevant for the origins of the modern ‘no-envy’ principle. As far as we know, the 1930 article is the first appearance of the ‘no-envy’ criterion in print.

Tinbergen’s 1930 article discusses the essential ingredients of early welfare economics as developed by Edgeworth and Pareto before presenting the ‘no-envy’ criterion. While perhaps of historical interest in the context of welfare economics, we leave aside a more detailed summary and discussion of these elements of the article, in order to focus on Tinbergen’s introduction of the ‘no-envy’ criterion. We discuss similarities and differences between how Tinbergen discusses the ‘no-envy’ principle in his 1930 article and his 1946/1953 book. We also compare it to Foley’s (1967) formulation, as the latter has so far been the most prominent source of ‘no-envy’. We will show that while all three sources present essentially the same ‘no-envy’ criterion, the issue of what the criterion is applied to is more ambiguous. On the one hand, both authors speak of the criterion being applied to an individual’s ‘place’ or ‘position’ within society. On the other hand, they also propose different precise objects to which the ‘no-envy’ criterion applies: Tinbergen (1930) considered hours worked and wages earned and Tinbergen (1946, 1953) considered income distribution (on the basis of hours worked and wages earned), whereas Foley (1967)
considered consumption bundles. Interestingly, Tinbergen’s early formulation of the ‘no-envy’ principle is in some ways more concrete than the later discussions by both himself and Foley. In addition to our discussion of Tinbergen and Foley on ‘no-envy’, we also briefly comment on more recent contributions to the ‘no-envy’ literature, and we make two general observations about the relations between Tinbergen’s contributions and conceptual discussions of fairness.

We proceed as follows. Section II introduces the ‘no-envy’ criterion and presents Tinbergen’s (1930) and (1946, 1953) formulations of it. Section III compares Tinbergen with Foley (1967) and discusses both in relation to other contributions in the ‘no-envy’ literature. Section IV makes the two general observations on conceptual discussions of fairness. Section V concludes.

II. INTRODUCING ‘NO-ENVY’

II.1. The Origins of the Modern ‘No-Env’ Criterion

If Alice does not prefer Bob’s allotment to her own, and Bob does not prefer Alice’s allotment to his own, the situation they find themselves in is ‘envy-free’. The idea of no individual preferring another individual’s allotment over theirs is typically referred to as the criterion of ‘no-envy’ or ‘envy-freeness’. This criterion is used in two different, but closely related, ways. Firstly, the ‘no-envy’ criterion is used to favour some allocations, distributions, or divisions of goods, or indeed whole states of social affairs over others, as they are ‘fairer’ (Arnsperger 1994; Thomson 2011). Secondly, many theorists see an absence of envy between individuals as important for, if not defining of, the concept of fairness. They explore in how far the ‘no-envy’ criterion is a constitutive aspect of the concept of fairness in general (Olson 2018; Fleurbaey 2021).

Philosophers and economists alike emphasize that ‘no-envy’ is supposed to be a technical concept that is independent of sentiment (Kolm 1996; D’Arms 2017; Olson 2018). Envy, as encapsulated in the ‘no-envy’ criterion, does not point to the presence of a feeling of resentment or jealousy, or to any psychological fact about the individuals involved that goes beyond the following: whether they prefer what the other has over what they have themselves. In short, ‘no-envy’ is just a commonly agreed upon label for a state of affairs in which individuals do not prefer to have what other individuals have. It is independent of whether individuals do actually experience any envy. Preference-based notions of ‘no-envy’ have

2 For a more detailed defense of the ‘impersonal’ nature of the preference-based ‘no-envy’ concept, see Olson (2018).
become key for many theories of fairness, reviewed in Thomson (2011) and Fleurbaey (2021, section 6).

While fairness theories that are based on ‘no-envy’ do not rely on psychological content beyond preferences, they are nevertheless to be classified as subjective fairness theories. They locate the meaning of fairness in the subjective judgement of the individuals who are concerned with the social state of affairs in question. In this sense, ‘no-envy’ based fairness theories are thus subjective, even if they do not rely on a broader psychological characterisation of the individuals concerned. Such subjective theorizing about fairness is to be distinguished from objective (or non-subjective) fairness theories in philosophy and economics. To appreciate the difference between subjective and objective fairness theories, consider the following example.

Owing Money: Jan borrowed 10 Euros from Duncan and 20 Euros from Serge, but he has only 15 Euros left. How should Jan divide his 15 Euros?

No doubt, both Duncan and Serge prefer more money to less, so that envy-freeness recommends that Jan allots them equal amounts. Now, while Duncan and Serge may have the same (subjective) preferences, they have different (objective) claims: Duncan has a claim to 10, and Serge to 20 Euros, in virtue of the promise of Jan to give back the money. According to objective theories, fairness recommends the proportional satisfaction of claims so that, in Owing Money, fairness requires that Serge receive twice as much as Duncan. Although Owing Money is more naturally analysed along the lines of an objective theory of fairness, the division problems studied by welfare economics typically do not involve agents with claims, but with preferences. For these problems, then, criteria of fair division have to be formulated in terms of preferences.

Preference-based ‘no-envy’ theories of fairness have been very influential. Indeed, there are now various strands of fairness-related literatures in philosophy and economics in which different variants of the so-called

---

3 The longstanding methodological debate about what kind of (if any) psychological content is implied by or should be attributed to the concept of preference can be safely left aside here.

4 In philosophy, objective theories of fairness are mainly based on Broome (1990) who characterizes fairness as proportional satisfaction of claims, which is not a subjective notion (see also Rescher 2002). In economics, objective theories of fairness are put forward in the axiomatic and game-theoretic literature of fair division problems (reviewed in Thomson 2015). For a more comprehensive taxonomy and review of fairness theories in philosophy and economics, see Wintein and Heilmann (forthcoming).
‘no-envy’ principle play an important role, most prominently in: the literature in economics (Young 1994; Kolm 1996; Thomson 2011), including the algorithmic literature (Brams and Taylor 1996), various strands of normative economics (Moulin 2003; Fleurbaey 2008; Fleurbaey and Maniquet 2011), and political philosophy, including egalitarianism (Olson 2020; Fleurbaey 2021).

Given the importance of the ‘no-envy’ criterion for philosophy and economics alike, it is interesting that its modern origins have been a matter of some ambiguity. Three attributions have become the norm in the literature: firstly, and most commonly, the ‘no-envy’ principle is attributed to Foley (1967). Secondly, references to either one of the two editions of Jan Tinbergen’s book Redelijke Inkomensverdeling (1946, 1953) are given. Most prominently, Thomson’s (2011, 395) extensive review notes that: “It is only in 1967 that an ordinal equity criterion designed for the evaluation of choices in concretely specified resource allocation models was first proposed: ‘no-envy’”.

However, in the same review, Thomson remarks in a footnote:

The idea had been formulated by at least one previous writer. Tinbergen (1953) devotes a few pages to a discussion of the no-envy test, explaining that he had developed it in conversations with the Dutch physicist Ehrenfest. However, it is thanks to Foley that the criterion has become known, and this author is usually credited with it. (Thomson 2011, 402n6)

Peyton Young, in his seminal 1994 book Equity: In Theory and Practice, also acknowledges that Tinbergen was the first to advance the modern ‘no-envy’ principle, when he writes that it was “first proposed in a very strong form by Tinbergen (1953)” (Young 1994, 11). There is also a third attribution, an anecdotal one, referring to conversations between Jan Tinbergen and his supervisor, Paul Ehrenfest, in the early 1920s. In an interview, published by Lambert (2012), Serge-Christophe Kolm recounts the following:

Here is an interesting story. In 1924 in Holland there was a student of physics who was 18, a young man named Jan Tinbergen interested in justice. He asked his professor, Paul Ehrenfest, [...] ‘What do you think a just wage is?’ Ehrenfest thought a little and said ‘Well, if I

---

5 Olson (2018) cites Denyer (2013) who discusses the earlier history of envy as a concept. For a comprehensive review of the various roles of envy in philosophy, including those beyond fairness, see also D’Arms (2017).

6 We thank Kristi Olson for bringing this interview to our attention.
prefer my wage and the job I do, to your wage and the job you do, and conversely, this is kind of fair.’ Jan Tinbergen personally told me that story [...] at a conference in Paris in 1961. (Lambert 2012, 70)


The criterion of Equity as no individual preferring any other’s allocation to her own, without any mention of equal liberty, has been introduced in economics by its mention by J. Tinbergen in application to equitable wages and occupations in *Redelijke Inkomensverdeling* (De Gulden Pers, Haarlem, 1946, in Dutch), following a suggestion from Tinbergen’s professor, the Dutch physicist Ehrenfest, in 1925. This previous suggestion by Ehrenfest was pointed out to me by Tinbergen in 1962 at a conference in Paris. (Kolm [1971] 1997, 6)

Naturally, the exact dates of this anecdotal attribution vary (for instance, Olson 2018 refers to 1921) and may be related to the fact that Tinbergen was a student of Ehrenfest from “around 1920” (Dekker 2021, 61) onwards. Nevertheless, there is strong evidence for the fact that Ehrenfest and Tinbergen had substantial conversations about what is now known as the ‘no-envy’ criterion.

