

A Theory of Chosen Preferences

By B. DOUGLAS BERNHEIM, LUCA BRAGHIERI, ALEJANDRO MARTÍNEZ-MARQUINA, DAVID ZUCKERMAN*

We propose and develop a dynamic theory of endogenous preference formation in which people adopt worldviews that shape their judgments about their experiences. The framework highlights the role of mindset flexibility, a trait that determines the relative weights the decision maker places on her current and anticipated worldviews when evaluating future outcomes. The theory generates rich behavioral dynamics, thereby illuminating a wide range of applications and providing potential explanations for a variety of observed phenomena.

JEL Codes: D11, D90, D91

Keywords: Chosen preferences, endogenous preferences, mindset flexibility, time inconsistency

The objective of this paper is to make progress toward answering one of the most critical open questions in Economics: *What determines preferences?* The notion that people are endowed with exogenously fixed preferences plays a central role in classical choice theory, but many critics of the classical paradigm, both from within and outside the field, have expressed skepticism concerning that premise. For example, Bowles (1998) emphasizes the role of social influences and likens the process of preference formation to the acquisition of an accent. Recent empirical research implies that preferences respond to both opportunities and experiences.¹

In defending the classical paradigm, some point out that it posits fixed *intertemporal* preferences, rather than fixed *momentary* preferences. For example, it can accommodate the possibility that the repeated consumption of pizza can

* Bernheim: Stanford University. bernheim@stanford.edu. Braghieri: Stanford University. lucabrag@stanford.edu. Martínez-Marquina: Stanford University. amm15@stanford.edu. Zuckerman: Stanford University. dzuckerm@stanford.edu. We would like to thank seminar participants at Harvard University, NYU, Northwestern University's Kellogg Graduate School of Management, Princeton University, Stanford University, University of Toronto, and UCSB, as well as three referees for helpful comments.

¹Various branches of the literature document how preferences change in response to opportunities (e.g., Cohn et al., 2015), traumatic events (e.g., Voors et al., 2012; Callen et al., 2014; Cameron and Shah, 2015; Danzer, Danzer and Fehr, 2016; Eckel, El-Gamal and Wilson, 2015; Hanaoka, Shigeoka and Watanabe, 2018; Guiso, Sapienza and Zingales, 2018), cultural influences (e.g., Ichino and Maggi, 2000; Guiso, Sapienza and Zingales, 2006; Benjamin, Choi and Strickland, 2010; Atkin, 2013; Cohn, Fehr and Marechal, 2014; Atkin, Colson-Sihra and Shayo, 2019), family influences (e.g., Fernández, Fogli and Olivetti, 2004; Washington, 2008), and institutional influences (e.g., Nunn and Wantchekon, 2011; Gruber and Hungerman, 2008; Fouka, 2020). For general discussions, see Bowles (1998) and Fehr and Hoff (2011).

lead a consumer either to develop a taste for pizza or to become sick of it.² Still, the classical paradigm cannot accommodate other potentially important forms of preference endogeneity, including the possibility that people exercise a degree of *direct* control over their tastes, and not merely *indirect* control through their consumption choices. Highly visible manifestations of deliberate preference selection include the adoption of philosophies and religions, as well as participation in certain forms of psychotherapy.

This paper proposes and develops a dynamic theory of preference formation in which people adopt *worldviews* that shape their judgments about their experiences. In other words, we endogenize preferences by making them the objects of constrained choice. We are certainly not the first to follow this path; see, for example, the discussion of Akerlof and Kranton (2000), Palacios-Huerta and Santos (2004), and other related papers in Section IV. We assume that society presents each of its members with a collection of *pure worldviews*, and that each individual can choose to subscribe to one or more of those perspectives; in the latter case, she holds a *mixed worldview* that evaluates outcomes according to a weighted average of the component worldviews.

The most novel aspect of our framework is that it highlights the role of a trait we call *mindset flexibility*. In settings with changing preferences, people can evaluate future outcomes either according to their current worldview, or according to the worldview they expect to hold when outcomes materialize.³ Mindset flexibility determines the relative weights attached to the current and future perspectives. Intermediate degrees of mindset flexibility give rise to time inconsistency, and (as we demonstrate) generate rich behavioral dynamics, thereby illuminating a wide range of applications and providing potential accounts for a variety of observed phenomena.

We begin by providing a collection of results concerning the dynamics of preference formation. In a simple infinite-horizon setting with two worldviews and two actions, we identify two cases that encompass the set of behavioral patterns potentially exhibited by sophisticated decision makers.

- In the first case, the consumer immediately jumps to one of the two pure worldviews. Both pure worldviews are absorbing states, so she makes no further changes. Whether she jumps to one worldview or the other depends on her initial mixed worldview.
- In the second case, the consumer moves gradually toward one of the pure worldviews. In other words, despite the absence of any new information, she changes her mind slowly, but not by enough to modify her behavior. Eventually she reaches a critical threshold at which she changes her behav-

²In that spirit, Becker and Murphy (1988) model the taste changes that accompany addiction by positing intertemporal complementarities.

³In principle, they could also evaluate those outcomes based on preferences they expect to hold in other future periods, or that they have held in the past. We addressed these possibilities in an earlier version of this paper (Bernheim et al., 2019).

ior and completes the transition to a pure worldview. Thus, our theory generates “changes of heart” that result from contemplation of non-trivial duration. Moreover, as in the first case, the consumer ultimately settles on a “purist” perspective.

We identify conditions under which the multiplicity of absorbing states present in the first case generalizes beyond models with two worldviews and two actions. This property proves important in some of our applications. We also briefly discuss naive behavior. In settings with multiple pure worldviews, we show that naive agents can act as “serial true believers,” leaping from one pure worldview to the next in quick succession until they find one that sticks.

A society’s menu of pure worldviews can expand either due to contact with other cultures, or to worldview invention and entrepreneurship. We identify two interesting patterns that can arise with the introduction of new worldviews. First, the addition of extreme worldviews can generate reactionary responses. In a setting where people are willing to migrate from a “conservative” worldview to a “moderate” one, the introduction of a “liberal” worldview (say from another culture) can harden traditionalism. Second, the addition of an intermediate worldview can provide a gateway that facilitates the migration of the population from one extreme to another. Thus, in a setting where people are unwilling to migrate from a “conservative” worldview to a “liberal” one, the introduction of a “moderate” worldview can make this transition possible.

After exploring the dynamics of preference adjustment, we turn to applications. The list of applications explored in this paper is reasonably long because we hope to convey the value of looking at a wide variety of applied phenomena through the lens of our theory, and to suggest that the theory has the potential to provide considerable unification. The following is a brief synopsis:

- *The origins of closed-mindedness.* While much of our analysis treats mind-set flexibility as a fixed personality trait, we acknowledge that different worldviews can potentially prescribe different degrees of open-mindedness or closed-mindedness. Accordingly, as an extension, we endogenize mind-set flexibility. We demonstrate that people are naturally drawn to inflexible outlooks because inflexibility helps them resolve the core time-inconsistency problem.
- *Social influences, conformity, and assimilation.* According to the leading economic theories of conformity, social equilibrium either homogenizes beliefs or promotes a shared understanding of social conventions and norms. Many psychologists, sociologists, and anthropologists envision a separate route to conformity involving the convergence of preferences. One leading school of thought traces its routes to Festinger (1954), who emphasized the role of “tension systems.” We show that our theory provides a formal account of the dynamics Festinger envisioned.

- *The “sour grapes effect” and related phenomena.* According to our theory, compelling or inducing a consumer to take a particular action for a limited period of time leads her to rationalize her experience through the choice of a consonant worldview, thereby generating a permanent change in preferences, and hence a durable change in behavior, that persists long after the inducement or compulsion is removed. Additionally, when the duration of compulsion is uncertain, people initially display a degree of resistance, acquiescing gradually. These principles have multiple applications.
- *Large-scale and small-scale risk aversion.* One of the best-known puzzles in behavioral economics concerns the incompatibility of risk aversion in observed small-stakes and large-stakes gambles. We show that our theory can resolve this puzzle by largely decoupling large-stakes risk preferences from small-stakes risk preferences.
- *Polarization.* Our theory provides a potential explanation for the phenomenon of polarization. Within our framework, there is a robust tendency for people to adopt pure worldviews that credit only one point of view, rather than mixed worldviews that credit many. Polarization follows directly from this “purist” tendency.

We organize the paper as follows. Section I sets forth the basic model. Section II explores the dynamics of preference formation. Section III investigates applications. Section IV discusses the relationship between our theory and other analyses of endogenous preferences found in the literature. We adopt the somewhat unconventional practice of discussing the literature in detail after presenting our analysis, rather than before, because this sequencing makes the nature of our contribution more apparent. However, we also mention some of the most closely related papers in context during the course of developing our theory. Section V offers some brief conclusions and directions for subsequent research. All proofs appear in an online appendix.

I. The Basic Model

We envision a consumer who lives for T periods, where T may be either finite or infinite. In each period t , her task is to select an action x_t from a set $X_t \subset \mathcal{X}$, where \mathcal{X} contains all possible consumption options. The manner in which she evaluates these options depends on her worldview – in other words, on the way in which she defines “the good life,” and how she differentiates between success and failure. The central tenet of our theory is that people exercise control over their worldviews, and that they adopt the perspectives they consider most conducive to their lifetime well-being. Of course, their evaluations of lifetime well-being depend in turn on their worldviews. The mutual causality between the selection of worldviews and the evaluation of well-being lies at the core of our theory.

We assume society presents the consumer with a fixed menu of *pure worldviews*, J . One can think of a pure worldview as akin to a philosophy or religion, though

we ultimately adopt more prosaic interpretations for some of our applications. For the moment, we take the set of possible worldviews to be finite: $J = \{1, \dots, n\}$, but the theory extends in a natural way to settings with infinitely many potential worldviews, and indeed some of our applications have this feature. When the consumer holds pure worldview j , she evaluates her momentary experiences according to the utility function $u_j : X \rightarrow \mathbb{R}$.⁴ To be clear, the same worldview may induce different utility functions for different consumers, depending on the fit between the principles of the worldview and the individual's internal wiring. Throughout this paper, we assume the consumer discounts utility from future experiences at the fixed rate δ .⁵

We allow for the possibility that consumers can subscribe to multiple worldviews simultaneously. For this purpose, we define a *mixed worldview* as a vector of utility weights, $\alpha = (\alpha^1, \dots, \alpha^n) \in \Delta^n$ (the unit simplex).⁶ When the consumer holds mixed worldview α , she evaluates momentary experience according to the utility function $U(\alpha, x) = \sum_{j \in J} \alpha^j u_j(x)$. Naturally, we can describe pure worldview j as a mixed worldview, $\alpha(j)$, such that $\alpha^j(j) = 1$ and $\alpha^i(j) = 0$ for $i \neq j$.

Our theory proceeds from the following four premises concerning the choice of worldviews.

First, we assume that the set of possible worldviews is limited to $\alpha \in \Delta^n$.⁷ We see constraints on worldviews as arising from two sources. One is neurobiological wiring: for example, even if we try to adopt the worldview that “nothing matters,” we likely will not end up with a utility function of the form $u_j(x) = C$ for all $x \in X$, where C is some constant. Each utility function u_j is a reflection of our motivational systems, and evolution has shaped those systems to ensure patterns of attraction and aversion that promote our survival. Evolution may preclude the full internalization of the “nothing matters” worldview because that perspective inherently defeats the biological purpose of our motivational systems.⁸ Likewise, evolution may limit our ability to derive enjoyment from consequences that are inherently inimical to survival. Still, a degree of flexibility to adjust our motivations, within bounds, in response to changing circumstances may itself be a useful adaptation (see Rayo and Becker, 2007, for a related evolutionary argument). The second constraint is that, because worldviews are complicated objects, it is difficult to invent them. We think of worldviews as collections of

⁴Accordingly, our notion of a worldview is closely related to Akerlof and Kranton's (2000) concept of a “narrative,” which they relate to identity. In our theory, a worldview is not necessarily connected to identity.

⁵As we mention below, incorporating the consumer's degree of patience into the chosen worldview is possible, but adds complexity.

⁶We use superscripts to denote the components of α so we can use subscripts to denote time.

⁷Our theory is also applicable to settings in which the set of feasible worldviews is a subset of Δ^n , as it might be, for example, if two pure worldviews clashed irreconcilably. The current paper does not examine applications involving such restrictions.

⁸Indeed, Rayo and Becker (2007) argue that evolution favors utility functions that exhaust the feasible range of motivational responses, in order to promote the greatest possible discernment.

principles, the adoption of which leads the consumer to conclude that certain options are better than others. Consumers may be understandably reluctant to embrace worldviews that encompass sufficiently incompatible principles, or at least averse to the resulting cognitive dissonance (Festinger, 1954), and complete internal consistency is difficult to achieve. In our theory, society offers a menu of prefabricated worldviews to its members, most of whom limit themselves to combinations of those options. Of course, successful innovation sometimes occurs, and our theory will allow us to examine the effects of injecting new worldviews into a social system (see Section II.B).

