

Does the Law Change Preferences?

*Jennifer Arlen and Lewis A. Kornhauser**

“I would prefer not”¹

Scholars have recently challenged the claim in classical deterrence theory that law influences behavior only through the expected sanction imposed. Some go further and argue that law may also “shape preferences,” changing people’s wants and values. In this Article, we analyze existing claims that criminal and civil law alter preferences and conclude that none suggest that the law shapes preferences. We first clarify this preference-shaping claim by elaborating the structure of rational choice theory generally and “preference” in particular. We then investigate three mechanisms of legal influence suggested by the preference-shaping literature: (1) the “serious harm” mechanism; (2) the “social norm” mechanism; and (3) the “self-improvement” mechanism. We then show that each of these mechanisms operates by changing the agent’s beliefs about the attributes or consequences of her choice options rather than by changing her preferences.

INTRODUCTION

How do laws that regulate harmful conduct influence behavior? This question lies at the heart not only of economic analysis of law, but also of policy analysis generally as well as of any social science that seeks to understand the effects of legal rules and institutions on society. From the beginning of

* The authors are, respectively, the Norma Z. Paige Professor of Law and the Frank Henry Sommer Professor of Law, at New York University School of Law. We would like to thank the journal’s editors, Oren Bar-Gill, Louis Kaplow, Shari Diamond, Kathy Spier, and participants at the Harvard Law and Economics Workshop, the Northwestern Faculty Workshop, the Stanford Law School Law and Economics Workshop and the Conference on How Law Changes What You Want: Positive and Normative Effects of Law on Values and Preferences, which was held at the University of California at Berkeley School of Law.

1 HERMAN MELVILLE, *BARTLEBY THE SCRIVENER: A STORY OF WALL STREET* (1853), reprinted in *THE PIAZZA TALES* 32, 48 (London, Sampson Low, Son & Co. 1856).

the second law and economics movement in the 1960s² through most of the 1980s, the economic analysis of law offered a simple answer to this question: classical deterrence theory. Classical deterrence theory posits that both civil and criminal law influence behavior through the threat of sanctions that raise the expected costs of the illegal actions and thereby provide people with an incentive to eschew self-interested but socially harmful behavior.³

Yet this explanation of how civil and criminal law influences behavior is incomplete. The imposition of criminal liability for harmful conduct appears to generate more compliance than would result from the imposition of civil liability or a Pigovian tax with the same expected cost.⁴ Criminal law appears to provide people with additional reasons to change their behavior than the simple threat of financial sanction or imprisonment.⁵

Scholars within and outside of the economic analysis of law claim that adoption and enforcement of criminal and civil laws regulating harmful conduct can deter such conduct through their negative expression of the value of the regulated conduct. Specifically, they claim that the criminal and civil laws' expression that certain conduct is socially undesirable (for example, because it harms others or violates social or moral norms) directly alters behavior by "shaping preferences," leading people to eschew socially undesirable conduct for reasons other than the threat of sanctions.⁶

2 The second law and economics movement is generally agreed to have begun with the publication of Ronald Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1 (1960) and Guido Calabresi, *Some Thought on Risk Distribution in the Law of Torts*, 70 YALE L.J. 499 (1961). The first law and economics movement occurred during the progressive era. For a detailed history, see BARBARA H. FRIED, *THE PROGRESSIVE ASSAULT ON LAISSEZ FAIRE* (1998).

3 Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

4 For example, American's seatbelt usage increased from 10% to more than 50% following the adoption of laws requiring seatbelts. Scholars argue that since noncompliance is rarely detected and sanctioned, these laws likely altered behavior through channels other than the threat of sanctions. Benjamin van Rooji, *Homo Juridicus: Questioning Legal Assumptions About Behavior* (working paper 2020) (on file with author).

5 See *infra* note 6; see also Lewis Kornhauser, *The Economic Analysis of Law*, in THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY § 4 (Edward N. Zalta ed., Fall 2017 Edition), <https://plato.stanford.edu/archives/fall2017/entries/legal-econanalysis/> offers a more recent account of the normativity of law.