In this article, we demonstrate that there is an additional, fourth, attribution: Tinbergen explicitly formulated the ‘no-envy’ principle in a 1930 article published in the Dutch journal *Mens en Maatschappij*, now translated into English in the *Erasmus Journal for Philosophy and Economics*.

**II.II. Tinbergen’s (1930) Formulation of the ‘No-Envy’ Principle**

The article “Mathematiese Psychologie” starts by, and mainly consists of, presenting the essential ingredients of early welfare economics as developed by Edgeworth (in *Mathematical Psychics*, 1881) and Pareto. Following these two authors, Tinbergen introduces the methodology of analysing the consumption of agents that have ordinal preferences over bundles of goods by using *indifference curves*. He introduces the main uses of indifference curves and presents material up to what is today known as Edgeworth’s box.

He opens the introduction of the ‘no-envy’ criterion by commenting that he is convinced that the methods of Pareto and Edgeworth can also be fruitfully applied to problems in normative economics. Most prominently,

---

7 For more detailed discussion about Ehrenfest and Tinbergen, see Boumans (1993) and Dekker (2021, chap. 4).

8 See also the next section for textual evidence from Tinbergen’s ([1930] 2021) article itself.
they can be applied to the problem of determining a fair income distribution: “The mathematical-psychological method seems to be useful for the analysis of the common conception of justice with regard to the rules governing the distribution of economic goods” (Tinbergen [1930] 2021, 216). Directly after this passage, there is a long footnote, which contains textual evidence of the role of Paul Ehrenfest: “The following problems and solutions were developed in very animated discussions with Prof. Ehrenfest” (216n5).

Tinbergen acknowledges the importance of being precise in the definitions used for such an analysis, which he starts as follows:

The determination of ‘just distributions’, for example, the distribution of the total production among all individuals, must of course be preceded by a definition of justice. As far as I can see, the common conception of justice entails the equal treatment of equally situated individuals. (Tinbergen [1930] 2021, 216)

Tinbergen then goes on to discuss two types of situations: one in which the individuals in question are equal (already alluded to in the above quotation), and one in which there are differences between them. On the equality case, Tinbergen’s comments are straightforward:

A) Two identical workers perform the same amount of labour of a certain kind. Which distribution of the product can be considered just? The criterion mentioned above gives the solution of equal distribution: equal wages. (Tinbergen [1930] 2021, 216–217)

Next, Tinbergen raises the problem of differences between individuals, and introduces the ‘no-envy’ criterion as the solution:

However, it cannot provide a solution in problems involving individuals in unequal circumstances, as, for example, when one extends the above case A in one or more of the following ways:

1) the quantities of work are unequal;
2) the workers have different ophelimity curves;*
3) the type of work is different.

As soon as one of these cases arises, in which the circumstances (in the broadest sense) of the individuals are different, the possibility of applying the principle of equal treatment is no longer there. The question is, how to extend the criterion of just distribution to these cases. In my opinion, the requirement that the in-
individuals should not have the desire to occupy each other’s position (in the broadest sense) should be adopted as such. We postulate this answer to the above question without any further justification other than that the first-mentioned ‘simple justice criterion’ must be included as a special case, which is indeed the case. (Tinbergen [1930] 2021, 217; italics in the original)

An open question for me here is whether it is sufficient for the psychological ‘identity criterion’ to hold for two people to be equal in terms of their ophelimity curves.

This passage is remarkable in a number of respects. Firstly, consider the careful introduction of three different dimensions on which differences between individuals might matter (we will return to this in the next section). Secondly, the footnote (included as footnote * in our article) that clarifies the role of ophelimity curves is remarkably careful as well. He seems to safeguard against taking them too literally: indifference curves are an important summary of an agent’s psychology (and useful, as they can be elicited by experiment), but also have limitations in how much of the psychology of an agent they can capture. Thirdly, and most importantly, Tinbergen presents the ‘no-envy’ criterion: ‘individuals should not have the desire to occupy each other’s position (in the broadest sense)’. Fourthly, he mentions how equality can be subsumed under it as a special case.

Tinbergen then goes on to illustrate how his ‘no-envy’ criterion can be applied. For this he considers two agents, who work for \(x_1\) and \(x_2\) hours respectively. They have (ordinal) preferences over pairs \((x, z)\) represented by two functions, \(w_1\) and \(w_2\). The simplest example that Tinbergen discusses is when the two agents have the same preferences, that is when \(w_1 = w_2\). Then, the ‘no-envy’ principle dictates that:

\[
\begin{align*}
    w(x_2, z_2) &\leq w(x_1, z_1) \\
    w(x_1, z_1) &\leq w(x_2, z_2)
\end{align*}
\]

(1)

Combining the ‘no-envy’ conditions for both agents for this case, we see that \(w(x_2, z_2) = w(x_1, z_1)\). Hence, if we know the hours \(x_1\) worked and wage \(z_1\) received by agent 1, we can find out which combinations of work-

\footnote{Tinbergen uses the Paretian term ‘ophelimity’—we use the current terminology of indifference curves in this article.}

\footnote{Whether Tinbergen also had in mind the idea that the presence of feelings of envy and the preference-based ‘no-envy’ criterion do not necessarily imply each other (discussed in section II.I) is not entirely clear from the passage and footnote. It is certainly consistent with it.}
Heilmann and Wintein

No Envy: Jan Tinbergen on Fairness

Figure 1: ‘No-envy’ (Tinbergen [1930] 2021, Figure 7).

Figure 2: ‘No-envy’ with different indifference curves (Tinbergen [1930] 2021, Figure 8).

...ing hours and wage for agent 2 would be fair and would result in ‘no-envy’. Tinbergen ([1930] 2021) explains, referring to Figure 1, that ‘no-envy’ requires that agent 2’s combination \((x_2, z_2)\) has to be located on the same indifference curve as agent 1’s \((x_1, z_1)\).

For different indifference curves, the ‘no-envy’ criterion reads:

\[
w_1(x_2, z_2) \leq w_1(x_1, z_1) \quad w_2(x_1, z_1) \leq w_2(x_2, z_2)
\]

We discuss equation (2) by referring to Figure 2.

Figure 2 displays two points, \((x_1, z_1)\) and \((x_2, z_2)\), for which (2) is satisfied. In order to satisfy (2), the point \((x_2, z_2)\) must be under the indifference curve of agent 1 (continuous line) through \((x_1, z_1)\) while the point \((x_1, z_1)\) must be under the indifference curve of agent 2 (dashed line) through \((x_2, z_2)\). Indeed, these conditions are fulfilled for the points, \((x_1, z_1)\) and \((x_2, z_2)\), that are displayed in Figure 2. In contrast to the case discussed above, in which the two agents have identical indifference curves, a particular combination \((x_1, z_1)\) of (hours worked, wage paid) for agent 1 does not uniquely determine the combination \((x_2, z_2)\) for agent 2 that results in envy-freeness. Indeed, with \((x_1, z_1)\) given as in Figure 2, the grey area marks possibilities for \((x_2, z_2)\) that result in envy-freeness.

This concludes the presentation of Tinbergen’s early 1930 formulation of the ‘no-envy’ criterion. Its remarkably careful presentation as well as concrete exposition via the example strikes us as a central finding in two ways. One, it further clarifies the somewhat ambiguous origins of the modern ‘no-envy’ criterion. Two, by combining a very general formulation (that mentions social positions ‘in the broadest possible sense’) with a concrete exposition of application (two agents with different wages and
hours worked), he anticipates different types of applications of it. We now turn to briefly introducing Tinbergen’s (1946, 1953) formulation.

II.III. ‘No-Envy’ in Tinbergen (1946, 1953)

Tinbergen’s discussion in his book *Redelijke Inkomensverdeling* does not contain comments on the theoretical aspects of welfare economics that were present in his 1930 article. However, with respect to ‘no-envy’, it is both conceptually and empirically richer than the 1930 article, containing a number of reflections on the relation between ‘no-envy’ and welfare. Partly, he gives broader context to his sparse discussion in the 1930 article, and partly, he also motivates some of the underlying ideas further. In particular, the three dimensions of inequality (quantity of work, preferences, type of work) mentioned in the 1930 article are discussed more thoroughly.

Tinbergen starts his discussion on ‘no-envy’ in the book by noting that a two-fold question should be answered: how to *produce* the largest amount of welfare, and how to *divide* the benefits and burdens associated with this production fairly (Tinbergen 1953, 51)? As the demands of production and the demands of fairness are typically not aligned, one needs to find a harmonious synthesis between them, for which Tinbergen reserves the term *redelijkheid*, which could be translated as “reasonableness”. It figures in the title of the book (*Redelijke Inkomensverdeling*, which could be translated as “Reasonable Income Distribution”). Reasonableness refers to the synthesis between the demands of production and fairness associated with income distributions.