Second, we assume that the internalization of a new worldview is not instantaneous. We cannot decide to live our lives by different principles immediately. Instead, the process of self-acclimatization and indoctrination to new principles takes time. For simplicity, we assume the consumer chooses the period t mixed worldview in period $t - 1$. Naturally, one could introduce more elaborate adjustment processes. The length of these periods likely depends on the application: it is reasonable to assume that people can adjust minor aspects of their worldviews more rapidly than core principles.

Third, we assume the consumer evaluates her current experience according to her current worldview. She does not attach importance to old worldviews she has discarded, nor is she concerned about her future attitudes toward memories of her current experience. In an earlier version of this paper (Bernheim et al., 2019), we relaxed the latter restriction, and showed how the accumulation of “memory capital” can lead consumers to become “set in their ways” as they age (because a change in worldviews degrades the value of memory capital).

Our fourth premise pertains to a critical question that lies at the core of our analysis: how does the consumer evaluate the future experiences she will have after changing her worldview? In two-period settings ($T = 2$), we can depict the range of possibilities comprehensively by assuming that, in period 1, she evaluates period-2 experience by applying some blend of her period-1 and period-2 worldviews. In other words, we assume she cares about the feelings she expects to have in the future, but also judges the experience that produces those feelings based on her current outlook. Formally, she evaluates the trajectory $\sigma_1 = ((\alpha_1, x_1), (\alpha_2, x_2))$ according to the following aggregator, V :

$$V(\sigma_1) = U(\alpha_1, x_1) + \delta[(1 - \lambda)U(\alpha_2, x_2) + \lambda U(\alpha_1, x_2)]$$

We interpret $\lambda \in [0, 1]$ as a *mindset flexibility* parameter. When $\lambda = 1$, the consumer is *perfectly mindset inflexible*, in the sense that she evaluates the future according to her current worldview. As an example, consider someone who is both highly religious and extremely unhappy. Upon noticing that heathens are generally much happier, she considers giving up her faith. However, she rejects this course of action because she recognizes that she would be deriving happiness from actions she currently regards as sinful, even though she understands that upon becoming a heathen she would have no objection to those actions. In contrast,

when $\lambda = 0$, the consumer is *perfectly mindset flexible*, in the sense that she evaluates the future according to her future worldview. As an example, someone who is bored and looking for a new hobby may “cultivate an interest” (for example, in stamp collecting) precisely because she wishes to become emotionally engaged with, and thereby derive pleasure from, an activity that she may initially enjoy only to a limited extent. The cases of greatest interest lie between the extremes of perfect mindset flexibility and inflexibility.

Modeling decision problems with longer T -period horizons requires a generalization of the two-period formulation. The simplest extension, which we adopt throughout the current paper, assumes that in period $t < t'$, the consumer attaches weight $1 - \lambda$ to period t' well-being as experienced under the period t' worldview, and weight λ to period t' well-being evaluated according to her period t worldview. Thus, in period t , she evaluates the continuation trajectory $\sigma_t = ((\alpha_t, x_t), (\alpha_{t+1}, x_{t+1}), \dots)$ according to the following aggregator, V_t :⁹

$$(1) \quad V_t(\sigma_t) = U(\alpha_t, x_t) + \sum_{k=1}^{T-t} \delta^k [(1 - \lambda)U(\alpha_{t+k}, x_{t+k}) + \lambda U(\alpha_t, x_{t+k})]$$

In effect, the preceding specification assumes that the consumer does not care about any future utility derived either from anticipation or memory. As a result, her period- t evaluation of her anticipated period- t' experience (for $t' > t$) depends only on the enjoyment she expects to derive from actions taken in period t' and her current judgment of those actions.

While appealing (at least as a reasonable starting point), this formulation excludes some interesting possibilities. For example, it cannot accommodate chosen worldviews that encompass patience (δ). In an earlier version of this paper (Bernheim et al., 2019), we proposed an alternative formulation of preferences with a recursive structure, in which the consumer places weight on her future views of continuation utility, rather than instantaneous utility. We then used that formulation to derive an explanation for the greater patience exhibited by the Depression cohort (Kapteyn, Alessie and Lusardi, 2005).

For most of our analysis, we treat λ as an exogenous personality characteristic. However, it is important to acknowledge that λ may differ from one application to the next, as our motivating examples of religion and postage stamp collection suggest. Moreover, many philosophies and religions take positions on the desirability of open-mindedness. Endogenizing λ is therefore of interest. We consider this possibility as an extension in Section III.A.

Critically, any $\lambda > 0$ implies that the consumer is time-inconsistent. That property plays a featured role in our analysis. As in other contexts, the implications of

⁹Readers may notice a resemblance between this formulation and models of projection bias (see Loewenstein, O'Donoghue and Rabin, 2003). From a formal perspective, the key difference is that consumers freely choose the stream of mixed worldviews in our model. Also, we do not interpret the model as depicting false beliefs about future attitudes or tastes.

time inconsistency differ according to the nature of consumers' expectations about their own future behavior. We distinguish between naive consumers, who assume incorrectly that their desired trajectory of worldviews will actually materialize (but who correctly account for the effect of worldview on the chosen action),¹⁰ and sophisticated consumers, who correctly anticipate their own future worldviews and choices. Sophisticated behavior corresponds to the Markov-perfect equilibrium of an intertemporal game where each "self" essentially chooses its successor. For the most part (but not exclusively), the current paper concerns sophisticated agents. An earlier version of this paper (Bernheim et al., 2019) contains additional results concerning naivete.

As formulated, our model of preferences assumes that changing worldviews is psychologically costless. As an extension, one could easily incorporate psychic costs of self-indoctrination, and indeed we explore this possibility in an application involving risk aversion; see Section III.D. It is worth noting, however, that one can also interpret our simple model as depicting a consumer with perfect mindset flexibility who incurs costs when changing worldviews that are proportional to the implied changes in evaluations.¹¹

Because preferences formally encompass subjective beliefs, it is natural to wonder whether one can also interpret our theory as describing the dynamics of processes involving chosen beliefs (Brunnermeier and Parker, 2005, Benabou and Tirole, 2016). While we do not in principle object to this interpretation, some caveats are in order. First, in our framework, there is no counterpart to *correct* beliefs, which effectively anchor chosen beliefs in Brunnermeier and Parker (2005). As a result, the theory predicts the eventual adoption of "purist" perspectives which, under the "chosen belief" interpretation, amounts to subjective certainty. That implication may be implausible in many applications involving beliefs. Second, equation (1) may not be the most natural formulation of intertemporal preferences under the "chosen belief" interpretation. The usual rationale for modeling preferences over beliefs is that people experience anticipatory emotions. In that case, it would seem a bit odd to assume that, regarding an experience in some future period t' , the consumer cares only about emotions experienced in t' and the current period t , but not those experienced in any intervening period ($t'' : t < t'' < t'$). These concerns do not necessarily arise when one interprets α as non-probabilistic preference weights.

¹⁰Our definition of naivete is not quite the same as O'Donoghue and Rabin's (1999), because we do not assume the consumer necessarily expects to have unchanging preferences. However, the two concepts are closely related.

¹¹Specifically, we can rewrite the consumer's utility as $V_t(\sigma_t) = U(\alpha_t, x_t) + \sum_{k=1}^{T-t} \delta^k U(\alpha_{t+k}, x_{t+k}) - \lambda \Gamma_t(\sigma_t)$, where $\Gamma_t(\sigma_t) = \sum_{k=1}^{T-t} \delta^k [U(\alpha_{t+k}, x_{t+k}) - U(\alpha_t, x_{t+k})]$ represents the costs of future changes in worldview.

II. The Dynamics of Preference Adjustment

The preceding models raise a variety of general questions about the dynamics of preference adjustment. Will people ultimately gravitate toward “purist” perspectives? Will they arrive at stable worldviews in a single step, or through gradual adjustment? Will their actions change immediately, or must new ideas “percolate,” with behavior changing after non-trivial periods of contemplation? Will people draw bright lines because they fear “slippery slopes”? How are the dynamics of preference formation affected by the introduction of new worldviews? We explore each of these issues in this section.

A. Some basic results

For the moment, we will focus on infinite horizon models ($T = \infty$) with two actions ($X_t = \{1, 2\} \forall t > 0$) and two possible pure worldviews ($J = \{1, 2\}$). We will assume that the worldviews imply different optimal actions, and label the alternatives so that action i is optimal under worldview i . We will also assume that worldview 2 offers the “happier” perspective in the sense that $u_2(2) > u_1(1)$. Because there are only two pure worldviews, we will simplify the notation by using α in place of α^1 .

Under our assumptions, there exists $\bar{\alpha} \in (0, 1)$ such that the consumer chooses action 1 when $\alpha > \bar{\alpha}$ and action 2 when $\alpha < \bar{\alpha}$ (equality implies indifference). The following formula is easily checked:

$$\bar{\alpha} = \frac{u_2(2) - u_2(1)}{[u_2(2) - u_2(1)] + [u_1(1) - u_1(2)]}$$

In this setting, a (pure-strategy) *Markov-perfect equilibrium* (MPE) is a subgame-perfect Nash equilibrium in which period- t choices depend only on the state variable, α_t . Formally, consider a sequence of functions (ϕ_t, z_t) for $t \in \mathbb{Z}^+$ (the positive integers) such that $\phi_t : [0, 1] \rightarrow [0, 1]$ maps the period- t mixed worldview (α_t) into the chosen worldview for period $t+1$ (α_{t+1}), and $z_t : [0, 1] \rightarrow X$ maps the period- t worldview (α_t) into the chosen action for period t (x_t). For any such sequence and for each period s , we can define a function $\tilde{V}_s(\alpha_{s+1}, x_s; \alpha_s)$ that maps the choice of the next period’s mixed worldview and the current period’s action into aggregate utility, conditional on the current period’s mixed worldview, under the assumption that (ϕ_t, z_t) will govern choices for all future periods $t > s$.¹² Then $((\phi_1, z_1), (\phi_2, z_2), \dots)$ is a Markov-perfect equilibrium iff $(\phi_s(\alpha_s), z_s(\alpha_s))$ maximizes $\tilde{V}_s(\alpha_{s+1}, x_s; \alpha_s)$ for all s and α_s . It is immediate that, in any MPE, we

¹²Specifically, for each s , α_s , α_{s+1} , and x_s , define $\alpha'_{s+2} = \phi_{s+1}(\alpha_{s+1})$, and then recursively $\alpha'_{s+k+1} = \phi_{s+k}(\alpha'_{s+k})$ for all $k > 1$ (where we have suppressed the dependence of α'_{s+k+1} on s , α_s , α_{s+1} , and x_s for the sake of notational compactness). Also define $x'_{s+1} = z_{s+1}(\alpha_{s+1})$, and $x'_{s+k} = z_{s+k}(\alpha'_{s+k})$ for $k > 1$. Let $\sigma'_s(x_s, \alpha_s, \alpha_{s+1}) = ((\alpha_s, x_s), (\alpha_{s+1}, x'_{s+1}), (\alpha'_{s+2}, x'_{s+2}), \dots)$. Then $\tilde{V}_s(\alpha_{s+1}, x_s; \alpha_s) \equiv V_s(\sigma'_s(x_s, \alpha_s, \alpha_{s+1}))$.

must have $z_t(\alpha) \in z^*(\alpha)$, where z^* is the static optimal action correspondence.¹³ A *stationary* MPE has the property that $(\phi_t, z_t) = (\phi_{t'}, z_{t'})$ for all t, t' .

Our characterization of Markov-perfect equilibria distinguishes between the following two cases.

- *Case 1:* $u_1(1) > u_2(1)$
- *Case 2:* $u_1(1) < u_2(1)$

For Case 2, worldview 2 leads to greater happiness irrespective of the action taken. In other words, it “happiness-dominates” worldview 1. For Case 1, neither worldview happiness-dominates the other.

Our characterization of stationary MPE for Case 1 references the following threshold weight:

$$\alpha^* = \bar{\alpha} + \left(\frac{1 - \lambda}{\lambda} \right) \left(\frac{u_2(2) - u_1(1)}{[u_2(2) - u_2(1)] + [u_1(1) - u_1(2)]} \right) > \bar{\alpha}$$

PROPOSITION 1: *Consider Case 1, and assume $\lambda \in [0, 1)$.¹⁴ A stationary MPE exists. Furthermore, the Markov policy function governing the evolution of worldviews takes the following form.¹⁵*

$$\phi(\alpha) = \begin{cases} 0 & \text{if } \alpha < \alpha^* \\ 1 & \text{if } \alpha > \alpha^* \\ 0 \text{ or } 1 & \text{if } \alpha = \alpha^* \end{cases}$$

As λ approaches unity (perfect mindset inflexibility), α^* converges to $\bar{\alpha}$, reducing the set of initial conditions for which the consumer chooses the happier worldview. As λ declines (greater mindset flexibility), α^* increases, expanding that set. For sufficiently small λ , we have $\alpha^* > 1$, which means the consumer chooses $\alpha_{t+1} = 0$ (all weight on worldview 2) regardless of α_t . A bit of algebra reveals that $\alpha^* < 1$ if and only if

$$(2) \quad \lambda u_1(2) + (1 - \lambda)u_2(2) < u_1(1),$$

which simply states that, from the perspective of worldview 1, the consumer is happier with worldview 1 (choosing action 1) than with worldview 2 (choosing action 2).