6 The extensive literature on the preference-shaping effect of law includes Oren Bar-Gill & Chaim Fershtman, *Law and Preferences*, 20 J. L. ECON. & ORGAN. 331 (2004); Robert Cooter, *Do Good Laws Make Good Citizens: An Economic Analysis of Internalized Norms*, 86 VA. L. REV. 1577 (2000) [hereinafter Cooter,

Evaluation of these claims depends on one's understanding of what "preference-shaping" entails. Leading authors have adopted a strong interpretation of preference-shaping. They claim that the law alters behavior by changing what people want and value.⁷ By contrast, others suggest that the law alters behavior by changing the perceived cost to people of engaging in prohibited conduct through channels other than expected sanctions, without actually changing people's preferences.⁸

The difference between these two understandings of preference-shaping has important implications for both legal policy and the economic analysis of law. The claim that the law can alter people's values has two important legal policy prescriptions. First, the claim that the mere adoption of a law can alter behavior by creating internal motivations to eschew harmful conduct implies that the law should reduce its reliance on monitoring, enforcement and sanctions,

Good Citizens]; Robert Cooter, *Expressive Law and Economics*, 27 J. LEGAL STUD. 585 (1998); Robert Cooter, *Models of Morality in Law and Economics: Self-Control and Self-Improvement for the Bad Man of Holmes*, 78 B.U. L. REV. 903 (1998); Kenneth G. Dau-Schmidt, *An Economic Analysis of the Criminal Law as a Preference-Shaping Policy*, 1990 DUKE L.J. 1 (1990); Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 U. CHI. L. REV. 1129, 1136-37 (1986); LYNN STOUT, *CULTIVATING CONSCIENCE: HOW GOOD LAWS MAKE GOOD PEOPLE* (2011); see also Kenworthy Bilz & Janice Nadler, *Law, Moral Attitudes, and Behavioral Change*, in THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 241 (Eyal Zamir & Doran Teichman eds., 2014); Yuval Feldman & Yotam Kaplan, *Preference Change and Behavioral Ethics: Can States Create Ethical People?*, 22 THEORETICAL INQUIRIES L. 85 (2021) (explaining why promoting explicit ethical preferences does not necessarily lead to ethical behavior); Ariel Porat, *Changing People's Preferences by the State and the Law*, 22 THEORETICAL INQUIRIES L. 215 (2021) (discussing why it may be optimal to change preferences); see also John C. Harsanyi, *Welfare Economics of Variable Tastes*, 21 REV. ECON. STUD. 204, 213 (1953-54).

7 Bar-Gill & Chaim Fershtman, *supra* note 6, at 332 ("different legal systems may affect not just the behavior of individuals, but who they are."); Cooter, *Do Good Laws*, *supra* note 6 (laws can cause people to change their moral values); Dau-Schmidt, *supra* note 6, at 2, 14 (criminal law seeks to establish new positive "norms of individual behavior by shaping the preferences of criminals and the public at large"); Stout, *supra* note 6, at 228 ("criminal law changes *what people want*, in the process shifting their behavior from purely selfish and asocial to unselfish and law-abiding.").

8 E.g., Daniel Kahan, *Social Influence, Social Meaning, and Deterrence*, 83 VA. L. REV. 349 (1997).

which are very costly.⁹ Second, some scholars argue that expressive laws that change preferences should be used sparingly because changing people's basic wants and desires infringes on free will and individual autonomy.¹⁰

In addition, the strong claim has profound implications for the economic analysis of law, as well as for economic theory. If the law can indeed change people's wants and desires, then a foundational tenet of neoclassical economics — that people have stable and exogenously determined preferences — crumbles.¹¹ It would also undermine the normative premise that policymakers should favor legal rules that maximize social welfare as determined by people's preferences, because preferences cannot necessarily be used as a metric to select between legal rules if the legal rules alter preferences.¹²