After re-affirming the *equal treatment of equals* principle for contexts in which individuals are similar, Tinbergen laments the fact that principles guiding dissimilar conditions are not available. Tinbergen (1953, 53) notes that the same amount (of income) for all will not do: rewarding easy, low-skilled work and hard, high-skilled work equally conflicts with our sense of fairness. Tinbergen suggests that this conflict can be explained by the principle that fairness requires that everyone is equally *happy*: if one puts in more effort, works hard, or is ill, fairness requires that one is compensated so that, after compensation, all are equally happy. Should we then say that fairness requires the *equal happiness of all*? Tinbergen argues that *equal happiness of all*, albeit conceptually attractive, is not tenable as a principle that specifies the demands of fairness:

This now seems a very clear description of our goal with respect to income distribution: an equal amount of happiness for all. On closer inspection, however, it is unusable for offering practical guidance. For
how to determine whether two people are equally happy? (Tinbergen 1953, 54; translation by the authors)\textsuperscript{11}

Tinbergen argues that we cannot determine whether two people are equally happy, which dismisses \textit{equal happiness for all} as a principle that specifies the demands of fairness. In modern terminology, we would say that Tinbergen dismisses the \textit{equal happiness for all} principle because it relies on the interpersonal comparison of utility which, according to Tinbergen (1946, 1953), is non-sensical. But then, if one abstains from such interpersonal comparisons, how are we to formulate the demands of \textit{fairness}?

Tinbergen again introduces the ‘no-envy’ criterion as an answer to precisely this question. The main idea is that each individual evaluates every other individual’s position in terms of their \textit{own} preferences, so that interpersonal comparisons are avoided:

\begin{quote}
We say that two people have been treated justly with respect to each other \textit{when neither of them wants to trade places with the other}. By ‘place’ we mean, figuratively, all circumstances: in particular the profession and the income, but also the personal circumstances, such as health, wellbeing and health of the family, degree of development, ability, etc. (Tinbergen 1953, 55; translation by the authors)\textsuperscript{12}
\end{quote}

Tinbergen goes on to discuss this principle in relation to income distribution, and links it to questions of how to deal with, on the one hand, differences in innate abilities and health and, on the other hand, differences in effort, desert, responsibility, motivation, and incentive. Both might present problems in terms of equalizing. Regarding the first set of issues, it may be wholly impossible to create a state of the world in which innate abilities and health are distributed such that no one wants to trade places (this point is already mentioned in Tinbergen [1930] 2021, section V). Regarding the second set of issues, it may not only be impossible but also undesirable to create a state of the world in which aspects such as effort, desert, responsibility, motivation, and incentive are distributed such that

\textsuperscript{11} The original in Dutch: “Dit lijkt nu een heel duidelijke omschrijving van ons doel inzake de inkomensverdeling: een even groot geluk voor ieder. Bij nader inzien is hij echter toch onbruikbaar, om als practisch richtsnoer te dienen. Want hoe moeten wij uitmaken, of twee mensen even gelukkig zijn?” (Tinbergen 1953, 54).

\textsuperscript{12} The original in Dutch: “Wij zeggen, dat twee personen rechtvaardig t.o.v. elkaar zijn behandeld, \textit{wanneer zij geen van beiden met de andere van ‘plaats’ willen ruilen}. Met ‘plaats’ bedoelen we dan overdrachtelijk alle levensomstandigheden: in het bijzonder het beroep en het inkomen, maar daarnaast ook de persoonlijke omstandigheden, zoals gezondheid, goote en gezondheid van het gezin, grad van ontwikkeling, bekwaamheid, enz” (Tinbergen 1953, 55; italics in the original).
no one wants to trade places. In relation to this, Tinbergen stresses that the principle of ‘no-envy’ is a *pro tanto* principle: it is but one criterion in judging social states of affairs. And so, for Tinbergen, the issue of a ‘reasonable’ income distribution is to seek balance between considerations of efficient production and fair division of welfare.

There is thus a slightly expanded discussion of the ‘no-envy’ criterion in the 1946/1953 book when compared to the 1930 article. Moreover, in the 1930 article, he gives a concrete example of two individuals to illustrate the principle (presented in section II.II), whereas in his book he repeatedly insists that the ‘no-envy’ criterion applies to *social groups*: “To avoid misunderstanding we repeat once more that [the ‘no-envy’ criterion] albeit explained via examples involving individual persons, is meant to apply to social groups” (Tinbergen 1953, 59).

Otherwise, both formulations of the criterion itself are non-committal and open about the domain of application. For instance, Tinbergen ([1930] 2021, 217) references social positions “in the broadest sense” and the above quotation from the 1946/1953 book refers to ‘figuratively, all circumstances’. It is helpful to keep these slight ambiguities in mind when we discuss other early contributions in the ‘no-envy’ literature.

### III. THE ‘NO-ENVY’ CRITERION AFTER TINBERGEN

#### III.I. Comparing Different Variants of the ‘No-Envy’ Criterion

The 1930 and 1946/1953 formulations of the ‘no-envy’ criterion by Tinbergen raise the question of how these compare to the other early proposals of it. We think it is useful to consider two dimensions of comparison.

Firstly, applying the ‘no-envy’ criterion can differ in terms of the alternatives over which the preferences are considered, that is, the ‘objects of envy’, so to speak. It is interesting to note that this is by and large an interpretative issue—one can always formulate the ‘no-envy’ criterion in terms of an allocation of *something*. These might be goods with a varying degree of complexity, or, as Tinbergen’s 1930 example has it, a combination of hours worked and wages received. In Tinbergen’s richer 1946/1953 discussion, there are some hints that he thinks of the allocation primarily in terms of income (also referenced in the title of his book), which he saw as being tied to societal ‘places’ or ‘positions’. Regarding the latter, however, it is important that he already mentions social positions in the 1930 article, as discussed above. So, already in Tinbergen’s early article (as well as in the book), there are different possible ‘objects of envy’ mentioned.

A second dimension on which ‘no-envy’ criteria may differ is the ecology of concepts within which they are conceived of and applied to. Simply
put, what are other theoretical concepts that the ‘no-envy’ criterion is used in conjunction with? Tinbergen’s sparse 1930 example makes no mention of any other concept (he even avoids explicitly mentioning utility).\textsuperscript{13} However, his 1946/1953 book conjures many other concepts, including efficiency of production, but also innate ability, endowments in relation to the social position of the individuals concerned, and effort. But he is not, in contrast to later work in economic theory, concerned with the efficiency of \textit{distribution} yet.

There are, no doubt, more dimensions on which different formulations of the ‘no-envy’ criterion might be compared. However, we are not aiming at a complete taxonomy of ‘no-envy’ criteria in this article that addresses all these issues. Rather, we want to stay focused on presenting and contextualising Tinbergen’s 1930 contribution. Arnspenger (1994, 157) mentions the following as “seminal references” for the ‘no-envy’ criterion: Tinbergen (1946), Foley (1967), Schmeidler and Yaari (1971), Kolm ([1971] 1997), and Varian (1975). We take Foley (1967) as the main source for comparison. We do so in part because he has been the main historical reference in the ‘no-envy’ literature. However, we will also see that the differences in the intended domain of application in Tinbergen (1930) and Tinbergen (1946, 1953) connect very well to issues in the interpretation of ‘no-envy’ within Foley (1967).

\textbf{III.II. Foley (1967) and Tinbergen (1930, 1946/1953)}

Foley (1967) has hitherto been considered the \textit{locus classicus} for the ‘no-envy’ criterion in the literature. He introduces it as follows.

There is an intuitive notion of inequality. If one person consumes more of every good (including leisure) than another, he is better off. If two people have identical preferences and one is in a position preferred by both, they cannot have equal welfare. This suggests a new way to define equality even when preferences are diverse: an allocation is equitable if and only if each person in the society prefers his consumption bundle to the consumption bundle of every other person in the society. (Foley 1967, 74)

It seems straightforward that the underlying idea is exactly the same as in Tinbergen (1930, 1946/1953). Interestingly, considering the above passage alone might suggest that Foley is interested in a narrow object of envy: that is, comparing preferences between consumption bundles. It

\textsuperscript{13} Observe, however, that Tinbergen uses the letter \(w\) in his 1930 article in the preference comparison. This might be taken as a hint at a welfarist interpretation.
seems to us that when viewed in isolation like this, it is easy to reduce Foley’s ‘no-envy’ criterion to such a quite narrow version. To see this, we discuss a typical example that is used in the literature to introduce Foley’s formulation of ‘no-envy’. Consider an economy with two goods, food and clothes, and two agents, Alice and Bob, whose (ordinal) preferences for the goods are represented by the following utility functions:

\[ U_A(f, c) = 2f + c \quad U_B(f, c) = f + 2c \]

Suppose that there are 4 units of food and 4 units of clothes in total. An allocation \((x_A, x_B)\) distributes the total endowment amongst Alice, who receives \(x_A = (f_A, c_A)\), and Bob, who receives \(x_B = (f_B, c_B)\). An allocation for this economy is envy-free just in case it satisfies equation (3):

\[ U_A(f_B, c_B) \leq U_A(f_A, c_A) \quad \text{and} \quad U_B(f_A, c_A) \leq U_B(f_B, c_B) \quad (3) \]

Clearly then, Foley’s notion of ‘no-envy’ if articulated by (3) is formally equivalent to Tinbergen’s ‘no-envy’ notion as articulated by (2). However, the interpretation of the objects of comparison is slightly different: Tinbergen (1930) introduced preferences over pairs of wages earned and hours worked, whereas the above formulation concerns the allocation of consumption bundles.