Figure 1 exhibits the structure of the policy function described in Proposition 1. Each value of α_t , shown on the horizontal axis, maps to either $\alpha_{t+1} = 0$ or

¹³The only flexibility in selecting $z_t(\alpha)$ involves the values of α for which the consumer is indifferent between at least two actions.

¹⁴When $\lambda = 1$, the agent only cares about future actions. Thus, she is indifferent among all worldviews that lead to the same actions. We exclude this case to avoid complicating the characterizations.

¹⁵The consumer is indifferent in the case of $\alpha = \alpha^*$. Consequently there are two stationary MPE policy functions, one with $\phi(\alpha^*) = 0$, the other with $\phi(\alpha^*) = 1$.

$\alpha_{t+1} = 1$, shown on the vertical axis. Two aspects of this equilibrium deserve emphasis.

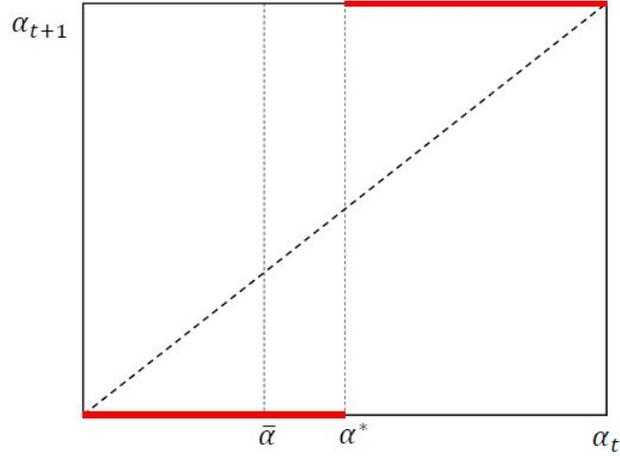


Figure 1. : The structure of sophisticated and naive choice in Case 1.

First, as long as the individual is not too mindset-flexible (specifically, λ is large enough to ensure that $\alpha^* < 1$), both $\alpha = 0$ and $\alpha = 1$ are absorbing states. This property generalizes immediately to settings with more than two pure worldviews: if the optimal action for a given pure worldview yields the greatest happiness under that worldview, then the worldview is an absorbing states for sufficiently large λ .¹⁶ This general property is worth stating as a separate result because it will play an important role in a number of applications:

PROPOSITION 2: *Suppose the number of pure worldviews and the number of actions are both finite. Suppose further that, for each pure worldview i , the optimal action, $z^*(\alpha(i))$, is unique.¹⁷ Consider any worldview i such that $u_i(z^*(\alpha(i))) > u_j(z^*(\alpha(i)))$ for all $j \neq i$. There exists $\lambda_i \in [0, 1)$ such that for $\lambda \geq \lambda_i$, i is an absorbing state in all stationary MPE.*

Under the additional assumption that a unique worldview i maximizes $u_j(z^*(\alpha(j)))$, the condition $u_i(z^*(\alpha(i))) > u_j(z^*(\alpha(i)))$ for all $j \neq i$ is necessarily satisfied.

¹⁶It is worth noting that the existence of multiple steady states remains possible when the action space is a continuum. Suppose $J = \{1, 2\}$, $X = [0, 1]$, u_i is twice continuously differentiable and concave, and $z^*(\alpha)$ is interior for all α . For λ sufficiently close to unity, the consumer considers large changes from any initial worldview detrimental (because the optimal action changes discretely). Moreover, under standard regularity conditions, the envelope theorem implies $\frac{\partial}{\partial \alpha} U(\alpha, z^*(\alpha)) = u_1(x^*(\alpha)) - u_2(x^*(\alpha))$. Consequently, if neither worldview happiness dominates the other, the consumer also considers small movements away from any pure worldview detrimental.

¹⁷Recall that $\alpha(i)$ is the mixed worldview that places all weight on pure worldview i .

Therefore, for λ close to unity, the existence of at least one absorbing state follows as a corollary with considerable generality.

Returning to Proposition 1, the second property of note is that the consumer always adjusts to a pure worldview in a single step. To illustrate the implications of this observation, suppose she starts out in a steady state with worldview 1, $\alpha = 1 > \alpha^*$. If the environment unexpectedly changes in a way that yields $\alpha^* > 1$ (for example, because $u_2(2)$ increases), she will repudiate worldview 1 and suddenly embrace worldview 2. Thus, the dynamics of preference adjustment exhibit a “come-to-god” structure, evoking “moments of revelation.”

Next we turn our attention to Case 2, in which worldview 2 happiness-dominates worldview 1. A consumer who starts out in worldview 1 taking action 1 would like to switch to worldview 2 without also switching to action 2, but knows that the second change would follow from the first. That said, she can move *toward* worldview 2 and achieve some of the benefits of a happier perspective without switching her action, provided α_t remains above $\bar{\alpha}$. Of course, once she shifts her worldview in that direction, she will become less averse to *eventually* switching her action, and consequently will likely take a further step toward worldview 2. Thus, a sophisticated consumer potentially recognizes that she faces a slippery slope. However, as we show next, she never responds by refusing to cross some bright line. Instead, she tiptoes her way up to the threshold $\bar{\alpha}$, and then jumps the rest of the way to worldview 2 in a single step.

Formally, we will show that stationary MPE strategies belong to a class of monotonic step functions, denoted \mathcal{M} , each of which is characterized by some finite sequence $(\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(\bar{\tau})})$ with $\alpha^{(1)} = \bar{\alpha} < \alpha^{(2)} < \dots < \alpha^{(\bar{\tau})} < 1$, and has the following structure:

$$\phi(\alpha) = \begin{cases} 0 & \text{for } \alpha \in [0, \alpha^{(2)}) \\ \alpha^{(1)} & \text{for } \alpha \in [\alpha^{(2)}, \alpha^{(3)}) \\ \alpha^{(2)} & \text{for } \alpha \in [\alpha^{(3)}, \alpha^{(4)}) \\ \dots & \\ \alpha^{(\bar{\tau}-1)} & \text{for } \alpha \in [\alpha^{\bar{\tau}}, 1] \end{cases}$$

Figure 2 illustrates a monotonic step function with five steps ($\bar{\tau} = 5$). The steps define a series of tiers, $[0, \alpha^{(2)})$, $[\alpha^{(2)}, \alpha^{(3)})$, $[\alpha^{(3)}, \alpha^{(4)})$, $[\alpha^{(4)}, \alpha^{(5)})$, $[\alpha^{(5)}, 1]$. Within the first tier, the consumer moves in one step to pure worldview 2. Within the second tier, she moves to $\alpha^{(1)} = \bar{\alpha}$, which lies within the first tier. Within all other tiers, she moves to the lowest value of α within the next lowest tier, and therefore eventually ends up at $\alpha = 0$.

Our next result shows that sophisticated consumers adopt monotonic step functions.

PROPOSITION 3: *Consider Case 2. There exists a stationary MPE for which the policy function governing the evolution of worldviews is a monotonic step*

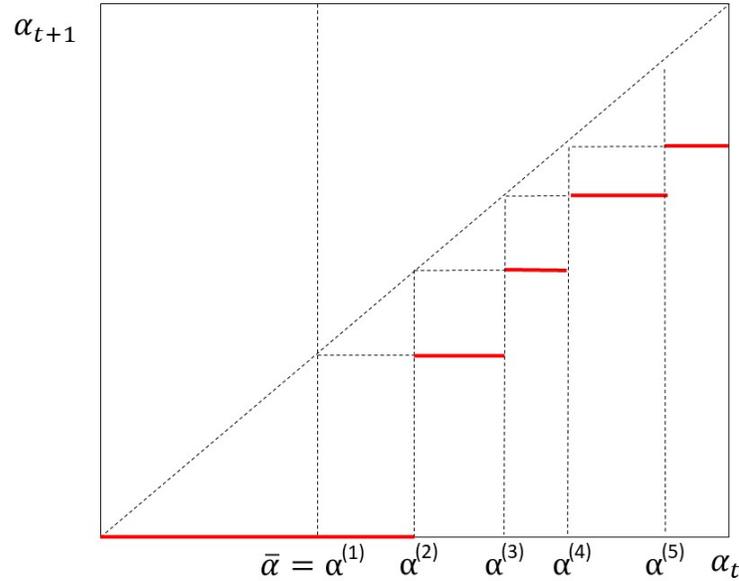


Figure 2. : The structure of sophisticated choice in Case 2.

function belonging to the set \mathcal{M} , and all stationary MPE policy functions coincide with it on a set of full measure. The number of steps is nondecreasing in λ , and for any $T \geq 1$ there exists λ_T such that the MPE policy function has T steps in an open neighborhood of λ_T . Finally, as $\lambda \rightarrow 1$, the number of steps increases without bound, and $\phi(\alpha) \rightarrow \alpha$ for all $\alpha > \bar{\alpha}$.

Several implications of this proposition deserve emphasis. At first, despite the absence of any new information, the consumer slowly changes her mind (that is, reduces α), but not by enough to modify her action. Eventually she reaches a critical threshold at which she changes her behavior and leaps to the alternative worldview. Thus, our theory generates “changes of heart” that result from contemplation of non-trivial duration. It raises the possibility that a change in the economic environment may affect behavior only after a delay, even if the consumer notices the change immediately, because she changes her mind slowly. Significantly, even highly mindset-inflexible individuals (large λ) display this pattern; they simply take longer to adjust. Indeed, one can make the period of “contemplative evolution” arbitrarily long (and the steps arbitrarily small) by taking λ sufficiently close to unity. Significantly, the consumer never responds to the existence of this slippery slope by drawing a “bright line.”

Cases 1 and 2 both have the property that the consumer ultimately ends up with a pure worldview, and arrives at it in a single dramatic step (e.g., from $\alpha \geq \bar{\alpha}$ to $\alpha = 0$). It is straightforward to show that consumers also converge to

purist perspectives in settings with an arbitrary number of worldviews, provided the consumer is sufficiently mindset-flexible (see Proposition 5 in Bernheim et al., 2019).¹⁸

We presented a parallel characterization of naive behavior in an earlier version of this paper (Bernheim et al., 2019). In Case 1, naive and sophisticated consumers make identical choices. In Case 2, the naive consumer follows a monotonic step function with no more than two steps, and always reaches the steady state in (weakly) fewer rounds than a sophisticated consumer. Intuitively, the naive consumer believes she can switch from high values of α to $\bar{\alpha}$, the lowest value consistent with choosing action 1, without triggering subsequent changes in her worldview. However, she always switches from $\bar{\alpha}$ to pure worldview 2 in the ensuing period.

Naivete gives rise to particularly interesting patterns when the consumer chooses from a rich set of “tailored” worldviews, by which we mean the following. Let X^I be the set of implementable actions – that is, actions for which there is some worldview (mixed or pure) that gives rise to that action. We will make the following assumptions:

- (i) For each $x \in X^I$, there exists a pure worldview, $\alpha^I(x)$, that implements x .
- (ii) For each $x \in X^I$, $u_{\alpha^I(x)}(x) > u_j(x)$ for all $j \neq \alpha^I(x)$.

Notice that, with two worldviews and two actions, Case 1 satisfies these assumptions. More generally, when they hold, a naive consumer is a “serial true believer,” in the sense that she moves rapidly from one pure worldview to another in a series of “come-to-god” moments, until she finds one that sticks.

PROPOSITION 4: *Under assumptions (i) and (ii), a naive agent is a “serial true believer.” She picks a pure worldview in every period, but changes it every period for a finite number of rounds, then stops changing.*

B. Adding worldviews

The introduction of a new worldview, either as a “foreign influence” or an internal “innovation,” can profoundly affect how members of society behave. In this section, we show how the existence of new worldviews can harden traditionalism in some settings, while providing a “gateway” for social change in others.

REACTIONARY RESPONSES TO NEW WORLDVIEWS

First we consider the possibility that the introduction of a “liberal” worldview may elicit a reactionary response among “conservatives,” undermining receptive-

¹⁸The argument is straightforward for the boundary case in which $\lambda = 0$. With perfect mindset flexibility, the consumer chooses future α and x to maximize $\sum_{j \in J} \alpha^j u_j(x)$. For any fixed x , this function is linear in α . Optimizing the choice of x conditional on each α yields the upper envelope of these linear functions, which is convex. Thus, the optimal value of α always involves a boundary solution, placing all weight on a single worldview. We conjecture that consumers also converge to purist perspectives for low levels of mindset flexibility, but we have not proven this conjecture.

ness to change.¹⁹ We illustrate this possibility in a setting where people initially choose between two actions and two worldviews. Action 1 is “traditional” and action 2 is “modern.” Worldview 1 is “conservative” and worldview 2 is “moderate.” We assume that the traditional action is optimal under both worldviews. Accordingly (as shown formally in the first part of the next proposition), people will select the worldview that allows them to derive the greatest happiness from that action – in other words, they will maximize $u_i(1)$ over the choice of i .