By contrast, if expressive law affects behavior by altering beliefs or social meaning without fundamentally changing what people value, then expressive law would be consistent with rational choice theory and deterrence theory, and would provide an important complement to (instead of serving as a perfect substitute for) legal sanctions. Such interventions also would not undermine autonomy. Thus, it is important to both legal policy and the normative validity of economic theory to evaluate whether claims that the law affects behavior

-
- 9 *E.g.*, Bilz & Nadler, *supra* note 6, at 241; Cooter, *Do Good Laws*, *supra* note 6; Sunstein, *supra* note 6, at 1137 (if new legal rules can change preferences then people will not attempt to circumvent the new rules); *see also* Gerard E. Lynch, *The Role of Criminal Law in Policing Corporate Misconduct*, 60 *LAW & CONTEMP. PROBS.* 23, 46 (1997) (to deter, people need to be induced to obey the law for reasons of conscience and conviction, and not out of fear of punishment). By contrast, if the law operates through other expressive channels such as norms, enforcement is needed in order to establish the norm. *See* Kahan, *supra* note 8, at 354-55, 378-82; *see also* van Rooji, *supra* note 4, at 1 (discussing how seatbelt compliance increased from 50% to 85% after states started advertising the legal and personal risk of noncompliance).
- 10 *E.g.*, Dau-Schmidt, *supra* note 6, at 22-23 (arguing that the law should reserve preference-shaping for limited situations because it “violate[s] individual sovereignty in [an] intimate way”); *see* Sunstein, *supra* note 6. For a discussion of why states may be able to legitimately alter people's preferences, *see* Sunstein, *supra* note 6, and Porat, *supra* note 6.
- 11 This possibility raises a host of concerns about the appropriate role of the state, as discussed in Porat, *supra* note 6.
- 12 *E.g.*, Dau-Schmidt, *supra* note 6, at 16 (the normative validity of Pareto optimality and social welfare are undermined if laws alter preferences); Sunstein, *supra* note 6, at 1145; *but see* Louis Kaplow & Steven Shavell, *Fairness versus Welfare*, 114 *HARV. L. REV.* 961, 1338 (2001) (the appropriateness of using the law to shape preferences depends on whether total welfare following the legal change would exceed total welfare prior to the legal change).

through expressive channels indeed involves the law altering fundamental preferences, changing what people want and value.

In this Article, we examine the various mechanisms through which the expressive law literature has asserted that the criminal and civil law alter behavior other than by imposing sanctions. We conclude that none of the hypothesized mechanisms support the proposition that the law changes preferences, in the sense of changing what people want and value. Instead, these mechanisms lead people to view the prohibited conduct as more costly to them than they previously thought. Thus, we argue that expressive effects of the law do not pose a threat to either normative or positive methods of economic analysis of law. As the law does not change preferences, at least in the short run,¹³ the normative basis of welfare economics is not threatened. Our analysis also reveals that expressive legal interventions do not present autonomy concerns. Finally, our analysis implies that the adoption of laws that alter behavior through expressive channels does not counsel against expenditures on detection and enforcement. To the contrary, as we will explain, detection and enforcement are essential to the law's ability to achieve its expressive aims by altering the perceived costs of engaging in prohibited conduct.

Our analysis proceeds in three steps. Part I explicates concepts that lie at the core of this debate: preferences, rational choice theory, classical deterrence, and general deterrence. To explicate the disconnect between the explanations of how legal preference-shaping occurs and the strong claims made about the implications of the preference-shaping role of law, we first elucidate the economic concept of preference. We distinguish between the choices people make and the underlying preferences that determine these choices. These preferences derive from the agent's assessment of which potential attributes are material to their well-being. These "criteria of evaluation" are integrated to form the agent's all-things-considered preferences that are used to rank the choices available to them. We show that changing preferences requires more than altering an individual's decisions. One must change either the individual's criteria of evaluation — what the individual values — or the process of integrating these criteria into the individual's all-things-considered preferences.