Was Foley more concretely minded than Tinbergen in the domain of application of the ‘no-envy’ criterion? Young (1994) suggests that he was. He writes that Tinbergen puts forward the ‘no-envy’ principle “in a very strong form” (Young 1994, 11). He takes issue with Tinbergen specifying ‘no-envy’ in terms of agents not wanting to trade places or ‘positions in the broadest sense’. For, as the notion of a place is very general and may involve non-transferable personal characteristics, such as height, health, intelligence, or age, Tinbergen’s ‘no-envy’ principle may be impossible to realize. Young illustrates this non-realizability via an example where person A envies person B for being tall, whereas B is indifferent between being tall or short. Then, a monetary compensation of A’s envy of B’s height will result in B being envious of A’s money. Hence, there is envy come what may, so that an envy-free arrangement does not exist. According to Young, this example illustrates that “there may be no reasonable way of eliminating envy in Tinbergen’s sense”, and he then quickly turns to “the more pragmatic formulation” (Young 1994, 11) due to Foley (1967) which, according to Young, defines envy-freeness not in terms of societal places but, more narrowly, only in relation to particular allocations of goods.
It is true that Tinbergen (1946, 1953) applies the ‘no-envy’ criterion to ‘social groups’. However, he discusses in quite some detail that people’s inability to move freely between associated social ‘places’ presents serious problems for realizing fair, envy-free income distributions. But also, Tinbergen (1930) contains concrete, and in that sense pragmatic, applications of the ‘no-envy’ criterion that we discussed in section II.I. Indeed, in the models discussed in Tinbergen (1930), the non-existence of envy-free allocations is not at stake.

The differences between Tinbergen and Foley should not be overemphasized. For Foley—even though one may suitably make his proposal concrete in the way introduced above—was not necessarily that much ‘more pragmatic’ than Tinbergen, as Young (1994) suggested. Foley explains some issues of interpretation of the consumption bundles in question, which we quote here at length. It strikes us as interesting in terms of the interpretation and intended domain of application of the ‘no-envy’ criterion:

Ask each person to imagine changing places with every other, not by exchanging incomes, but by experiencing the material aspects of that person’s life. If no one is willing to change, the allocation is equitable. The farmer might be willing to take on the bricklayer’s apartment, TV, dinners, steady working hours, and Saturdays in a bar in exchange for his own homestead, home-grown food, and freedom from supervision if the city air was clean but not if it was dirty. In applying the test there are some simple qualifications. First, the consumption patterns compared must be lifetime plans to wash out differences of time preference and other noncomparabilities. The postponement by professional men of marriage and of even modest incomes must be taken into account when the salesman views their high rates of consumption later in life. Second, the comparison can be pushed only so far, and some common sense procedures are necessary when the comparison involves orchestra conductors, painters, chess masters, and so on. The difficulty here is that if a gas station attendant has the desire to be a painter but not the ability, it may be necessary to make the painter’s life very unattractive in other ways before the gas station attendant will prefer his own; so unattractive, perhaps, that the painter will envy the attendant while the attendant is still envying him. These cases must be interpreted flexibly; either equivalents to the talents must be postulated which the gas station attendant does possess, or reasonable alternatives framed that abstract from the glamour and prestige of certain activities. (Foley 1967, 75)
In other words, it seems that Foley (1967) is—quite similarly to Tinbergen (in both 1930 and 1946/1953)—also aware of the fact that a too narrow interpretation of what is compared in the preference used for the ‘no-envy’ criterion may not be opportune. In relation to this, it is also important to note that Foley (1967) did not himself give the concrete example that we used above to illustrate his view. In contrast, Tinbergen (1930) did provide two concrete cases of specific hours worked and wages earned in order to illustrate the ‘no-envy’ principle.

It is perhaps not desirable to get too bogged down in discerning the precise meaning that Tinbergen and Foley were trying to convey in what were not fully worked-out theoretical developments, but rather statements of an ethical principle that seemed plausible to them. What seems important to us is that both authors are aware that there are different kinds of possible objects of envy (ranging from concretely specified consumption bundles to social states in the broadest sense). This issue seems particularly salient as the literature on fair allocation in economic theory has developed addressing the envy-free (and Pareto-optimal) allocation of goods, whereas Tinbergen’s key concern has remained a fair, or reasonable, income distribution.\footnote{For an elaborate discussion on the latter point, see Dekker (2021, 375–392).}

III.III. ‘No-Envy’ after Tinbergen and Foley

Varian (1975) expanded on Foley (1967) by proving results on the relation between fairness and efficiency. This marks a clear shift in the literature. Whereas Tinbergen ([1930] 2021) and Foley (1967) mainly stated the ‘no-envy’ principle, and included some thoughts on the different types of objects of envy, the ecology of concepts they considered alongside ‘no-envy’ did not yet play a major role for them. As mentioned in section II.III, Tinbergen (1946, 1953) discussed the demands of production efficiency alongside the fair distribution of welfare. However, Tinbergen also refrained from offering a theoretical discussion of their relation along the lines of Varian (1975) and other authors after the latter.

In order to appreciate Varian’s (1975) contribution, we continue the above example of an economy with two goods, food and clothes, and two agents, Alice and Bob, whose (ordinal) preferences for the goods are represented by the following utility functions:

\[
U_A(f, c) = 2f + c \quad U_B(f, c) = f + 2c
\]
As before, suppose that there are 4 units of food and 4 units of clothes in total. Now, quite a few envy-free allocations for the example at hand exist. In particular, allocation $a = ((3, 0), (1, 4))$ is envy-free: Alice receives $(3, 0)$ which gives her a utility of 6 so that she does not envy Bob’s $(1, 4)$ which would also give her a utility of 6. Likewise, Bob’s $(1, 4)$ yields him a utility of 9 so that he does not envy Alice’s $(3, 0)$ which would only give him a utility of 3. Not only is $a$ envy-free, but it is also (Pareto)-efficient—that is, there is no allocation $x$ which is strictly preferred to $a$ by one individual while $x$ is weakly preferred to $a$ by all other individuals. The notions of efficiency and envy-freeness do not always coincide, as they do in allocation $a$. For instance, $((2, 2), (2, 2))$ is envy-free but not efficient, whereas $((0, 0), (4, 4))$ is efficient but not envy-free.

From Varian’s (1975) contribution onwards, a large literature developed in economic theory that explored the relations between efficiency (via the Pareto principle) and fairness (via the ‘no-envy’ criterion), which is reviewed in Arnsperger (1994), Kolm (1996), and Thomson (2011). Fleurbaey (2008) makes seminal contributions at the intersection of philosophy and economics, and Olson (2018) and Fleurbaey (2021, section 6) review philosophical contributions.

The strand of literature on envy-freeness succeeding Varian (1975) is by and large concerned with existence results: it is concerned with the question under what conditions envy-free allocations may or may not exist. The question of how rational, self-interested individuals may end up in an envy-free allocation, that is by bargaining, trading or, more generally by ‘following the rules of the games’, is not discussed. However, in the literature on ‘algorithms for fair division’ that addresses these very questions, the notion of ‘no-envy’ plays a prominent role, as acknowledged by two of its most prominent contributors:

Our approach to fair division is distinctive not only in combining properties, algorithms, and applications but also in elevating the property of ‘envy-freeness’, and procedures that generate envy-free allocations, to a central place in the study of fair division. Roughly speaking, an envy-free division is one in which every person thinks he or she received the largest or most valuable portion of something—based on his or her own valuation—and hence does not envy anyone else. Although the concept of envy-freeness has been used in the mathematics literature on fair division for almost 40 years (Gamow and Stern, 1958) and in the economics literature for almost thirty years (Foley, 1967), only recently have several algorithms been developed that guarantee
envy-freeness in a wide variety of situations. (Brams and Taylor 1996, 2)

As demonstrated in this article, the concept of envy-freeness has been used in the economics literature for more than 90 years now. And, as mentioned in the above quotation, the 'no-envy' criterion is indeed applicable to a wide variety of problems. In this analysis, envy-freeness plays an important role. However, as this literature assumes that preferences are interpersonally comparable, further, non-ordinal fairness criteria become available and are studied. To illustrate these non-ordinal criteria, we briefly consider the most well-known problem studied in the algorithms for fair division literature: the so-called 'cake-cutting problem'.

A single rectangular cake has to be divided, in two pieces that result from a single vertical cut, amongst Alice and Bob. The cake need not be homogeneous—that is, it may be topped with chocolate on the left but with strawberries on the right—and Alice and Bob need not have the same preferences for chocolate and strawberry. What properties should a division \((A, B)\) of the cake, allotting piece \(A\) to Alice and piece \(B\) to Bob, satisfy? The following two familiar properties are attractive:

\[
\text{Envy-freeness: } U_A(B) \leq U_A(A) \quad \text{and} \quad U_B(A) \leq U_B(B)
\]

\[
\text{Efficiency: } \text{For no } (A', B') : U_A(A) \leq U_A(A') \text{ and } U_B(B) < U_B(B') \\
\text{or } U_A(A) < U_A(A') \text{ and } U_B(B) \leq U_B(B')
\]

Envy-freeness and efficiency only require ordinal utilities which need not be interpersonally comparable. Cake divisions may also satisfy two further attractive properties, which are more demanding, concerning the comparability of utilities. \textit{Proportionality} requires that both Alice and Bob receive, according to their own valuation, at least half of the cake’s value, whereas \textit{equitability} requires that Alice and Bob receive the same utility from their cake pieces:

\[
\text{Proportionality: } U_A(A) \geq (1/2)U_A(\text{cake}) = 1/2 \quad \text{and} \quad U_B(B) \geq (1/2)U_B(\text{cake}) = 1/2
\]

\[
\text{Equitability: } U_A(A) = U_B(B)
\]

\footnote{The cake preferences of Alice and Bob are assumed to be expressed by additive and normalized utility functions \(U_A\) and \(U_B\). Additive means that for any two pieces of cake \(X\) and \(Y\), \(u_i(X \cup Y) = u_i(X) + u_i(Y) - u_i(X \cap Y)\). Normalized means that receiving nothing yields a utility of 0 whereas receiving the whole cake yields a utility of 1.}
Arguably, envy-freeness, efficiency, proportionality, and equitability are desirable properties for an allocation to have. Moreover, one may even argue that, for problems with interpersonally comparable preferences, fairness requires more than just envy-freeness and also, say, proportionality or equitability.