Imagine that, historically, the conservative outlook has been dominant and stable ($u_1(1) > u_2(1)$), but that contact with other “modern” cultures has changed the environment in a way that now favors the moderate outlook ($u_2(1) > u_1(1)$). Were this the only development, the population would migrate to the moderate worldview in a single period. Suppose, however, that contact with modern cultures also introduces a “liberal” worldview, $i = 3$, that favors the modern action ($u_3(2) > u_3(1)$). We will assume that this worldview delivers the greatest happiness ($u_3(2) > u_2(1)$), and indeed that people would switch immediately from the moderate worldview to the liberal one if these were the only two options: $u_2(1) < \lambda u_2(2) + (1 - \lambda)u_3(2)$.²⁰ To simplify some of the arguments, we will also assume that worldview 3 does not happiness-dominate worldview 2 ($u_3(1) < u_2(1)$). Finally, we assume that conservatives are highly opposed to the modern action ($u_1(2)$ is low).

The following proposition consists of two parts. The first part verifies the assertion stated above: under the indicated conditions, and in the absence of the liberal worldview, people migrate immediately from the conservative worldview to the moderate worldview. The second part demonstrates that the introduction of the liberal worldview delays that migration. Greater aversiveness of conservatives to modern actions (i.e., a lower value of $u_1(2)$) lengthens the delay, and can make it arbitrarily long. In effect, the presence of the liberal worldview creates an environment that resembles Case 2 of the simple model: knowing that a moderate outlook will lead to liberalism and hence to modern actions, conservatives trade off the lure of greater happiness against their repulsion from modernism, and therefore tiptoe toward the moderate perspective instead of jumping to it quickly.

PROPOSITION 5: *(i) When worldview 3 is not available: The consumer places zero weight on worldview 1 after the first period in all stationary MPE.*

(ii) When worldview 3 is available: For all $T > 0$, there exists K_T with $\lim_{T \rightarrow \infty} K_T = -\infty$ such that if $u_1(2) < K_T$, a consumer who starts with pure worldview 1 places strictly positive weight on that worldview for at least T periods in all stationary MPE.

¹⁹For an example of a reactionary response, see Fouka (2020).

²⁰This inequality implies that condition (2) is violated.

NEW WORLDVIEWS AS GATEWAYS

We saw in the previous section that the introduction of a new *extreme* worldview can engender resistance to change if its presence causes people to fear that they might drift too far from their traditional values. Here we consider the opposite possibility, that the introduction of an *intermediate* worldview may provide an attractive stepping stone, thereby undermining traditionalism.

Suppose in particular that there are initially two worldviews, $J = \{1, 3\}$, respectively “conservative” and “liberal,” as well as two actions, $X = \{1, 2\}$, respectively “traditional” and “modern,” where the traditional (modern) action is optimal under the conservative (liberal) worldview. Imagine also that Case 1 prevails with $\alpha^* < 1$, so that both pure worldviews are absorbing states, but the “liberal” worldview is happier. Under these assumptions, local movements toward liberalism (worldview 3) starting from conservatism (worldview 1) make the consumer worse off, and her mindset flexibility is insufficiently high to justify a leap from one end of the spectrum to the other.

Now suppose we introduce a “moderate” worldview ($i = 2$) that happiness-dominates worldview 1, but not worldview 3, while reversing the preference ordering over the actions so as to favor the modern alternative. In particular, $u_2(2) > u_2(1)$, $u_2(1) > u_1(1)$, and $u_2(2) < u_3(2)$. Ignoring the liberal worldview for the moment, Case 2 would prevail between the conservative and moderate worldviews. Consequently, we would expect consumers to end up at the moderate worldview in finite time. Note, however, that upon reaching worldview 2, they would in fact jump immediately to worldview 3 (inasmuch as both perspectives favor action 2). According to worldview 1, it is better to end up at worldview 3 than worldview 2. Therefore, consideration of the eventual jump to worldview 3 only makes the transition between worldviews 1 and 2 more attractive.

The preceding intuition is incomplete because it does not consider all mixed worldviews (e.g., those placing positive weight on both conservative and liberal worldviews). Even so, our conclusion is valid:

PROPOSITION 6: *Consider a setting with two actions, 1 and 2, and initially two worldviews, 1 and 3. Assume that Case 1 prevails (with action 1 optimal under worldview 1, action 2 optimal under worldview 3, $u_1(1) < u_3(2)$, and $\alpha^* < 1$). There exists a class of pure worldviews the introduction of which gives rise to MPE in which all consumers transit to worldview 3 in finite time, irrespective of their initial worldviews.*

Figure 3 illustrates the adjustment path starting from the conservative worldview ($i = 1$), under the assumption that all three worldviews are available. The triangle represents the unit simplex, and each point in the simplex corresponds to a set of weights for the three pure worldviews. At first, the consumer stays on the edge of the simplex connecting conservative ($i = 1$) and moderate ($i = 2$) worldviews. She “tiptoes” toward the moderate worldview, as in Case 2 of the basic model.

However, before reaching that pure worldview, she jumps to the liberal worldview ($i = 3$).

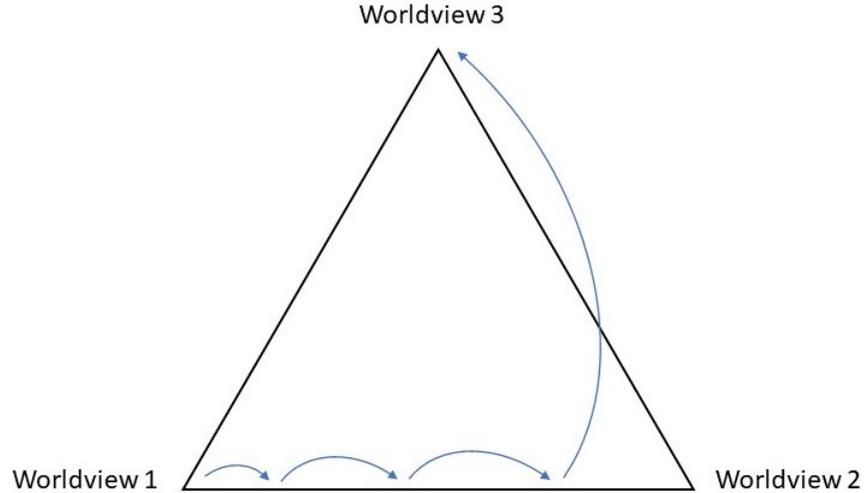


Figure 3. : The worldview simplex displaying the sequence of worldviews an agent takes who begins with all weight on Worldview 1 (lower left vertex). Here the agent moves to placing all her weight on Worldview 3 in four steps.

III. Applications

A. *The origins of closed-mindedness*

So far we have treated the mindset-flexibility parameter λ as a fixed personality trait. Yet many religions encourage mindset inflexibility by asserting uncompromising accounts of “absolute truth” that reject cultural relativism, while some philosophies applaud open-minded adaptability. For certain purposes, it is therefore appropriate to treat λ as part of the worldview. Here we ask whether our theory predicts a tendency for people to gravitate toward open- or close-minded worldviews.

We extend the two-action two-worldview setting described in section II.A to include many versions of each “basic” worldview, one for each degree of mindset flexibility λ within some closed set $\Lambda \subseteq [0, 1]$. Let $\bar{\lambda}$ denote the largest value of $\lambda \in \Lambda$. We assume $\bar{\lambda} < 1$, perhaps because our neurobiological wiring precludes complete inflexibility.²¹

²¹We also rule out the case of $\bar{\lambda} = 1$ for a practical reason: it raises some technical issues that complicate the analysis (see below).

To understand the consumer's incentives with respect to mindset flexibility, recall that she suffers from time inconsistency. Assuming she is sophisticated, she anticipates that a change in her current worldview will in some circumstances lead to further changes, contrary to her contemporaneous wishes. Choosing a worldview with a relatively inflexible mindset is essentially a (partial) commitment to future worldviews, which (partially) resolves the time-inconsistency problem. To illustrate, suppose in particular that Case 2 prevails. A consumer who starts out with worldview 1 would like to shift to the mixed paradigm $\bar{\alpha}$ and stay there forever, choosing action 1. With $\lambda < 1$ she takes smaller steps because she knows the slope is slippery. However, intuition suggests that, with $\lambda = 1$, she can safely move to $\bar{\alpha}$ without the risk of drifting into the territory where she would pick action 2. While this intuition is not quite complete,²² it points us in the right direction. Consequently, as our next proposition demonstrates, the theory yields an endogenous explanation for the prevalence of inflexible worldviews.

PROPOSITION 7: *Consider a setting with two potential actions and two basic worldviews, either of which can be paired with any $\lambda \in \Lambda$. Suppose action i is the unique optimum for basic worldview i , and that basic worldview 2 happiness-dominates basic worldview 1 (Case 2 of the simple model). Let (α, λ) denote the initial worldview. Under the stated conditions, a stationary MPE exists. Furthermore, for all stationary MPE, if the consumer selects $\alpha_{t+1} > \bar{\alpha}$ in period t , she also selects maximal inflexibility ($\lambda_{t+1} = \bar{\lambda}$). Finally, for all $\alpha_t > \bar{\alpha}$, there exists $\underline{\lambda} < 1$ such that if $\bar{\lambda} > \underline{\lambda}$, then $\lambda_t \in (\underline{\lambda}, \bar{\lambda}]$ implies $\alpha_{t+1} > \bar{\alpha}$.*

It is worth emphasizing that Proposition 7 does not establish a general preference for inflexibility. Rather, it states that the consumer will be maximally inflexible if her choice of worldview for the subsequent period involves a value of α exceeding $\bar{\alpha}$. (The final portion of the proposition also establishes that there are conditions under which the consumer makes such choices.) She has no reason to seek a commitment in Case 1, or in Case 2 if she begins with a low α and would in any event move immediately to pure worldview 2. In such cases the optimal choice of λ becomes indeterminate.

The preceding analysis presupposes the absence of uncertainty. More generally, uncertainty provides the consumer with a reason to preserve a degree of flexibility. As in other settings with dynamic inconsistency, the optimal strategy then reflects a balance between these competing objectives.

²²When $\lambda = 1$, the consumer only cares about her future worldviews to the extent they affect her future actions. Consequently, she is indifferent between choosing $(\alpha = \bar{\alpha}, \lambda = 1)$ and $(\alpha > \bar{\alpha}, \lambda \leq 1)$, provided the two continuation paths lead to the same trajectory of actions. Focusing on cases with $\bar{\lambda} < 1$ avoids this complication.

B. Social influences, conformity, and assimilation

Accumulating evidence shows that people often adjust their behavior to conform with the choices of others, either partially or fully.²³ Economic theories variously attribute this tendency to factors such as community sanctions (Akerlof, 1980), social signaling (Bernheim, 1994), and the information conveyed by others' choices (Banerjee, 1992, Bikhchandani, Hirshleifer and Welch, 1992). In all of these theories, preferences are fixed. Social equilibrium either standardizes inferences or promotes a shared understanding of social conventions and norms.

Many psychologists, sociologists, and anthropologists envision a separate route to conformity involving the convergence of preferences. One leading school of thought traces its roots to Festinger (1954), who emphasized the role of “tension systems.” According to this view, people inherently dislike being in a state of disagreement with other members of their social group. One common response to “social dissonance” is to conform – i.e., mimic observable manifestations of the predominant attitude. That response creates unpleasant dissonance between the individual’s actions and their true attitudes. According to Festinger, the individual resolves this internal dissonance by gradually bringing their outlook in line with their observable behavior. Thus a predominant attitude is eventually internalized and becomes a general motivation for behavior.

Our theory of chosen preferences yields a formal account of Festinger’s hypothesized dynamic. For the purposes of interpreting our model and results, it helps to have in mind a concrete application of cultural assimilation. Imagine that Antonio moves to the United States where he finds himself immersed in a community obsessed with American football. His friends and co-workers gather on weekends to watch televised NFL and college games, and they talk incessantly about the sport. To avoid being treated as an outsider, Antonio joins their gatherings and forces himself to learn enough about football to carry on conversations without embarrassment. Though he initially derives no real pleasure from the games themselves, he is determined to make the best of a situation he cannot reasonably avoid, and deliberately cultivates an appreciation of football. Over time, as his intentionally acquired taste for football takes root, he assimilates, and starts watching games in private for his own personal enjoyment.²⁴

Formally, as in Section II.A, we focus on an infinite horizon model ($T = \infty$) in which society offers two possible pure worldviews, $J = \{1, 2\}$ (e.g., one that values

²³The classic reference is Asch (1956). As examples of more recent work, Chen et al. (2010) document conformism in voluntary contributions to an online community, and Fouka (2019) examine assimilation efforts by German immigrants to the U.S.

²⁴Other examples abound. A liberal who moves to a conservative area might find it expedient to express conservative views publicly while nevertheless continuing to vote for liberals. Over time, she might make the best of her circumstances by deliberately nudging herself toward some conservative perspectives in order to feel more comfortable socially, thereby eventually turning herself into a conservative voter. Similarly, a lover of classical music who moves to Nashville might simulate an appreciation of country music in order to fit in, despite listening to Mozart in private. Over time, he might cultivate a sincere taste for country music in order to make the best of his circumstances, thereby eventually turning himself into a bona fide Tim McGraw fan.

football and one that does not). We modify our earlier model by assuming that the consumer makes a collection of simultaneous choices in each period. Payoffs are additive across these decisions. As before, each choice involves two options (e.g., whether to watch a football game), and the instantaneous intrinsic payoff to action i under worldview j is $u_j(i)$.