Part II presents a deeper analysis of the concept of preferences and explains what it would mean for the law to change preferences. We set out the common structure of rational choice models, and then situate both classical deterrence theory and general deterrence theory within rational choice theory. Rational choice models are very flexible. We show that rational choice and general

13 See *infra* note 104 and accompanying text (discussing how the law may affect preferences over many years by changing the environment).

deterrence theories can accommodate, once their structure is understood, many additional ways for the law to change behavior beyond government-imposed sanctions.

Finally, Part III identifies the three channels through which the literature claims that the law shapes preferences, showing that none involves changing preferences and all are compatible with deterrence theory. According to the first mechanism, call it the “serious harm” mechanism, a law regulating or prohibiting specific conduct expresses the seriousness of the harm that the conduct causes. In light of this expression, an agent may adjust her behavior to reflect an updated belief about the consequences of her action. The second mechanism, the “social norm” mechanism, claims that a regulatory law may either establish a social or moral norm or make an otherwise ignored social or moral norm salient to the actor, thereby inducing her to change her behavior. The third mechanism, the “self-improvement” mechanism, asserts that a legal rule may change the agent’s willingness to comply with the law.¹⁴ We assume that each of these mechanisms can alter behavior and then assess whether any of them indeed involves changing people’s preferences. We show that none of these mechanisms operates to change “preferences,” in the sense of changing what people value.¹⁵ Rather, each of these mechanisms relies on changing agents’ beliefs about the attributes of the prohibited conduct, causing them to conclude that the conduct would have greater undesirable characteristics than they previously understood. Thus, each mechanism aligns with a channel for

14 *E.g.*, Bar-Gill & Fershtman, *supra* note 6, at 332; Cooter, *Good Citizens*, *supra* note 6; Sunstein, *supra* note 6, at 1136. *See also* Stout, *supra* note 6, at 228 (“[B]y changing the social context, *criminal law changes what people want*, in the process shifting their behavior from the purely selfish and asocial to unselfish and law-abiding.”). For an excellent discussion of the literature, see Feldman & Kaplan, *supra* note 6; Porat, *supra* note 6.

15 Our conclusion that the leading theories of how the law can change preferences do not, in fact, entail a change to people’s preferences does not imply that preferences are immutable. For example, physical injury, trauma, and other such interventions can be expected to change people’s preferences. *See* Jennifer Arlen, *Reconsidering Efficient Tort Rules for Personal Injury: The Case of Single-Activity Accidents*, 32 WM. & MARY L. REV. 41 (1991). Changes to the social or educational context in which people are raised also may cause people to value different things than they would have otherwise. *See* text accompanying note 104. None of these channels for changing values provides support for claims that the law changes preference in the short run or can eschew enforcement by immediately altering people’s underlying wants and desires. *See id.*

altering behavior that is consistent with rational choice theory.¹⁶ We conclude that the “preference-shaping” effect of law is thus consistent with both rational choice and general deterrence theory, and is a complement to legal sanctions, instead of a perfect substitute for it.

I. PREFERENCES AND RATIONAL CHOICE THEORY

The claim that the law alters preferences by changing what people value, and thus who they are, presents a direct challenge to both rational choice theory and to deterrence theory. Evaluating this claim requires that we both identify the understanding of the term “preferences” required to support this strong claim and explicate the structure of rational choice theory. In this Part, we first distinguish different meanings of the word “preference,” identifying the conception associated with the strongest claims of the various meanings. We then outline the common structure of the models that comprise rational choice theory. We then situate classical deterrence theory within this structure. Finally, we offer a richer account of the structure of rational choice models, and an expanded theory of deterrence that will enable us to demonstrate how existing preference-shaping claims are consistent with both rational choice and general deterrence theory.

A. Disentangling Choices, Tastes, Preferences, and Criteria of Evaluation

Linguistic confusion besets discussions of both preference-shaping and rational choice theory. The confusion arises in part because economists have adopted words as technical terms that also have lay meanings. Discussion thus often conflates the lay with the technical meaning.