However, what really sets apart the literature on algorithms for fair division is not the specification of non-ordinal ‘fairness criteria’, but the focus on algorithms. To illustrate, suppose that Alice and Bob have to divide the cake but there are no parents, government, or impartial observer to do the cutting and dividing for them, what should they do? They could, for instance, use the cut-and-choose algorithm which, intuitively, seems to be a fair way of solving the problem. According to cut-and-choose, one individual, say Alice, cuts the cake after which Bob is allowed to choose a piece, leaving the other piece for Alice. Cut-and-choose is intuitively fair and it can be shown that, when Alice and Bob cut and choose rationally, it will result in an allocation that is envy-free and proportional, but not necessarily equitable. Further, the resulting allocation is efficient with respect to the class of all single-cut allocations.

As discussed in sections II.II and II.III, Tinbergen did not consider preferences to be interpersonally comparable. Moreover, he did not consider or discuss algorithms in the sense of the literature discussed above. Indeed, Tinbergen ([1930] 2021, 218–219) assumes that when an envy-free allocation exists, it will be reached “automatically”. However, Tinbergen was concerned with conditions under which envy-free allocations could be reached. Indeed, he writes that an envy-free allocation exists under certain conditions: “When any individual can indeed, if so desired, reach a state (in the broadest sense) equal to that of any other individual: when there is ‘free mobility’ in all respects” (Tinbergen [1930] 2021, 219).

He then notes that this free mobility may be absent in a wide variety of circumstances, due to a wide variety of causes and, in the last section of Tinbergen ([1930] 2021), presents various examples of this absence. This vindicates our earlier remark, made in section III.II, that Tinbergen was well aware of the difficulties surrounding the concept of social states in the broadest sense.

IV. TINBERGEN AND CONCEPTUAL DISCUSSIONS OF FAIRNESS

We now turn to two general observations about the relation between Tinbergen’s contributions and conceptual discussions of fairness.

Firstly, in our translation of Tinbergen’s 1930 article (Tinbergen [1930] 2021), we have stayed as closely as possible to the original, and therefore
chosen to use the words ‘just’, ‘justly’, or ‘justice’ where one might also have used the words ‘fair’, ‘fairly’, or ‘fairness’. It is worth mentioning that the usage of these words does not follow consistent rules in the contemporary literature in philosophy and economics. Furthermore, there are also a number of other closely related labels that are used in the various strands of literatures, and the usage of ‘justice’ and ‘fairness’ varies. One trend in terminology that seems to prevail, though, is that ‘fairness’ is more often used for smaller and more easily delineated contexts, whereas ‘justice’ is often used for broader, more encompassing problems of larger scale. Peyton Young comments:

Fairness in [a] global sense is concerned with the proper distribution of resources, rights, duties, opportunities, and obligations in society at large. This grand theme has animated political philosophers since antiquity, from Plato’s and Aristotle’s conceptions of the ideal state, to the social contract theories of Hobbes, Locke, and Rousseau, to the more modern theories of Rawls, Nozick and Walzer. I shall refer to these as theories of social justice. (Young 1994, xi)

We agree with Young that it is useful to distinguish between ‘local’ and ‘global’ fairness problems, and also with his assertion that problems of a local scale can be solved without invoking ‘global’ theories (whether one wants to refer to them as concerned with ‘fairness’ or ‘justice’). Here, it is interesting that the ‘no-envy’ criterion might be applied to contexts of different scale. Indeed, this issue is linked to the different suggestions for the object of envy in the texts by Tinbergen and Foley, which we analyzed in sections II and III. Simply put, one might use the ‘no-envy’ criterion in local settings (such as cake cutting, or in general the division of a very simple bundle of goods), or one might conceive of it globally, in terms of judging the fairness of social structures (such as when conceiving of whole life histories of individuals that touch on all kinds of social arrangements).

Secondly, we comment on a structural coherence between Tinbergen’s starting point in theorizing about fairness and that of key fairness theories in philosophy and beyond. There is broad agreement between fairness theorists in both philosophy and economics that fairness requires (1) ‘the equal treatment of equals’. However, a fairness theory also needs to specify what fairness requires for (2) ‘the treatment of unequals’. Tinbergen also comments on the different status of (1) and (2), both in his 1946/1953 book and in his 1930 article:

16 For a more detailed discussion, see Young (1994) and Wintein and Heilmann (forthcoming).
However, one principle is generally accepted; one could call it ‘same monks, same robes’. Those who are in similar condition should also be treated equally. There is no equally clear principle about dissimilar conditions. (Tinbergen 1953, 50; translated by the authors)\textsuperscript{17}

As far as I can see, the common conception of justice entails the equal treatment of equally situated individuals. […] However, it cannot provide a solution in problems involving individuals in unequal circumstances. (Tinbergen [1930] 2021, 216–217)

Tinbergen’s ([1930] 2021) ‘no-envy’ principle is thus a solution for (2) which, as he remarks, includes (1) as a special case.

So, Tinbergen’s ‘no-envy’ principle substantiates (1) and (2) for cases where individuals have ordinal preferences. But, more generally, there is broad consensus on the fundamental two-fold fairness concern in the literature on fairness at large, such as in Aristotle’s \textit{Nichomachean Ethics}. The two-fold fairness concern is also visible in the algorithmic literature cited in section III.III, via the principles equitability and proportionality for interpersonally comparable utilities.

Indeed, the two-fold fairness concern also extends across the divide between subjective and objective fairness theories. As remarked in section II.I of this article, there are theories that view fairness as something subjective (such as those theories based on a ‘no-envy’ criterion) and there are also theories that conceptualize fairness as objective. Now, objective fairness theories also strive to give meaning to the two-fold fairness concern of (1) equal treatment of equals and (2) unequal treatment of unequals. Consider Broome’s (1990) influential account of objective fairness, according to which fairness requires to ‘satisfy claims in proportion to their strength’. This principle clearly gives substance to (2); and it includes (1) as a special case, as it follows that claims with equal strength require equal satisfaction.

We find it remarkable that both the distinction between local and global fairness as well as the two-fold fairness concern are explicitly visible in Tinbergen’s early contributions. We also think it shows that, more generally, conceptual discussions of fairness can fruitfully be viewed as competing specifications of such general desiderata of fairness theories. We develop more specific discussions of this theme elsewhere (Heilmann and Wintein 2017; Wintein and Heilmann 2018, 2020, forthcoming). Here,\textsuperscript{17}

\textsuperscript{17}The original in Dutch: “Een principe is wel algemeen aanvaard; men zou het door de uitspraak: ‘gelijke monniken, gelijke kappen’ kunnen aanwijzen. Wie in hetzelfde geval verkeren, moeten ook gelijk behandeld worden. Over ongelijke gevallen bestaat echter geen duidelijk principe” (Tinbergen 1953, 50).
we stress the intriguing parallels between Tinbergen's and many other
contributions in the interdisciplinary discussions about fairness in phi-
losophy, economics, and beyond.

V. Conclusion
We have revealed that Jan Tinbergen introduced ‘no-envy’ as a fairness
criterion in his article “Mathematiese Psychologie” published in 1930 in
the Dutch journal Mens en Maatschappij and translated as “Mathematical
Psychology” in 2021 in the Erasmus Journal for Philosophy and Economics.
This fact calls for a revision in the attribution of the ‘no-envy’ criterion in
the fairness related literature.

We discussed Tinbergen's (1930, 1946/1953) and Foley’s (1967) for-
mulation of the ‘no-envy’ criterion. We found that there are similarities:
both authors speak of the criterion being applied to an individual’s ‘place'
or ‘position’ within society, but also slight differences in the interpreta-
tion of the object of envy: Tinbergen (1930) considered hours worked
and wages earned, Tinbergen (1953) considered income, and Foley (1967)
considered consumption bundles. We also briefly commented on how
Tinbergen's notion of ‘no-envy’ connects to that used by the more recent
literature on ‘no-envy’ and, more broadly, fairness discussions in general.

Jan Tinbergen’s contributions to theories of redistribution are well
known—here we have shown that he has made crucial, and early, con-
tributions to the conceptual discussion of fairness as well.