Choices fall into two categories, public (e.g., whether to attend a gathering to watch a game with friends) and private (e.g., whether to watch a game in private in the absence of a gathering). We use θ to denote the ratio of the number of public decisions to the number of private decision. Publicly choosing action i generates a utility bonus or penalty of $s_i(\mu)$, where μ denotes the fraction of social contacts who select action 1. Thus, the total payoff from action i under worldview j is $u_j(i) + s_i(\mu)$. To capture the notion that observable non-conformity creates unpleasant dissonance,²⁵ we assume that an increase in μ raises s_1 and reduces s_2 . Let $S(\mu) = s_1(\mu) - s_2(\mu)$. For the sake of simplicity, we also assume symmetry ($s_1(\mu) = s_2(1 - \mu)$), so that $S(\mu)$ is strictly positive for $\mu > 0.5$ and strictly negative for $\mu < 0.5$.

We assume that, in the absence of social considerations, Case 1 of Section II.A would prevail: $u_2(2) > u_1(1)$ (worldview 2 is happier than worldview 1) and $u_1(1) > u_2(1)$ (worldview 2 does not happiness dominate worldview 1), with $\alpha^* < 1$. Accordingly, without the pressure to conform, actions 1 and 2 would both be absorbing states, one of which the consumer would reach in a single step, regardless of her starting point.

We also assume that some prevailing equilibrium establishes the value of μ , which we treat as fixed. In effect, we study the behavior of an individual who joins a large and stable population. We take action 2 to be the prevalent choice, which implies $S(\mu) < 0$.

A consumer who arrives in period t with mixed worldview α_t will behave as follows. For each private choice, she will select action 1 if $\alpha_t > \bar{\alpha}$, action 2 if $\alpha_t < \bar{\alpha}$, and either action if $\alpha_t = \bar{\alpha}$. For each public choice, she will select action 1 if $\alpha_t > \bar{\alpha}^P$, action 2 if $\alpha_t < \bar{\alpha}^P$, and either action if $\alpha_t = \bar{\alpha}^P$, where

$$\bar{\alpha}^P = \frac{u_2(2) - u_2(1) - S(\mu)}{[u_2(2) - u_2(1)] + [u_1(1) - u_1(2)]}$$

Finally, we will assume that social pressure is so strong that the consumer chooses action 2 for her public decisions regardless of her mixed worldview: $u_1(1) - u_1(2) + S(\mu) < 0$, which implies $\bar{\alpha}^P > 1$.

The following proposition shows how the behavior and attitudes of a consumer who starts out with worldview 1 evolve over time. As long as the consumer is

²⁵It is important to emphasize that our conclusions do not depend on the particular mechanism that generates *public* conformity. We have chosen a simple representation for convenience, but alternatively could have employed other mechanisms, such as social sanctions, as in Akerlof (1980), or social signaling, as in Bernheim (1994). In either case, appending our chosen preference mechanism would generate Festinger's assimilation dynamics.

not too mindset-flexible and neither public nor private decisions are too prevalent, he immediately conforms publicly but not privately (e.g., joins gatherings to watch football games, but does not watch alone), and then shifts gradually from worldview 1 to worldview 2 (e.g., develops a taste for football) to eliminate the tension between his true attitudes and his public actions. This gradual adjustment is particularly notable in light of the fact that the parameter values fall into Case 1, which yields immediate adjustment for the settings considered in section II. Eventually he assimilates, in the sense that he completely internalizes the prevailing attitude by adopting pure worldview 2, at which point he conforms privately as well as publicly (e.g., watches football games by himself). Thus, the evolution of his attitudes and behavior broadly matches the pattern that Festinger described.²⁶ The proof simply involves rewriting the model so that it falls within the class examined in Section II.A, demonstrating that Case 2 in fact prevails in the transformed model for intermediate values of θ , and applying Proposition 3.

PROPOSITION 8: *Assume $\bar{\alpha}^P > 1$. For λ sufficiently close to unity, there exist $\theta_1 > 0$ and $\theta_2 > \theta_1$ such that the following characterization holds for $\theta \in (\theta_1, \theta_2)$. There exists a stationary MPE involving a monotonic step function with at least two steps, and all MPE policy functions coincide with it on a set of full measure. In this equilibrium, the consumer eventually transitions to worldview 2 regardless of her starting point. A consumer who starts out with worldview 1 selects action 1 for all her private decisions for a finite number of periods until she fully adopts pure worldview 2 (the final step), at which point she switches to action 2. She selects action 2 for all of her public decisions in every period. The number of steps in the MPE policy function is nondecreasing in λ , and for any $T \geq 1$ there exists λ_T such that the MPE policy function has T steps in an open neighborhood of λ_T . As $\lambda \rightarrow 1$, the number of steps increases without bound.*

C. The “sour grapes” effect and related phenomena

According to Elster (1983), “the preferences underlying a choice may be shaped by... constraints,” a phenomenon known as *adaptive preferences*. As motivation, he recounts the parable of the hungry fox who, upon discovering himself unable to reach some otherwise appealing grapes, declares them sour and hence undesirable.²⁷ Our theory provides a formal account of this “sour grapes” effect and

²⁶If the fraction of public decisions is small, both worldviews are absorbing states. The consumer therefore conforms publicly, but never changes her worldview nor conforms privately. In Festinger’s language, fear of social dissonance deters public individuality, but the resulting internal dissonance is too weak to precipitate a true change in attitudes. In contrast, if the consumer is sufficiently mindset-flexible or the fraction of public decisions is large, she not only conforms publicly, but also converts to the prevailing worldview instantaneously to avoid severe internal dissonance. The proof of the proposition 8 encompasses these cases.

²⁷Elster references the parable to raise the possibility that the fox might resolve not to desire the grapes, or to form a positive desire not to eat them, rather than the possibility that the fox might form a false belief that the grapes are sour.

highlights some intriguing dynamic implications. Intuitively, a temporary restriction on a consumer's opportunity set leads her to make the best of her experience through the choice of a consonant worldview (the sour grapes effect). Moreover, due to the "stickiness" of worldviews, the resulting changes in preferences and behavior outlast the restriction, and can even be permanent in settings where multiple worldviews are absorbing states (recall Proposition 2).

For the purpose of illustration, we will simplify by focusing on Case 1 of the two-action two-worldview model, and assume $\alpha^* < 1$. Our first key observation, summarized as part (a) of the next proposition, describes the effect of temporarily removing one of the two actions from the opportunity set. For the purpose of this result, we focus on *quasi-stationary* MPE, which satisfy stationarity after the first K periods (i.e., once the structure of the problem becomes stationary). Intuitively, if the consumer realizes she will have to choose action j in the first K periods, and if K is large, she will make the best of this constraint by adopting worldview j , which effectively rationalizes her outcome. But since worldview j is an absorbing state, she will stay with it once the constraint is removed. Thus, a temporary intervention can create a permanent change in behavior. The same idea applies to temporary subsidies or taxes that increase or decrease the attractiveness of an action; see part (b) of the proposition.

PROPOSITION 9: *Consider Case 1 of the basic model, with $\lambda < 1$ and $\alpha^* < 1$.*

(a) *Suppose action i is not available in periods $1, \dots, K$. If K is sufficiently large, then all quasi-stationary MPE have the property that the consumer chooses pure worldview $j \neq i$, along with action j , in every period.*

(b) *Suppose a subsidy for action j increases $u_1(j)$ and $u_2(j)$, or a tax decreases $u_1(i)$ and $u_2(i)$, by an amount σ for K periods. If $u_i(j) + \sigma > u_i(i)$ (for $i \neq j$) and K is sufficiently large, then all quasi-stationary MPE have the property that the consumer chooses pure worldview j , along with action j , in every period.*

This result has many potential applications. For example, a common assumption in the literature on intergenerational preference transmission is that parents can select preferences for their children (e.g., Bisin and Verdier, 2001), but the transmission mechanism is typically unexplained. Nor is it clear why children stick with these preferences once they are on their own. Our theory provides a more complete account of the durable transmission of cultural preferences. To illustrate, we reinterpret the two-action two-worldview model as follows. Suppose action 1 involves going to church, while taking action 2 involves avoiding church. Action 1 is optimal under a religious outlook (worldview 1), while action 2 is optimal under a non-religious outlook (worldview 2). Applying Proposition 9, part (a), we see that a child who is forced to attend church every Sunday will tend to make the best of her experience by adopting a religious worldview. Assuming this worldview is not happiness-dominated, it is an absorbing state. Consequently, the child will remain religious even after she is on her own.²⁸

²⁸The simple model obviously does not capture all the complexities that can come into play. For

In an earlier version of this paper (Bernheim et al., 2019), we discussed other applications that potentially involve sour grapes effects along the lines of those described in Proposition 9, including the following. (1) *Authoritarian oppression without coercion*. People who suffer from oppression often express agreement with the values that justify their oppressors (Nussbaum, 2001, Sen, 1985). Our theory provides an explanation:²⁹ through temporary coercion, the authority can induce a psychological reorientation that ensures lasting acquiescence.^{30,31} (2) *The cultivation of habits*. A commitment to perform an activity, such as washing hands, for a period of time can result in it becoming habitual. According to our theory, the commitment induces people to adopt durable worldviews that rationalize the activity, for example by embracing the importance of hygiene. (3) *Densensitization*. Systematic exposure to an aversive condition can reduce the degree of aversion (Wolpe, 1961). In our theory, the exposure encourages people to adopt durable worldviews that minimize the importance of its negative consequences. (4) *Thriftiness of the Depression cohort*. The cohort that lived through the Great Depression is often described as more thrifty and patient than its successors (Kapteyn, Alessie and Lusardi, 2005). Our theory leads to the following explanation: during a severe and prolonged economic downturn, the anticipation of eventual improvement implies that people can make the most of their plight by cultivating more patient worldviews, the persistence of which accounts for their greater patience.³² (5) *Cultural impediments to development*. Noting that the introduction of institutions and opportunities does not automatically improve economic outcomes, some economists attribute the persistently poor quality of life in developing countries to cultural factors such as “aspira-

example, it cannot account for children who rebel against their parents’ cultural preferences. A complete theory of cultural transmission obviously requires a richer model.

²⁹For an alternative interpretation of oppression without coercion involving endogenous beliefs, see Benabou (2013).

³⁰Chen and Yang (2019) document a practical example involving censorship in China. Although enforcement of internet censorship is lax and easily circumvented, the consumption of foreign media is low. However, in an experiment, people are willing to use an internet service that provides access to censored material when they receive a small incentive payment, and consumption of foreign media remains persistently high after the incentive is removed. Seen through the lens of Proposition 9, part (a), strict censorship in China drove the population to the isolationist worldview that discounts the value of foreign media. Censorship survived with lax enforcement because that worldview was an absorbing state. Chen and Yang’s experiment temporarily induced subjects to consume foreign media, thereby shifting them toward globalist worldviews, as described in Proposition 9, part (b). Chen and Yang offer a quote from Neil Postman that evokes our interpretation: “What Orwell feared were those who would ban books. What Huxley feared was that there would be no reason to ban books, for there would be no one who would want to read them.”

³¹Proposition 9 points to an interesting incentive problem: if the coercion is of sufficient duration to induce people to change their worldviews, this response will occur before the coercion ends (immediately in the simple model). But if coercion is costly, the authority will dispense with it as soon as people switch worldviews. Anticipating that response, each individual will expect coercion to end quickly, and consequently will not switch to begin with. In Bernheim et al. (2019), we show that there is an equilibrium involving periodic and unpredictable crackdowns targeted at a population segment that never fully acquiesces.

³²As we mentioned in section I, the simple formulation of preferences adopted in this paper does not permit us to endogenize patience (the discount factor). For this purpose, the recursive formulation developed in our earlier working paper (Bernheim et al., 2019) is required.

tion failures” (Genicot and Ray, 2017). According to our theory, people adopt worldviews that de-emphasize the value of opportunities that are not available, and those attitudes can prove durable even with institutional changes that create those opportunities.³³ (6) *Overcoming prejudice*. Involuntary intergroup contact can reduce intergroup prejudice (Pettigrew and Tropp, 2006, Rao, 2019). According to our theory, this contact fosters the durable adoption of worldviews that celebrate diversity. Likewise, positive and humorous (but non-stereotypical) depictions of minorities in entertainment media can reduce prejudice without conveying real information (Ford, 1997, Schiappa, Gregg and Hewes, 2006). Our theory suggests that people respond to such entertainment by adopting durable worldviews (inclusive attitudes) that maximize the enjoyment they derive from it. (7) *Coping with religious doubts*. Blaise Pascal argued that a temporary commitment to participate in rituals can lead to lasting religious faith (Pascal, 1958; see also Dennett, 2006). In our theory, the expectation of sufficiently sustained participation leads people to make the most of their experience by cultivating faith, a durable worldview. (8) *Marketing strategies*. Temporary inducements to begin using products may have strong effects on demand. In our theory, these inducements can encourage the adoption of worldviews that attribute greater value to the benefits the products provide.