The term “preference” causes some of the greatest confusion because it has not only multiple lay usages but also multiple connotations in economics itself.¹⁷ In lay usage, the term “preference” means a “taste,” “desire” or “inclination”

16 Our conclusion that the identified mechanisms of preference-shaping are consistent with rational choice theory does not imply that we believe people always conform to rational choice theory. An extensive empirical, experimental and theoretical literature reveals that people both regularly and systematically deviate from rational choice theory in certain situations. We simply conclude that the existing claims that the law shapes preferences do not, as currently elucidated, represent a challenge to either rational choice theory or to deterrence theory.

17 The term “utility” is the only term that perhaps causes more confusion than “preference.” These two technical concepts are intimately linked, as we will discuss below. *See infra* notes 42 and 48.

for something. So, someone might say “Alice has a preference for vanilla ice cream,” meaning simply that Alice likes vanilla ice cream. Thus, Alice has an inclination for vanilla, rather than, strawberry (or chocolate, mocha, or other) ice cream.¹⁸

For economists, “preference” sometimes means “revealed preference,” i.e., simply the set of choices that the agent has made in a series of different environments; confusingly, “revealed preference” also can refer to a specific choice.¹⁹ To avoid confusion, throughout this Article, we never use “preference” in this sense. The agent’s choices are just that — “choices.”

More often, economists use the term “preference” to refer to the technical concept of a ranking of objects in some domain.²⁰ But this usage is itself

18 These three terms have subtly different meanings. A desire for vanilla ice cream suggests an operative mental state at a specific time. A taste for vanilla ice cream suggests an appreciation of it, while an inclination for vanilla ice cream suggests a disposition to choose vanilla when confronted with a choice.

19 “Revealed preference” refers to the theory introduced by Paul Samuelson that states that a choice of *a* over *b* reveals a preference for *a* over *b*. Paul Samuelson, *A Note on the Pure Theory of Consumer Behavior*, 5(17) *ECONOMICA* 61 (1938). The theory assumes that the domain of preference corresponds to the domain of choice. In many contexts, as illustrated below in, for instance, our election example, these two domains differ. See *infra* notes 20 and 44. For further discussion, see Marcel K. Richter, *Revealed Preference Theory*, in 7 *THE NEW PALGRAVE DICTIONARY OF ECONOMICS* 151 (Steven N. Durlauf & Lawrence E. Blume eds., 2d ed. 2008).

20 The domain of a preference is the set of objects over which the agent has preferences. The standard (single-period) model of consumer theory assumes that the agent has narrowly self-interested preferences over consumption bundles. The set of possible consumption bundles constitutes the domain of preference. Yet the domain of preference can differ from the domain of choice. See *infra* note 44. For example, models of consumption over time assume that the agent has preferences over streams of consumption bundles occurring over her lifetime, one in each period. In these models, the agent cannot choose her lifetime consumption at time zero. Instead, in each period she chooses a consumption bundle. Thus, her domain of choice — the consumption bundle in each period — differs from her domain of preference — a stream of consumption bundles. To facilitate analysis, the models often assume that the preferences over streams of consumption are separable in time. Separability implies that the agent has a derived preference over consumption bundles. But this assumption is very strong. For example, an agent’s preferences over what car to purchase in period 1 is likely to depend on whether she wants to buy a house or pay for graduate school in period 4. For further discussion, see Jean-Pierre Benoît & Lewis Kornhauser, *Assembly-*

ambiguous, as an agent may have multiple rankings of a single domain.²¹ Primarily, “preference” in this sense refers to an agent’s all-things-considered ranking that guides her choice among the options available to her.²² Throughout this Article, the term “preferences” refers to this (motivational) all-things-considered ranking of elements in the domain of preference.²³ For reasons that will become apparent in Part I.C., we will sometimes refer to these as “fundamental preferences” or “basic preferences.”²⁴

Occasionally, “preference” for an economist refers to the criteria of evaluation that a person employs when deliberating.²⁵ These criteria are integrated into the agent’s all-things-considered ranking — her basic preference. Moreover,

Based Preferences, Candidate-Based Procedures, and the Voting Rights Act, 68 S. CAL. L. REV. 1503 (1995).