References
Brams, Steven J., and Alan D. Taylor. 1996. Fair Division: From Cake-Cutting to Dispute
Edward N. Zalta. Article published December 18, 2002; last modified December 22,
Denyer, Nicholas. 2013. “The Political Skill of Protagoras.” In Politeia in Greek and
Roman Philosophy, edited by Verity Harte and Melissa Lane, 155–167. Cambridge:
Cambridge University Press.
Edgeworth, Francis Y. 1881. Mathematical Psychics: An Essay on the Application of


**Conrad Heilmann** is associate professor of philosophy and co-director of the Erasmus Institute for Philosophy and Economics (EIPE) in the Erasmus School of Philosophy and Core Faculty of the Erasmus Initiative “Dynamics of Inclusive Prosperity” at Erasmus University Rotterdam. He is the co-editor (with Julian Reiss) of The Routledge Handbook of Philosophy of Economics.
Contact e-mail: <heilmann@esphil.eur.nl>

**Stefan Wintein** is assistant professor of theoretical philosophy at the Erasmus School of Philosophy (ESPhil) and member of the Erasmus Institute for Philosophy and Economics (EIPE), both at Erasmus University Rotterdam in the Netherlands. Together with Conrad Heilmann, he runs the Fairness project: http://fairness-research.org. Contact e-mail: <wintein@esphil.eur.nl>

DOUGLAS W. ALLEN
Simon Fraser University

This is an odd book. At just eighty pages, it is shorter than most modern applied-micro papers; the hardback cover is thicker than the text! However, with around 250 references, ranging across the past century and several schools of thought, the number of ideas per page is quite staggering. The terse, dense writing, combined with translation issues related to articles/number/plurals between Italian and English, make this a very difficult read. Perhaps this is why, after reading it, I am not fully confident that I understand it. It also partially explains my harsh review.

The structure of the book is simple enough. It lays out four ‘institutional’ notions of a transaction as described by Ronald Coase, John Commons, Robert Lee Hale, and Oliver Williamson. It then describes three “dimensions” (17) of a transaction: legal, competitive, and political. This is followed by a chapter on each dimension and a short concluding chapter calling for a research agenda. My conjecture of the book’s purpose is that the author wants to integrate ideas and expressions of some older institutional writers (most notably Commons and Hale, but also Adolf Berle, Wesley Hohfeld, and Thorstein Veblen) with the works of Coase, Oliver Hart, Williamson, and a few other ‘new’ institutionalists.

That’s all well and good, but I had trouble with the details. This stemmed from the common failure in this genre to carefully define terms. Here is a book on ‘the transaction’, written within the context of ‘institutional economics’, yet on the first page the words “institution” and “transaction costs” appear with no definition. Nor is one coming. A few pages later “property rights” show up in a sentence about “market transactions” but with no definition and no explanation of the connection (4). And away we go.

This is problematic because the author obviously has a conception of ‘transaction’ in mind, and it influences the way he is understanding Coase and others. However, because this understanding is implicit, the book reads as being inconsistent.
Consider the introductory discussion of Coase’s “ideas of transaction” where the author asserts that the “scope” of Coase's 1960 paper was to “question the claim that state intervention in the economy is necessary or desirable” (2). However, in section VI of his 1960 paper, Coase states:

*An alternative solution is direct Government regulation.* Instead of instituting a legal system of rights [...] the government may impose regulations [...]. Thus, the government [...] may, to deal with the problem of smoke nuisance, decree that certain methods of production should or should not be used [...] or may confine certain types of business to certain districts [...]. (Coase 1960, 17; emphasis added)

It could be argued that Coase gave a justification for state intervention in the economy.

The author wants to claim that Coase “identifies three types of transactions: 1. market transactions, 2. transactions within a firm, and 3. transactions within a super-firm” (4). However, Coase literally does not do this. Coase (1960) believed that ‘transactions' only took place in a ‘market exchange'. Consider the following from Coase:

Within the firm individual bargains between the various cooperating factors of production are eliminated and *for a market transaction is substituted an administrative decision.* [...] It does not, of course, follow that the administrative costs of organising a transaction through a firm are inevitably less than the costs of the market transactions which are superseded. [...] This solution would be adopted whenever *the administrative costs of the firm were less than the costs of the market transactions* that it supersedes and the gains which would result from the rearrangement of activities greater than the firm’s costs of organising them. (Coase 1960, 16–17; emphasis added)

Coase did not see the firm as a collection of ‘transactions', but rather as an organization where production takes place through direction. The ‘transaction' is 'superseded' by a firm. As such, Coase (perhaps confusingly) makes a distinction between the ‘costs of market transactions' and ‘administrative costs of the firm'. Likewise in Coase's mind, ‘super-firms' like governments, do not engage in market exchanges and therefore transactions, but rather governments direct production through statute, regulations, and zoning, among other things.

I happen to believe, probably along with the author, that production involves a system of property right exchanges, and that these can be de-
fined as various forms of a transaction. In this light, Coase could be interpreted as ‘identifying three types of transactions’. Had the author developed a definition and theory of ‘a transaction’ such an interpretation would make sense, but as it stands the text comes across as muddled.

A bigger problem is the book’s treatment of the Coase Theorem. Coase wrote the first four sections of the 1960 paper not to describe the real world, not to make some type of normative policy suggestion, but to criticize the neoclassical model of economics. The Coase Theorem simply states: if transaction costs are zero, then the allocation of resources is independent of the distribution of economic property rights (Allen 2000). Since the neoclassical model (whether in its perfectly competitive form or any other version) implicitly assumes zero transaction costs, then it cannot have anything to say about the distribution of economic property rights. In other words, it cannot be used for understanding organizations and institutions. As Coase stated at the end of section VI: “the problem is one of choosing the appropriate social arrangement for dealing with the harmful effects” (1960, 18).

Consider this claim by the author: “the difference between the Coase theorem world and the perfect competition framework is represented by this legal dimension, consisting of the active role of the public official actor in defining and enforcing rights” (20–21). Rights in ‘the Coase theorem world’ just exist without cost. They are not defined or enforced by anyone. If such a public official existed, he would be in a real world, not the imaginary world of zero transaction costs. For this reason, it is also not a “corollary to the Coase theorem […] that the law […] should be designed by a public official actor to lower transaction costs” (21). If we were in a ‘Coase theorem world’, transaction costs would be zero, and could not be lower.

This sort of troubling logical inconsistency is all over the place when it comes to Coase, but I will point out one particular instance of importance. Chapter 3 is devoted to the “legal dimension of transactions” (33), and a great deal of time is spent on the concept of ‘positional’ goods—goods where relative consumption matters along with the level of consumption. The climax of the chapter comes with the declaration that specific laws are positional. Therefore, the allocation of property rights is positional, and as a result there are not just concerns over social costs, but also of positional status. Each allocation of rights “creates or redistributes legal positions that are characterized by positional concerns” (45). This means that “no (property) right is a ‘free lunch’ [… and therefore
rights are] never neutral, as the Coase theorem instead suggests” (46). This conclusion does not follow, and results from having a positive transaction cost slip into the analysis: namely the ‘political actor’ is switching rights around without a compensation (that is, a theft), and this cannot happen under zero transaction costs. Allen (2015) discusses this in the more famous case of the Coase Theorem holding only if income is held constant.

I found it odd that there was so much emphasis on Hart and Williamson in the book. Chapter 4 on the “competitive dimension of transactions” (49) is entirely devoted to the different treatments by Hart and Williamson on specific investments and the problem of hold-up. The problem of hold-up is at best a special case of a transaction cost problem, and Hart and Williamson (though important players) are hardly all there is when it comes to incomplete contracts and asymmetric information. Conspicuous in their absence from the book are Armen Alchian, Yoram Barzel, and Stephen Cheung—the very ABC’s of property rights.

Which leads to the book’s big claim, found in the final two-page chapter, which states that the book leads to a ‘remarkable research proposal’. What is this proposal? That “future works should study these transaction costs theoretically and empirically” (79). For someone who first defined transaction costs in 19881 and who has worked (along with hundreds of others) on the theoretical and empirical content of transaction cost economics ever since, this research proposal came as a surprise. What have all the institutional economists been up to for the past 30 years?

The book is not without some merit. Many of the ideas of the old institutionalists and legal realists were solid, and there is often a strong correspondence with modern institutional ideas. For example, Hale’s work on legal rights can be linked quite readily to the works of Alchian, Barzel, and others. For those unfamiliar with the work of Commons, this book provides a very brief introduction. Unfortunately, it doesn’t go much beyond this.

REFERENCES

1 See Allen (1988, chapter 2; also, cf. 1991).


**Douglas W. Allen** is the Burnaby Mountain Professor of Economics at Simon Fraser University, Burnaby, British Columbia. He is the author of *The Institutional Revolution: Measurement and the Economic Emergence of the Modern World*.

Contact e-mail: <doug_allen@sfu.ca>

ERICA YU
Erasmus University Rotterdam

*Open Democracy* is a bold exploration of how we can move beyond a purely electoral conception of democratic representation. Using normative democratic theory and real-world examples of innovations in citizen representation, Hélène Landemore argues for a vision of democracy that is more faithful to popular rule, more likely to tap into democratic reason, and more stable and durable than electoral democracy.