In a related application, we also showed in Bernheim et al. (2019) that our theory provides a potential explanation for the endowment effect. In the canonical experiment (Knetsch, 1989; Kahneman, Knetsch and Thaler, 1990), half of the subjects, chosen at random, receive coffee mugs, while the others do not. Contrary to the predictions of standard theory, considerably fewer than half the mugs change hands once a market for mugs unexpectedly opens. Furthermore, the willingness to accept (WTA) for mugs among those who receive them is roughly twice the willingness to pay (WTP) among those who do not. These effects are substantially attenuated when trading opportunities are anticipated (Marzilli-Ericson and Fuster, 2011).

Our theory explains these patterns as follows. A participant who receives a mug, and who does not anticipate subsequent trading opportunities, has an incentive to convince herself that mugs are desirable.³⁴ But why does that effect persist once the trading opportunities materialize, at which point she learns that her opportunities are actually independent of her initial mug assignment? According to our theory, the answer is that, once a worldview is adopted, it becomes “sticky” (much as Proposition 9 suggests). In particular, having already talked herself into the view that mugs are desirable, she does not completely talk herself out of it. In contrast, when she recognizes from the outset that her opportunities will be independent of her initial mug assignment, she has no incentive to convince

³³However, our theory also suggests that it may be possible to overcome cultural inertia through temporary measures that strongly incentivize new practices; see, e.g., Dupas (2014) for an example.

³⁴In the psychology literature, some refer to this account of the endowment effect as “motivated taste change” or the “mere-ownership effect.” See Morwedge and Giblin (2015) and Marzilli-Ericson and Fuster (2014) for reviews of evidence bearing on the validity of these theories.

herself that mugs are especially desirable in the first place. We also showed that our theory is empirically distinguishable from the standard explanation for endowment effects, which posits reference-dependent preferences along with loss aversion (Kahneman, Knetsch and Thaler, 1991). Specifically, under our theory, receiving one mug inflates the WTP for additional mugs commensurately with the WTA for the first mug. The same implication does not in general follow under reference-dependence and loss aversion, because additional mugs are not part of the endowment.³⁵

Proposition 9 has the somewhat unrealistic implication that people change their worldviews immediately in response to temporary conditions involving restrictions or opportunities. Next we show that this feature is a consequence of the rather stark assumption that the duration of the condition is fixed and known. If instead the duration is uncertain, people change their worldviews gradually rather than instantaneously. Furthermore, if the condition happens to end before the consumer's mixed worldview reaches a critical threshold, she returns to her original outlook. However, if the condition persists long enough, she fully "converts" to the alternative worldview, and retains that outlook indefinitely. Thus, for example, if an authoritarian regime attempts to impose a change in behavior through coercion, and if there is also a chance the regime will collapse, people will display a degree of resistance, acquiescing gradually rather than instantaneously. If the regime ends sufficiently quickly, people will return to their original views, but a long-lived oppressive regime can have durable effects on behavior even after it collapses.

To keep the analysis reasonably simple, we will focus on the two-action two-worldview model and study a temporary restriction that excludes action 2 from the choice set. The restriction persists from one period to the next with probability $1 - p$, and consumers maximize expected utility. Decisions and events unfold as follows. If the restriction is still in force at the outset of period t , the consumer (trivially) chooses action 1. Next, she learns whether the restriction will continue into period $t + 1$. Conditional on that information, she chooses her mixed worldview for period $t + 1$. In the event she learns the restriction will remain in force, she believes it will end in period $t + k$ with probability $(1 - p)^{k-2}p$.

We assume that $u_2(2) > u_1(1) > \max\{u_2(1), u_1(2)\}$ (Case 1 of the basic model): worldview 2 potentially yields the greatest happiness, but it does not happiness dominate worldview 1. Furthermore, we impose condition (2), so that $\alpha^* < 1$. Assuming that worldview 1 (rather than worldview 2) potentially yields the greatest happiness changes nothing of substance, with one minor exception noted below.

The following proposition shows that this model gives rise to an MPE wherein the consumer *gradually acquiesces* to worldview 1 by adjusting her mixed world-

³⁵Notably, Morewedge et al. (2009) conducted a two-mug experiment and obtained results consistent with our theory. However, the evidence on this point is mixed; see Marzilli-Ericson and Fuster (2011) who find evidence against what they call "motivated taste changes."

view in steps. The MPE policy mapping governing the evolution of the consumer's worldview belongs to a class of monotonic step functions, denoted \mathcal{M}' , similar to those examined in Proposition 3, each of which is characterized by some finite sequence $(\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(\bar{\kappa})})$ with $\alpha^{(1)} = \alpha^* > \alpha^{(2)} > \dots > \alpha^{(\bar{\kappa})} > 0$, and has the following structure:

$$\phi(\alpha, 1) = \begin{cases} \alpha^{(\bar{\kappa}-1)} & \text{if } \alpha \in [0, \alpha^{(\bar{\kappa})}] \\ \alpha^{(\bar{\kappa}-2)} & \text{if } \alpha \in (\alpha^{(\bar{\kappa})}, \alpha^{(\bar{\kappa}-1)}] \\ \vdots & \vdots \\ \alpha^{(1)} & \text{if } \alpha \in (\alpha^{(3)}, \alpha^{(2)}] \\ 1 & \text{if } \alpha \in (\alpha^{(2)}, 1] \end{cases}$$

The use of this policy function implies that, during the processes of adjustment, the consumer resists adopting worldview 1 in the hope that the restriction will end, but nevertheless progressively comes to see things more and more from that perspective. If the restriction ends soon enough, she returns immediately to her original worldview. If it persists, she eventually reaches a point at which she gives up and fully embraces worldview 1. However, in contrast to Proposition 3, she switches from action 2 to action 1 *before* making the final leap.³⁶ Once she takes that leap, she continues to select action 1 and to embrace worldview 1 even after the restriction ends. In applications involving authoritarian regimes, oppression without coercion thereby emerges as the ultimate consequence of gradual acquiescence.

PROPOSITION 10: *Assume $u_2(2) > u_1(1) > \max\{u_2(1), u_1(2)\}$. There exists a stationary MPE for which the policy function governing the evolution of worldviews while the restriction on the availability of action 2 persists is a monotonic step function belonging to the set \mathcal{M}' . Once the restriction ends, the continuation equilibrium described in Proposition 1 prevails.*

D. Reconciling large-stakes and small-stakes risk aversion

One of the best-known puzzles in behavioral economics concerns a tension between small-stakes and large-stakes risk aversion. Hansson (1988) and Rabin (2000) provide calibration results showing that it is impossible to reconcile even modest risk aversion in the context of small gambles with sensible behavior for gambles involving large stakes. They show that small-stakes behavior implies a lower bound on risk aversion for large-stakes gambles, and the implied bound is unreasonably high. This insight has been widely construed as implying that the tension between low- and high-stakes risk aversion establishes the empirical rele-

³⁶The reason for this difference is that, in the current instance, the last step before full acquiescence is to α^* rather than to $\bar{\alpha}$. If we had assumed that worldview 1 is happier than worldview 2, then we would have $\alpha^* < \bar{\alpha}$, and the consumer would not switch actions until reaching pure worldview 2, analogously to Proposition 3.

vance of reference dependence and loss aversion. Here we explore another possible explanation involving chosen preferences.

Before presenting our formal analysis, we provide a brief intuitive explanation of our results. We consider a model in which a consumer makes choices with uncertain material payoffs. Potential worldviews vary according to the importance they attach to materialism (e.g., “money can’t buy happiness” versus “anyone who says that money can’t buy happiness doesn’t know where to shop”). We capture this variation through a parameter, θ . An increase in θ makes the consumer more materialistic, in the sense that the utility associated with a high payoff rises while the utility associated with a low payoff declines. Under this assumption, no worldview generally happiness-dominates another. We write the utility associated with worldview θ and payout x as $U(x, \theta)$. We assume the consumer can modify her worldview at some fixed (psychological) cost after learning her payoff. In our simple model, any revision she decides to undertake maximizes her ex post happiness. Thus, when she receives a high payoff, she adjusts her θ upward, whereas if she receives a low payoff, she adjusts her θ downward.

It is easiest to build intuition for this setting by focusing on the special case in which the consumer is perfectly mindset-flexible ($\lambda = 0$). Suppose for the moment that worldview modifications involve no other costs. Then the consumer is necessarily *less* risk averse than she would be if her worldview (θ) were fixed. To understand why, consider Figure 4, Panel A. Since the consumer is free to set θ ex post, she chooses the value $\bar{\theta}(x)$ that delivers the greatest utility given her payoff x . As a result, the ex post utility she actually experiences corresponds to the upper envelope of all her possible fixed-worldview utility functions. Because the upper envelope is necessarily less concave than the functions from which it is derived, the consumer’s ability to change her worldview makes her less risk averse. Indeed, as shown in Panel B, the upper envelope may even be convex, which means her flexibility makes her risk-loving. As we show below, all of these conclusions carry over to situations with imperfect mindset flexibility ($\lambda \in (0, 1)$), though the quantitative impact on risk aversion is muted.

Now consider the effect of introducing a positive fixed cost of worldview modifications. If the optimal worldview for the realized payoff turns out to be close to the consumer’s initial worldview, she won’t be inclined to change it. Assuming her initial worldview is already calibrated to her level of material well-being, realizations in small-stakes lotteries will be too small to induce her to change θ . As a result, a fixed-worldview utility function governs her small-stakes risk aversion. In contrast, realizations in lotteries with sufficiently large stakes will generally induce her to optimize θ ex post, which means the upper envelope governs large-stakes risk aversion. Because the curvature of the fixed-worldview utility functions does not tie down the curvature of the upper envelope, our theory effectively de-couples large-stakes and small-stakes risk aversion, except that the latter places an upper bound on the former.

We formalize this intuition as follows. Consider a consumer who potentially

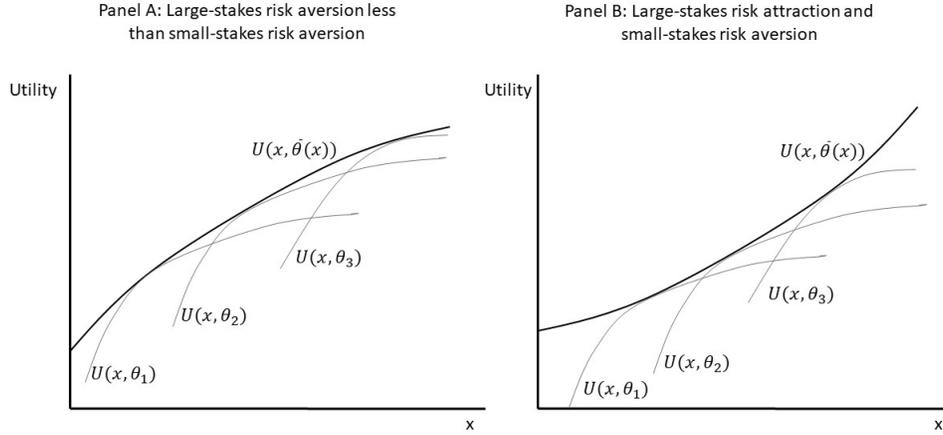


Figure 4. : Large-stakes and small-stakes risk aversion

faces choices involving gambles. She derives ex post happiness from her consumption, $x \geq 0$, which depends on her worldview as follows:

$$U(x, \theta) = \frac{\theta x^{1-\alpha}}{1-\alpha} - C(\theta),$$

where $\alpha > 0$ (excluding unity) and C is strictly increasing and continuous. As described above, the parameter θ rotates the consumer's utility function to reflect the importance she attaches to material payoffs. We adopt the following functional form for C :

$$C(\theta) = \frac{k\theta^\eta}{\eta}$$

with $\eta \neq 0$.

Decisions and events unfold in three stages:

Stage 1: The consumer starts out with materialism parameter θ_1 . She learns that she has various lottery options available and makes a choice.

Stage 2: After learning the payoff from the lottery, the consumer chooses θ_2 .

Stage 3: The consumer experiences ex post happiness based on θ_2 . If $\theta_2 \neq \theta_1$, she also incurs a utility penalty of $c > 0$. We interpret c as a fixed psychological cost of changing worldviews. We mentioned these costs briefly in Section I, but up to this point have assumed them away.³⁷

³⁷Our analysis generalizes to cases in which these costs are given by some increasing function $\Gamma|\theta_1 - \theta_2|$, provided (i) $\lim_{z \downarrow 0} \Gamma(z) > 0$, so that even small changes in worldviews entail non-negligible

In stages 1 and 2, the consumer maximizes the expectation of her realized utility, placing weight λ on her current evaluation of the outcome, and weight $1 - \lambda$ on her future valuation. Because there are no stage 2 or 3 choices other than θ , the consumer tries to maximize ex post happiness when changing her worldview. In other words, she chooses θ_2 to solve $\max_{\theta'} U(x, \theta')$. We will use $\bar{\theta}(x)$ to denote the corresponding maximizer. Considering the psychic cost associated with changing underlying attitudes, the consumer adopts a new worldview in stage 2 iff

$$(1 - \lambda) [U(x, \bar{\theta}(x)) - U(x, \theta_1)] > c$$

We also assume the consumer's initial worldview resulted from optimization within a stable environment that did not include the lottery currently under consideration. Accordingly, using y to denote her anticipated income in the absence of the lottery, her initial worldview satisfies $\theta_1 = \bar{\theta}(y)$.

The following proposition shows that $U(x, \theta_1)$ governs decisions involving small-stakes gambles, but that the following objective function governs the evaluation of large-stake gambles:

$$W(x) = \lambda U(x, \theta_1) + (1 - \lambda) U(x, \bar{\theta}(x))$$

It also characterizes risk aversion for W .³⁸ In particular, for any utility function u , the degree of relative risk aversion at x is defined as

$$r_u(x) = -\frac{xu''(x)}{u'(x)}$$

Notice that $r_U(x) = \alpha$ for all x – in other words, U is a CRRA utility function. For notational convenience, we will use $r_{\bar{U}}(x)$ to denote the degree of relative risk aversion for the function $U(x, \bar{\theta}(x))$. This quantity is of interest because, for small λ , it approximates $r_W(x)$, the degree of relative risk aversion for large-stake gambles.

PROPOSITION 11: (i) For $\alpha < 1$ (low risk aversion), $\bar{\theta}(x)$ is an interior solution iff $k > 0$ and $\eta > 1$ (case (a)). For $\alpha > 1$ (high risk aversion), $\bar{\theta}(x)$ is an interior solution iff $k < 0$, and $\eta < 1$ (case (b)). In both of these cases, for any y , there exists $\bar{c} > 0$ ($\bar{c} = +\infty$ in case (a)) with the following property: for any $c < \bar{c}$ we can select finite numbers σ_L, σ_S with $\sigma_L \geq \sigma_S$ such that consumer does not change her worldview if $|x - y| < \sigma_S$ (“small stakes”) and does change her worldview if $|x - y| > \sigma_L$ (“large stakes”).

(ii) In both cases (a) and (b): When all payoffs involve small stakes, choices over lotteries are governed by $U(x, \theta_1)$, and the degree of relative risk aversion is

costs, and (ii) $\Gamma(z)$ does not increase too rapidly with z . Condition (i) is required to ensure that the consumer does not change worldviews after receiving small payoffs. Condition (ii) is required to ensure that she changes her worldview sufficiently when she receives large payoffs.

³⁸As noted above, W does not apply when x is sufficiently close to y . However, when evaluating large-stakes gambles, one can proceed as if W applies globally.

α . When all payoffs involve large stakes, choices over lotteries are governed by $W(x)$, and the degree of relative risk aversion at any payoff x , $r_W(x)$, is strictly increasing in λ , and can take on any value between $r_{\bar{C}}(x) = \frac{\alpha\eta-1}{\eta-1}$ (which is strictly less than α) and α , depending on the value of λ . Furthermore, $r_{\bar{C}}(x)$ can take on any value between $-\infty$ (infinite risk attraction) and α (small-stakes risk aversion), depending on the value of η .

To illustrate the proposition's implications, suppose $\alpha = \frac{1}{2}$ and $\eta = 2$. In that case, as the consumer approaches perfect mindset flexibility, she becomes risk-neutral for large-stakes gambles: $r_W(x) \rightarrow 0$ for all x . With intermediate mindset flexibility, her risk aversion reflects a blend of risk aversion of degree α , and risk neutrality.

The comparative statics of large-stakes risk aversion with respect to the parameters λ and η are intuitive. Greater mindset flexibility (lower λ) decreases risk aversion because it shifts weight from fixed-worldview to flexible-worldview utility, and the latter function is, as we have noted, less concave. Greater curvature of the $C(\theta)$ function (higher values of η) imply that utility functions with large values of θ overtake functions with low values of θ only at high levels of x . Hence, the flexibility to switch between worldviews increases the second derivative of W (and thus reduces risk aversion below α) to a smaller degree when η is larger.

As we have emphasized, our theory almost completely decouples small-stakes and large-stakes risk aversion. The only connection is that the former provides an upper bound for the latter. Thus, we resolve the puzzle raised by Hansson (1988) and Rabin (2000) without invoking reference dependence and loss aversion. Moreover, our resolution is robust with respect to the introduction of random variation in y ("background risk"), provided a meaningful portion of that variation involves stakes that are too low to motivate a change of worldview. In contrast, if one assumes that a loss-averse consumer evaluates prospects relative to a single reference point, the rationalization of small-stakes risk aversion in settings with background risk requires an additional "patch" such as the assumption of narrow framing.

E. Polarization

The degree of polarization within the American public has been the subject of much commentary and research (see, e.g., Boxell, Gentzkow and Shapiro, 2017). Americans are sharply divided on a variety of issues, particular those offering binary choices. In such settings, the public must of course divide itself between the "pros" and the "cons," but one might nevertheless expect the perceived utility differential to have a single-peaked distribution. On the contrary, the distribution often appears to have two peaks, one at the extreme positive end of the spectrum and one at the extreme negative end. As a result, the two camps frequently exhibit a complete lack of understanding and sympathy for each others' positions, and they tend to characterize their differences in terms of good and evil. As

an example, consider the phenomenon of “mommy wars” (Steiner, 2007). Many women perceive that they face a binary choice between having a career and staying at home to raise their children. A debate rages over the “right” choice for the “good mother,” with much anger on both sides.

Our theory provides a potential explanation for these types of polarization phenomena. As we established in Section II.A, there is a robust tendency within this framework for people to adopt pure worldviews that credit only one point of view, rather than mixed worldviews that credit many. Polarization follows directly from this tendency.

To illustrate, we reinterpret the basic two-worldview, two-action model of Section II.A as an account of the “mommy wars.” Specifically, suppose worldview 1 emphasizes the importance of having a career, while worldview 2 stresses the value of raising a family. Correspondingly, action 1 involves holding a full-time job outside the home, while action 2 involves staying at home with children. Action 1 is optimal under worldview 1, and action 2 is optimal under worldview 2.³⁹ Consider a population for which the distribution of α is initially single-peaked and concentrated on the interior of the unit interval. The population may be heterogeneous in other respects (in other words, with respect to $u_i(j)$, λ , and δ), with some grand CDF, F , governing the joint distribution of all relevant parameters. According to Proposition 1 and 3, the population will migrate toward the two extremes, in some cases quickly and in others gradually (depending on whether one worldview happiness-dominates the other in the eyes of any particular individual). Ultimately, the entire distribution of worldviews will collapse to two atoms, one at worldview 1 and worldview 2, with actions to match.⁴⁰ Those who pursue careers and those who stay at home to raise children will both rationalize their experience by adopting a worldview under which their choice is virtuous and the alternative is not.

Might this polarization result be an artificial consequence of the stark assumption that there are only two pure worldviews? That is in fact not the case: our conclusions extend directly to settings with a continuum of pure worldviews. Imagine that people derive happiness from two goods, x and y , a “home good” and a “career good.” Here we posit that for all values of ρ in some interval $[\rho_1, \rho_2]$ (with $\rho_1 < \rho_2$), there is a pure worldview that generates the utility function $V(x, y; \rho) = \rho v(x) + (1 - \rho\phi)w(y)$ (where ϕ parametrizes the rate at which the weight on career goods declines as weight shifts to home goods). We also suppose that non-convexities in opportunity sets force women to choose between a career-oriented bundle, (x_1, y_1) , and a home-oriented bundle, (x_2, y_2) . Defining $u_i(j) = \rho_i v(x_j) + (1 - \rho_i\phi)w(y_j)$, we see that this model fits within our original

³⁹To be clear, our conclusions do not depend on which worldview is happier (that is, on whether $u_1(1)$ or $u_2(2)$ is greater).

⁴⁰Specifically, those for whom worldview 1 happiness-dominates worldview 2 would migrate to worldview 1 (by Proposition 1), those for whom worldview 2 happiness-dominates worldview 1 would migrate to worldview 2 (again by Proposition 1), and all others would migrate to one of the two pure worldviews depending on their preference parameters and initial worldview (by Proposition 3).

framework, except that it adds the intermediate worldviews. However, for any $\rho \in (\rho_1, \rho_2)$ and $j = 1, 2$, we have $V(x_j, y_j; \rho) = \alpha_\rho u_1(j) + (1 - \alpha_\rho) u_2(j)$, where $\alpha_\rho \equiv \frac{\rho_2 - \rho}{\rho_2 - \rho_1} \in (0, 1)$. Thus, mixtures of the two extreme (pure) worldviews render the existence of all other intermediate (pure) worldviews redundant. It follows that the introduction of the latter into the model changes nothing of substance.

IV. Related Literature

As we mentioned in the introduction, this paper is not the first to envision settings in which people choose their preferences. Most notably, in Akerlof and Kranton's (2000) theory of identity, people choose among "narratives," which are analogous to our notion of worldviews. One can think of their model as a static version of ours wherein the consumer has perfect mindset flexibility ($\lambda = 0$). Because imperfect mindset flexibility creates the dynamics upon which we have focused, it is central to almost all of our results. The overlap between our formal analyses is therefore minimal. We would characterize the relationship between our work and that of Rotemberg (1994), Becker and Mulligan (1997), Ng and Wang (2001), and Palacios-Huerta and Santos (2004) similarly. Welsch (2005) explores the tension between initial and adopted preferences, but his analysis is also static. As a result, he does not consider the dynamic inconsistencies that arise in settings wherein adopted preferences govern the selection of subsequent preferences, which is the central focus of our analysis.

A related set of models examine the intergenerational transmission of preferences under the assumption that each generation shapes the preferences of its successor; see, for example, Bisin and Verdier (2001), Tabellini (2008), and Doepke and Zilibotti (2017). Because these analyses depict the preferences of any particular generation as fixed once chosen, they focus on an entirely different set of applications and issues. Still, as a formal matter, there is obviously an equivalence between choosing preferences for the next generation and for the next "self." However, these studies do not focus on a formal counterpart to mindset flexibility, or on the dynamics it generates.

Another related line of research concerns the choice of expectations rather than preferences; see Brunnermeier and Parker (2005) and Benabou and Tirole (2016). One can of course think of beliefs as an aspect of preferences. That said, there are important formal differences between our framework and the ones developed in these papers. For example, in Brunnermeier and Parker's theory, people care about outcomes evaluated according to the true probabilities and the probabilities they choose to believe. In our framework, there is no counterpart to correct beliefs. Instead, people care about outcomes evaluated according to their current worldview and their future worldview, which subsequently becomes their current worldview. This process gives rise to dynamics not present in existing models of chosen expectations.

Other lines of literature are more distant from our theory, but are nevertheless related. Some studies depict preference evolution as a mechanical phenomenon,

rather than a chosen one; see, e.g., Bowles (1998), Bar-Gill and Fershtman (2005), and Ashraf and Bandiera (2017). Mechanical models of habit formation, such as Becker and Murphy (1988), fall within this broad category. There is also a connection between our work and Ostling (2009), who endogenizes moral values. To some extent, our analysis of multiple worldviews recalls other settings in which decision makers entertain multiple objectives, such as Kalai, Rubinstein and Spiegel (2002). Finally, as we have noted, the simple model of chosen preferences developed in Section I invokes a preference structure resembling the one employed in the Loewenstein, O'Donoghue and Rabin (2003) model of projection bias. As a formal matter, we depart from their framework by making the mixed worldview an object of choice.

V. Conclusions

We have developed a theory of preference formation that treats preferences as reflections of chosen worldviews. A key element of the theory is that, when preferences change, there is a tension between evaluating future (or past) consequences according to the current worldview and the worldview prevailing when those consequences materialize. The relative weight assigned to each reflects the decision maker's mindset flexibility. Intermediate degrees of mindset flexibility give rise to time inconsistency. Our analysis shows that the theory generates rich predictions concerning dynamic behavior. We have argued that the framework provides a useful unifying lens for understanding choices in a wide variety of diverse practical applications.

A natural criticism of our theory is that it involves "too many degrees of freedom." While we appreciate this concern, we see our theory as a framework for building more specific models of applied phenomena by adding structure. Analogous concerns arise in the context of theories that envision reference-dependent utility. Without a theory of reference points, such theories are notoriously hard to test, and yet the notion of reference points has proven useful in a variety of applications. It is also important to recognize that, even without additional structure, our theory has testable implications. First, at any moment in time, preferences are fixed, so behavior should respect standard axioms. Second, our theory has the strong implication that preference changes give rise to time inconsistency. Third, it implies that preferences should adapt to anticipated constraints. The last two implications clearly distinguish our theory from models in which preferences appear to evolve due to intertemporal complementarities and substitutabilities.

Another potential criticism of our theory is that we have not provided a fully satisfactory answer to our opening question, what determines preferences? In a sense, we have simply pushed that question back one level by introducing exogenous preferences that depend on chosen worldviews as well as chosen actions. We accept this characterization while nevertheless maintaining that our analysis provides useful insights into the evolution of the preferences over actions that prevail at each moment in time.