The book begins with the all-too-familiar observation that “democracy is in crisis, or so we are incessantly told” (xiii). The symptoms of this crisis include a decline in voter turnout, the decline of parties as vehicles for mass participation, polarization, extremism, and populism (26–27). Landemore seeks to answer two important questions about this so-called crisis of democracy: How did we get here, and how do we get out? On the first question, Landemore provides a simple answer: representative democracy—characterized by electoral representation—has failed to deliver on the democratic promise of popular rule. On the second question, Landemore develops an ambitious proposal for an ‘open democracy’ which will bring power back to the people through novel (that is, non-electoral) forms of democratic representation based on random and self-selection.

After an introductory chapter, Landemore develops her claim that the crisis of democracy can be traced to representative democracy being designed on the basis of electoral premises. The crisis that we face, she argues, is not merely the result of external shocks such as globalization, technological change, or the rise of economic inequalities brought about by capitalism. After all, we see that representative democracies are not completely powerless in the face of these challenges.¹ More fundamen-

¹ Landemore argues that the fact that different policies lead to different outcomes in different countries shows that representative democracies have the capacity to deal with these external shocks (32).
tally, representative democracy seems to systematically fail to meet citizens’ expectations as to how these problems should be addressed. The crisis, then, seems to come from a problem which is internal to representative democracy.

“The main problem”, Landemore argues, “is that representative democracy was designed on the basis of electoral premises” (25). Elections separate out a set of individuals who are deemed ‘fit to rule’ from the rest of the citizenry. These individuals are then given opportunities to participate in collective deliberations and decisions about public policy, while the ordinary citizen is effectively excluded from these processes. Even with guarantees of full enfranchisement and no restrictions on who can run for office, representative democracy fails to deliver on the democratic promise of popular rule—defined as a regime where all members of society are equally able to participate in all processes of collective deliberation and decision about public policies (33).

This gives rise to two undesirable effects. First, basing representative democracy on elections “limits [its] epistemic potential” (25). This follows from Landemore’s epistemic argument for democracy: all things being equal, the rule of the many is at least as good as, and occasionally better than, the rule of the few at identifying the common good and providing solutions to collective problems (2012, 3). As electoral representation entails a ‘rule of the few’, its epistemic potential is limited.

Second, elections result in partisan politics which are not conducive to deliberation. The principle of periodic elections entails that parties have to regularly compete with each other and differentiate themselves in terms of the values and interests that they claim to promote. This competitive environment leads to partisanship and a lack of willingness to cooperate with other parties, which goes against the deliberative virtue of open-mindedness.

The solution, then, is to open up the entire decision process to all citizens. Shall we then turn to direct democracy, which is becoming easier to implement through digital technologies? In chapter 3, Landemore argues that we should not. Direct democracy is not able to deal with the scale problem of deliberative democracy: “deliberative decisions appear to be illegitimate for those left outside the forum, while bringing in more

---

2 Landemore (2012) argues that this is so because including more people in a decision-making process increases cognitive diversity, which has been shown to lead to better outcomes in problem-solving and prediction problems (Hong and Page 2004).

3 See Lindner and Aichholzer (2020) for a recent introductory overview of the theoretical and conceptual foundations of electronic democracy.
than a few people in would quickly turn the event into speech-making, not deliberation" (Parkinson 2003, 181; my emphasis). Representation is still essential in order to have deliberation that is both democratic and able to meet certain procedural standards. We do not need to dispense with representation, but to radically rethink it.

In chapters 4 and 5, Landemore explores the potentials of lottocratic and self-selected representation to deliver on the democratic promise of popular rule. Lottocratic representation is based on random selection: all citizens are given an equal chance to be selected as a representative through lotteries. Self-selected representation is based on individuals choosing to become a representative. Landemore argues that both lottocratic and self-selected representation are more democratic and can be at least as legitimate as electoral representation.

A form of representation is democratic to the extent that access to the assembly is inclusive and equal. Electoral representation, on the one hand, presents significant barriers to entry to positions of power by privileging a certain political elite. Lottocratic representation, on the other hand, rests on a strict principle of equality and impartiality between citizens (90): all have an equal chance to be chosen to represent their fellow citizens, no matter who they are. Even self-selected representation is more democratic than electoral representation as it is open in a way that electoral representation is not: anyone who wants to become a representative can become one.

A form of representation is legitimate to the extent that representatives are deemed morally entitled to make binding decisions for the rest of the polity. The legitimacy of electoral representation rests on citizens authorizing the transfer of their power to their representatives through the votes that they cast. Similarly, lottocratic and self-selected representation can be legitimized through the authorization of the mechanisms of selecting representatives by at least a majority of those to be represented.

Having shown that lottocratic and self-selected representation are more democratic and at least as legitimate as electoral representation, Landemore then formulates a new conception of representative democracy based on these new notions of representation in chapter 6 which she calls ‘open democracy’. Open democracy is defined according to five institutional principles. Participation rights ensure actual access to power instead of mere opportunities to provide or withdraw consent. Deliberation is a process which must involve all participants (either directly or indirectly), and not just gifted orators. The majoritarian principle ensures
that where deliberation fails to produce consensus, disagreements are able to be resolved. Democratic representation is not necessarily defined in electoral terms in open democracy, but it “embraces a richer ecology of various forms of democratic representation” (142) such as in lottocratic and self-selected terms. It is important to note, however, that the central form of democratic representation advocated for by Landemore is lottocratic. Finally, transparency ensures that citizens are aware of the responsibilities and duties of their representatives and can hold them to account.

Chapter 7 then explores how the Icelandic constitutional reform process from 2010 to 2013 shows how open democracy can be implemented, what its demonstrated benefits are, and what lessons we can learn. Most significantly, Landemore argues that the high levels of public participation through lottocratic representation and crowdsourcing resulted in more expansive rights provision in the drafted constitution as a greater number of interests and perspectives were taken into account.

Chapter 8 then deals with various challenges that can be put forward concerning the viability of open democracy. The most serious of these challenges include the feasibility of open democracy in big and heterogeneous countries, and the lack of competence and accountability of non-elected representatives.

Landemore concludes by offering a concrete vision of what an open democracy could look like:

[The open] mini-public is an all-purpose, randomly selected body open to the input of the larger public via citizen initiatives and rights of referral as well as a permanent online crowdsourcing and deliberative platform, and ultimately connected to a demos-wide referendum on central issues [...]. This general purpose mini-public would form the center of a network of other mini-publics, some of them single-issue, others generalist, operating at various sub-levels of the polity. Combined, they would form a web of connected mini-publics all staffed with randomly selected citizens. (218–219)

Landemore has written a thoughtful and engaging appraisal of representative democracy which engages with both ideal theory and not-always-ideal real-world cases. It is noteworthy how Open Democracy does not shy away from engaging with skeptics of both, normative theory and practical implementation. Landemore does not only want to convince those who already think that randomly-selected deliberative mini-publics
are essential for democracy, but also those who stand by elections as the constitutive feature of democracy.

Landemore coherently and convincingly argues that representative democracy based on electoral representation fails to deliver on the democratic promise of popular rule, as its fundamental design is based precisely on drawing a distinction between the rulers and the ruled. That being said, there are two concerns which call into question whether lottocratic representation—which Landemore argues should be the central form of representation in open democracy—is indeed better able than electoral representation to achieve popular rule and enjoy its purported benefits.

The first point concerns Landemore’s argument for lottocratic representation being more democratic than elections: “Given enough rotation and a small enough population, actual access to power is strictly equalized over the long term” (90). However, as Landemore rightly points out, this implies that lottocracy’s advantage over elections depends on there being a large number of seats available and frequent enough rotation to ensure that the chance of being selected as a lottocratic representative is greater than the chance of the average citizen being elected into office (91). Her solution is then to increase opportunities to be selected for a lottocratic assembly by combining a lottocratic assembly at the national level with the creation of a multiplicity of local assemblies that set the agenda for the national one (92).

However, Landemore now needs to weigh the benefits of achieving actual (and not merely hypothetical) equalization of access to power with the costs entailed by putting up these multiple localized assemblies. If, for example, the marginal costs of putting up a lottocratic localized assembly exceed its marginal benefits in terms of its contribution to the quality of collective decisions, would it still be worth pursuing in order to achieve actual equalization of access to power? If Landemore argues that it is, then her commitment to the instrumental value of democracy in terms of its epistemic benefits is compromised. If she argues that it is not, then the case for lottocratic representation being more democratic than electoral representation is weakened as there is no guarantee that the chance of being selected as a lottocratic representative is always greater than the chance of the average citizen being elected into office.

The second point is that lottocratic representation may fail to bridge the distance between the rulers and the ruled because these forms of representation still do not address the underlying lack of interest of citizens
to engage in making decisions for the rest of the polity. Some of the symptoms of the crisis of democracy that Landemore points to include declining voter turnout and party membership. Thus, citizens are already choosing not to exercise their already little power over collective decisions in an electoral democracy. Simply giving them more opportunities to influence these decisions does not mean that they will choose to undertake the responsibilities of representatives.

This has been illustrated in the case of an experiment done in Rotterdam where neighborhood councils were formed by selecting representatives by lot (van Buuren et al. 2020). These councils were supported by civil servants who were tasked with providing information and advice to the representatives. A key observation in this experiment was that the civil service support ended up doing most of the actual decision-making work on behalf of these lottocratic representatives because of their lack of confidence in their knowledge, skills, and connections to make these decisions themselves (van Buuren et al. 2020, 26).