Many other applications are potentially worth pursuing. For example, in an earlier draft of this paper (Bernheim et al., 2019), we noted that our theory provides a potential explanation for the observation that people become “set in their ways” and resistant to change as they age (Blau, 1973). Our explanation follows from the assumption that, at each moment in time, people care about past experiences in addition to current and future experiences. As time passes, these experiences create “memory capital.” Because memory capital reflects past actions chosen to serve past objectives, its psychological value declines when someone changes their worldview. Older people have greater memory capital, and hence more to lose from changing their worldviews. Redefining success and failure is relatively “cheap” for the young, but to the elderly it can imply a wasted life. As a second example, we would point to the work of Atkin (2013), which suggests that theories of preference endogeneity may contribute to our understanding of international trade.

In closing, we would highlight three potential directions for future research. First, we have largely ignored issues pertaining to the generation of worldviews. Society’s menu of available worldviews expands as a result of both cross-cultural contact and “worldview entrepreneurship.” The latter phenomenon manifests itself most visibly through cults and offshoot religions. Modeling this entrepreneurship phenomenon would be of significant interest. Second, it is important to develop a better understanding of the process through which people internalize chosen worldviews and the costs of doing so. Arguably, the purpose of much psychotherapy is to help people identify and then assimilate happier worldviews; in some cases, there may also be an element of invention. Finally, our analysis has been completely silent on the critical issue of welfare. Preference endogeneity raises challenging normative questions; see, e.g., the discussions in Elster (1983), Nussbaum (2001), and Baber (2007). For example, how should one evaluate welfare in the context of applications involving oppression without coercion, as in Section III.C? Are women in gender-oppressive cultures who prefer to wear the burqa making optimal choices, or are they undermining their own well-being? To the extent one interprets preference endogeneity as reflecting intertemporal complementarities and substitutabilities, there is no scope within choice-based welfare analysis for declaring such decisions sub-optimal. In contrast, because our framework depicts preference endogeneity as giving rise to dynamic inconsistency, it accommodates the view that the optimality of a given action depends on the evaluator’s perspective. A mechanical application of the Bernheim and Rangel (2009) framework for choice-oriented behavioral welfare economics is certainly feasible, but may not prove discerning without further elaboration and refinement.

REFERENCES

- Akerlof, George A.** 1980. “A Theory of Social Custom, of Which Unemployment May Be One Consequence.” *Quarterly Journal of Economics*, 94(4): 749–

775.

- Akerlof, George A., and Rachel E. Kranton.** 2000. "Economics and Identity." *Quarterly Journal of Economics*, 115(3): 715–753.
- Asch, Solomon E.** 1956. "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied*, 70(9).
- Ashraf, Nava, and Oriana Bandiera.** 2017. "Altruistic Capital." *American Economic Review: Papers and Proceedings*, 5(107): 70–75.
- Atkin, David.** 2013. "Trade, Tastes and Nutrition in India." *American Economic Review*, 103(5): 1629–1663.
- Atkin, David, Eve Colson-Sihra, and Moses Shayo.** 2019. "How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption." *Working Paper*.
- Baber, Harriet E.** 2007. "Adaptive Preference." *Social Theory and Practice*, 33(1): 105–126.
- Banerjee, Abhijit.** 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107(3): 797–817.
- Bar-Gill, Oren, and Chaim Fershtman.** 2005. "Public Policy with Endogenous Preferences." *Journal of Public Economic Theory*, 7(5): 841–857.
- Becker, Gary S., and Casey B. Mulligan.** 1997. "The Endogenous Determination of Time Preference." *The Quarterly Journal of Economics*, 112(3): 729–758.
- Becker, Gary S., and Kevin M. Murphy.** 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96(4): 675–700.
- Benabou, Roland.** 2013. "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies*, 80(2): 429–462.
- Benabou, Roland, and Jean Tirole.** 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, 30(3): 141–164.
- Benjamin, Daniel J., James J. Choi, and Joshua A. Strickland.** 2010. "Social Identity and Preferences." *American Economic Review*, 100(4): 1913–1928.
- Bernheim, B. Douglas.** 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–877.

- Bernheim, B. Douglas, and Antonio Rangel.** 2009. "Beyond Revealed Preference: Choice-theoretic Foundations for Behavioral Welfare Analysis." *Quarterly Journal of Economics*, 124(1): 51–104.
- Bernheim, B. Douglas, Luca Braghieri, Alejandro Martínez-Marquina, and David Zuckerman.** 2019. "A Theory of Chosen Preferences." *Working Paper, Stanford University*.
- Bikhchandani, Sushi, David Hirshleifer, and Ivo Welch.** 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100(5): 992–1026.
- Bisin, Alberto, and Thierry Verdier.** 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." *Journal of Economic Theory*, 97(2): 298–319.
- Blau, Z. S.** 1973. *Old Age in a Changing Society*. Oxford, England: New Viewpoints.
- Bowles, Samuel.** 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economic Literature*, 36(1): 75–111.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro.** 2017. "Is the Internet Causing Political Polarization? Evidence from Demographics." *NBER Working Paper No. 23258*.
- Brunnermeier, Markus, and Jonathan Parker.** 2005. "Optimal Expectations." *American Economic Review*, 95(4): 1092–1118.
- Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger.** 2014. "Violence and Risk Preferences: Experimental Evidence from Afghanistan." *American Economic Review*, 104(1): 123–148.
- Cameron, Lisa, and Manisha Shah.** 2015. "Risk-taking Behavior in the Wake of Natural Disasters." *Journal of Human Resources*, 50(2): 484–515.
- Chen, Yan, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li.** 2010. "Social Comparisons and Contributions to Online Communities: A Field Experiment on Movie-Lens." *American Economic Review*, 100(4): 1358–1398.
- Chen, Yuyu, and David Yang.** 2019. "The Impact of Media Censorship: 1984 or Brave New World." *American Economic Review*, 109(6).
- Cohn, Alain, Ernst Fehr, and Michel Andre Marechal.** 2014. "Business Culture and Dishonesty in the Banking Industry." *Nature*, 516: 86–89.
- Cohn, Alain, Jan Engelmann, Ernst Fehr, and Michel Andre Marechal.** 2015. "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals." *American Economic Review*, 105(2): 860–885.

- Danzer, Alexander M., Natalia Danzer, and Ernst Fehr.** 2016. "The Behavioral and Psychological Consequences of a Nuclear Catastrophe: The Case of Chernobyl." *Working Paper*.
- Dennett, Daniel C.** 2006. *Breaking the Spell*. New York: Penguin Group.
- Doepke, Matthias, and Fabrizio Zilibotti.** 2017. "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica*, 85(5): 1331–1371.
- Dupas, Pascaline.** 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica*, 82(1): 197–228.
- Eckel, Catherine C., Mahmoud A. El-Gamal, and Rick K. Wilson.** 2015. "Risk Loving After the Storm: A Bayesian-Network Study of Hurricane Katrina Evacuees." *Journal of Economic Behavior & Organization*, 69(2): 110–124.
- Elster, Jon.** 1983. *Sour Grapes*. Cambridge University Press.
- Fehr, Ernst, and Karla Hoff.** 2011. "Tastes, Castes, and Culture: The Influence of Society on Preferences." *Economic Journal*, 121: F396–F412.
- Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti.** 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *Quarterly Journal of Economics*, 119(4): 1249–1299.
- Festinger, Leon.** 1954. "A Theory of Social Comparison Processes." *Human Relations*, 7(2): 117–140.
- Ford, T. E.** 1997. "Effects of Stereotypical Television Portrayals of African-Americans on Person Perception." *Social Psychology Quarterly*, 60: 266–275.
- Fouka, Vasiliki.** 2019. "How Do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I." *American Political Science Review*, 113(2): 405–422.
- Fouka, Vasiliki.** 2020. "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I." *The Review of Economic Studies*, 87(1): 204 – 239.
- Genicot, Garance, and Debraj Ray.** 2017. "Aspirations and Inequality." *Econometrica*, 85(2): 489–519.
- Gruber, Jonathan, and Daniel M. Hungerman.** 2008. "The Church Versus the Mall: What Happens When Religion Faces Increased Secular Competition?" *The Quarterly Journal of Economics*, 2(123): 831–862.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales.** 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.

- Guiso, Luigi, Paola Sapienza, and Luigi Zingales.** 2018. "Time Varying Risk Aversion." *Journal of Financial Economics*, 128(3): 403–421.
- Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe.** 2018. "Do Risk Preferences Change? Evidence from the Great East Japan Earthquake." *American Economic Journal: Applied Economics*, 10(2): 298–330.
- Hansson, B.** 1988. "Risk Aversion as a Problem of Conjoint Measurement." In *Decision, Probability, and Utility.*, ed. P. Gardenfors and N.-E. Sahlin, 136–158. Cambridge University Press.
- Ichino, Andrea, and Giovanni Maggi.** 2000. "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm." *The Quarterly Journal of Economics*, 115(3): 1057–1090.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler.** 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98(6): 1325–1348.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler.** 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5(1): 193–206.
- Kalai, Gil, Ariel Rubinstein, and Ran Spiegler.** 2002. "Rationalizing Choice Functions by Multiple Rationales." *Econometrica*, 70(6): 2481–2488.
- Kapteyn, Arie, Rob Alessie, and Anamaria Lusardi.** 2005. "Explaining the Wealth Holdings of Different Cohorts: Productivity Growth and Social Security." *European Economic Review*, 49: 1361–1391.
- Knetsch, Jack L.** 1989. "The Endowment Effect and Evidence of Nonreversible Indifference Curves." *American Economic Review*, 79(5): 1277–1284.
- Loewenstein, George, Ted O'Donoghue, and Matthew Rabin.** 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics*, 118(4): 1209–1248.
- Marzilli-Ericson, Keith M., and Andreas Fuster.** 2011. "Expectations as Endowments: Evidence on Reference-dependent Preferences from Exchange and Valuation Experiments." *The Quarterly Journal of Economics*, 126(4): 1879–1907.
- Marzilli-Ericson, Keith M., and Andreas Fuster.** 2014. "The Endowment Effect." *Annual Review of Economics*, 6: 555–579.
- Morewedge, Carey K., Lisa L. Shu, Daniel T. Gilbert, and Timothy D. Wilson.** 2009. "Bad Riddance or Good Rubbish? Ownership and Not Loss Aversion Causes the Endowment Effect." *Journal of Experimental Social Psychology*, 45(4): 947–951.

- Morwedge, Carey K., and Colleen C. Giblin.** 2015. "Explanations of the Endowment Effect: An Integrative Review." *Trends in Cognitive Sciences*, 19(6): 339–348.
- Ng, Yew-Kwang, and Jianguo Wang.** 2001. "Attitude Choice, Economic Change, and Welfare." *Journal of Economic Behavior and Organization*, 45: 279–291.
- Nunn, Nathan, and Leonard Wantchekon.** 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review*, 101(7): 3221 – 3252.
- Nussbaum, Martha C.** 2001. "Symposium on Amartya Sen's Philosophy: 5 Adaptive Preferences and Women's Options." *Economics and Philosophy*, 17: 67–88.
- O'Donoghue, Ted, and Matthew Rabin.** 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103–124.
- Ostling, Robert.** 2009. "Economic Influences and Moral Values." *The B.E. Journal of Economic Analysis and Policy*, 9(1).
- Palacios-Huerta, Ignacio, and Tano J. Santos.** 2004. "A Theory of Markets, Institutions, and Endogenous Preferences." *Journal of Public Economics*, 88(3–4): 601–627.
- Pascal, Blaise.** 1958. *Pascal's Pensees*. New York: E. P. Dutton.
- Pettigrew, Thomas F., and Linda R. Tropp.** 2006. "A Meta-analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology*, 90(5): 751–783.
- Rabin, Matthew.** 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–1292.
- Rao, Gautam.** 2019. "Familiarity does not breed contempt: Generosity, Discrimination, and Diversity in Delhi Schools." *American Economic Review*, 109(3): 774–809.
- Rayo, Luis, and Gary S. Becker.** 2007. "Evolutionary Efficiency and Happiness." *Journal of Political Economy*, 115(2): 302–337.
- Rotemberg, Julio.** 1994. "Human Relations in the Workplace." *Journal of Political Economy*, 102(1): 684–717.
- Schiappa, E., P. B. Gregg, and D. E. Hewes.** 2006. "Can One TV Show Make a Difference? Will and Grace and the Parasocial Contact Hypothesis." *Journal of Homosexuality*, 51: 15–37.

- Sen, Amartya.** 1985. *Commodities and Capabilities*. Amsterdam: North Holland.
- Steiner, Leslie Morgan.** 2007. *Mommy Wars*. New York: Random House.
- Tabellini, Guido.** 2008. "The Scope of Cooperation: Values and Incentives." *The Quarterly Journal of Economics*, 123(3): 905–950.
- Voors, Maarten J., Eleonora E.M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest.** 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review*, 102(2): 941–964.
- Washington, Ebonya.** 2008. "Female Socialization: How Daughters Affect Their Legislator Fathers." *American Economic Review*, 98(1): 311–332.
- Welsch, Heinz.** 2005. "Adaptation of Tastes to Constraints." *Theory and Decision*, 57: 379–395.
- Wolpe, J.** 1961. "The Systematic Desensitization Treatment of Neuroses." *Journal of Nervous and Mental Disease*, 132: 189–203.