Perhaps lottocracy is merely a ‘shortcut’ to a democracy which is able to deliver on its promise of popular rule, as Lafont (2019) argues. Though Landemore is right to move away from Lafont’s push for full mass participation, she still has to engage with Lafont’s key claim that a necessary condition for popular rule (or, in Lafont’s words, “self-government”) entails that “all citizens can equally own and identify with the institutions, laws, and policies to which they are subject” (Lafont 2019, 3; my emphasis). Without empowering citizens to actually take charge of collective decisions, lottocratic and self-selected representation are not able to guarantee this.

To conclude, Open Democracy provides a much-needed radical rethinking of democratic representation. It engages deeply with normative democratic theory and real-world examples of innovations in democratic representation. However, more needs to be done in order to argue that open democracy—with lottocratic representation at its core—delivers on the promise of popular rule to a greater extent than electoral democracy.

---

4 Whether or not this lack of interest and engagement is due to electoral democracy is beside the point. The mere availability of opportunities to participate in collective decision-making is not enough to guarantee that people will actually participate.

5 This case is particularly interesting as these lottocratic representatives were given access to the relevant expertise, but still chose to delegate their decision-making power to these experts instead of being empowered to make the decisions themselves.
REFERENCES


Erica Yu is a PhD candidate in philosophy at Erasmus University Rotterdam. Her research interests lie in the intersection of social choice theory, deliberative democracy, and political representation.

Contact e-mail: <yu@esphil.eur.nl>
PHD THESIS SUMMARY:
Galbraith’s Integral Economics (1933–1983)

ALEXANDRE CHIRAT
PhD in Economics, November 2020
Université Lumière Lyon 2, Laboratoire Triangle

My dissertation studies the genesis, construction, and reception of John Kenneth Galbraith’s integral economics. This term refers to his theoretical project—thought of as an alternative to conventional economics—which proposes integrated ‘pattern models’\(^1\) of the functioning of the economic system of post-war American industrial society.\(^2\) Galbraith’s notion of “conventional economics” combines highly diverse economic analyses. But they share three core postulates: (i) the hypothesis of consumer sovereignty, (ii) the hypothesis of citizen sovereignty, and (iii) the hypothesis of profit maximization (Galbraith 1973a, 5). These postulates lead to the exclusion of power outside economics; and it is against these that Galbraith has built his own theories of corporation, competition, and consumption. My dissertation studies these issues in four separate parts.

The first part of the dissertation accounts for Galbraith’s participation in original institutional economics from intellectual, theoretical, and epistemological points of view.\(^3\) This allows me to situate his integral economics within the secular “struggle” (Yonay 1998) between original institutional economics and neoclassical economics (Rutherford 2011) and to show that his theory of the corporation, which lies at the center of his

---

\(^1\) A ‘pattern model’ is a holistic, systemic, and evolutionary model. The term characterizes the institutionalist methodology:

Thus, institutionalism is holistic because it focuses on the pattern of relations among parts and the whole. It is systemic because it believes that those parts make up a coherent whole and can be understood only in terms of the whole. It is evolutionary because changes in the pattern of relations are seen as the very essence of social reality. (Wilber and Harrison 1978, 71)

\(^2\) From a cross-reading of the works of Raymond Aron and John Kenneth Galbraith, I have shown that Galbraith is one of the main theorists of the concept of industrial society (Chirat 2019).

\(^3\) The fact that I consider Galbraith as an institutionalist rather than a Keynesian explains many of the differences between my reconstruction and those by Dunn (2011), Laguérodie (2007), and Parker (2005).
pattern models, is heir to the theories developed by Veblen (1904), Clark (1923), Berle and Means ([1932] 1991), and Berle (1959).

The second part examines Galbraith’s period of intellectual development (1933–1952). Academically, Galbraith quickly switched from agricultural economics at Berkeley to the study of the whole economic system at Harvard (Galbraith 1936, 1939, [1948] 1956). He endeavored to combine the insights of what he considered to be the three revolutions in economics of that period, namely Berle and Means ([1932] 1991), Chamberlin (1933), and Keynes ([1936] 2013). But his theoretical project was also nourished by the practical lessons of a decade of extra-academic experiences, notably with the Office of Price Administration (Galbraith 1947, 1952), the United States Strategic Bombing Survey, and *Fortune Magazine*.

The third part, drawing on new archival materials as well as those published by Holt (2017), focuses on the construction of Galbraith’s American trilogy, which forms the core of his integral economics (1952–1967). I show that *American Capitalism* ([1952] 1980) is constructed as a reappropriation of the works in the emerging field of industrial organization at Harvard. It provides a “bimodal model” of the American economy, divided into a competitive part and an oligopolistic one. The latter is analyzed in terms of “countervailing power” which is presented as an alternative to the classical competitive model (Galbraith [1952] 1980, 108). I then demonstrate how various specialized works lead Galbraith to envisage the writing of “a new treatise on political economy”. The *Affluent Society* (1958) conveys this ambition by providing a theory of consumption based on the rejection of the principle of consumer sovereignty (Chirat 2020a). *The New Industrial State* (1967), Galbraith’s masterpiece in terms of his ambition to provide an alternative to the neoclassical synthesis, provides a theory of both the entrepreneur and the modern corporation, which were missing from the two earlier works (Baudry and Chirat 2018).

---

4 Galbraith considered that Berle and Means’ analysis of the separation between ownership and control in modern corporations directly challenges the relevance of the neoclassical theory of the firm. He reappropriated Chamberlin’s analysis in terms of oligopoly and monopolistic competition to challenge the classical competitive model as well as welfare economics. He praised Keynes’ work for providing theoretical justification for reasoning in terms of underemployment equilibrium and thus undertaking macroeconomic public policies.

5 For a reconstruction of the emergence of the so-called Harvard tradition in industrial organization, see Chirat and Guicherd (2021).

6 Unpublished manuscript, John Kenneth Galbraith personal papers, JFK Library, Series 9, Box 750.
The fourth and final part focuses on the impact of Galbraith's integral economics as a paradigm. I show how Galbraith (1970, 1973a, 1973b) strives to perfect a theoretical project that generates both enthusiasm and controversy at the precise moment when economics enters its “second crisis” (Robinson 1972), leading to a schism with the creation of the Association for Evolutionary Economics and The Union for Radical Political Economy. While he manages to integrate into a consistent whole his theories of corporation, competition, and consumption through a general theory of power (Galbraith 1983), I conclude that his integral economics fails to meet the requirements of a scientific revolution since it fails to establish itself as a dominant framework in the discipline. External reasons explain this failure, such as the fact that his integral economics runs counter to the double movement of specialization and formalization in post-war economics. But internal reasons play a role too since the historical nature of Galbraith's pattern models make them hard to replicate.

My dissertation is not merely a study of Galbraith's project itself. I have also used his integral economics as a new lens for looking at the dynamics of American economics between the 1930s and the 1970s. First, I argue that Galbraith's early works both illustrate and support the thesis of “interwar pluralism” (Morgan and Rutherford 1998, 3–4); an argument that does not contradict the recognition of Galbraith’s involvement in the secular struggle between institutionalist and neoclassical economists (Yonay 1998).

Second, I demonstrate how Galbraith, during the construction stages of his post-war alternative project, drew on the thinking of various economists. For instance, his theory of the corporation borrows from the managerial theory of the firm of Baumol (1959) and Marris (1964) as well as the behavioral theory of the firm of Simon (1962). Despite their epistemological and political disputes, Galbraith’s pattern model of the planning system draws on Theodore Schultz’s (1961) pioneering economic analysis of education. In brief, I explain that his trilogy is, without a doubt, heterodox regarding its theoretical and epistemological dimensions. This heterodoxy notwithstanding, an essential factor in the production of his integral economics was the perpetuation of forms of pluralism

---

8 For more details on the convergence between Baumol and Galbraith, see Chirat (2020b).
9 For more details on the convergence between Schultz and Galbraith, see Chirat and Le Chapelain (2020).
until the 1970s, notably through the acceptance of variety and the maintenance of dialogue between economists with different methods or objectives.

This partial maintenance of pluralism in the post-war period obviously did not prevent Galbraith from entering into controversies with economists as diverse as Demsetz, Friedman, Hayek, Heller, Meade, Samuelson, Solow, and Sweezy. The careful study of the debates generated by the publication of each opus of his trilogy was particularly interesting. First, in aiming at providing a history of American thought through the lens of Galbraith’s project, it lends a voice to a plurality of points of view among the profession. Second, the study of these controversies reveals the emergence and dynamics of the fault lines within economics. Finally, as I propose a reconstruction of Galbraith’s work that is both historical and rational, his integral economics also constitutes a prism for understanding the changes in twentieth-century capitalism that it intends to report on.

REFERENCES


**Alexandre Chirat** obtained his PhD in Economics at the Lumiè re University Lyon 2 (TRIANGLE). He is currently working on the publication of his thesis as a book for Classiques Garnier and planning an English translation. His research in the history of economic thought mainly focuses on twentieth-century American economics, theories of the firm, and the ‘managerial revolution’. In parallel, he has begun a new research agenda on the political economy of populism and democracy. Alexandre currently teaches at the University of Besançon (CRESE). Contact e-mail: <chirat.alexandre@gmail.com>