- 21 Agents may have multiple rankings in a single domain because preferences serve both an explanatory and an evaluative role. Preferences represent an evaluation of objects of choice with respect to facets of choice (e.g., cost, appearance etc.) that the agent cares about. Explanatorily, the preferences represent her motivations; normatively, they represent her well-being. When an agent’s preferences are not narrowly self-interested these motivational and evaluative rankings may diverge. A parent, for instance, may sacrifice her well-being for the well-being of her child by working three jobs so the child may go to college. The parent’s motivational preferences include both her own and the child’s well-being, but her “well-being” preferences (used in the Pareto criterion or in utilitarian evaluation) concern only her own well-being.
- 22 See, e.g., DANIEL HAUSMAN, PREFERENCE, VALUE, CHOICE, AND WELFARE (2012). Note that we focus on what we might call the agent’s explanatory or motivational preferences/ranking.
- 23 In most economic models, the domain of preference is assumed to be identical to the domain of choice. For further discussion, see *supra* note 20 and *infra* notes 24 and 44.
- 24 In analyzing a problem, the analyst attributes preferences (and hence a domain of preference) to each agent. The attributed preferences typically are not “fundamental” or “basic” in some absolute sense, only basic or fundamental to the choices being analyzed. Preferences that are “basic” in a model of consumer choice are unlikely to be “basic” (or even relevant) in a model of electoral choice. The analyst’s choice of the domain of preference balances considerations of realistic psychology and analytic tractability. To derive truly fundamental preferences, one might imagine that each agent has preferences over life histories, but the gap between the grand scale of this domain of preference (life histories) and the relevant choices made by an agent at each period of time would make this model completely intractable.
- 25 See, e.g., Till Grune-Yanoff & Sen Ove Hansson, *Introduction*, in PREFERENCE CHANGE 1, 1-26 (Grune-Yanoff & Sen Ove Hansson eds., 2007).

each criterion of evaluation ranks the same objects ranked by the basic preference, but does so by evaluating each option along a single dimension (or on the basis of a single attribute). One might say, for example, that “Bob has a preference for norm compliance,” meaning he evaluates his actions on the basis of, among other attributes, whether they comply with relevant norms.²⁶ Throughout this Article, we call these rankings “criteria of evaluation.” In order for an intervention to change preferences by changing what people want and value, it would need either to alter their criteria of evaluation — to change the set of attributes they value the most — or to change the way the criteria of evaluation are integrated into their all-things-considered rankings.

To explicate the differences between these concepts, and how we are using these terms, it is useful to apply them to a decision of an individual, Liza, who is ordering a meal in a restaurant. Liza likely does not treat her choice of appetizer, main course and dessert as independent choices. Her domain of choice thus would be the whole meal. The choices she actually makes will depend on her preferences — that is, her all-things-considered ranking of the possible complete meals she can create from the menu. These preferences integrate the criteria of evaluation against which she evaluates a meal. These criteria of evaluation might include taste, creaminess, cholesterol count, calories, price, texture, social consciousness, and the way in which dishes consumed in different courses might complement each other.²⁷ Her preference ranking thus depends on both her criteria of evaluation and her assessment of the attributes (e.g., the price, calorie content, etc. of each possible meal). Based on her preferences and her beliefs about the attributes of each possible meal, Liza chooses the meal that ranks highest in her all-things-considered preference. Her actual choice of meal results from her preferences, but it is not, in our terminology, a preference itself; it is her “choice.”

Thus, we can conclude that an intervention alters or shapes Liza’s preferences if it causes her to alter her criteria of evaluation — for example, by causing her to value a new attribute, such as whether the food was locally sourced.

26 *Id.* at 1-26.

27 Suppose that Liza went to the restaurant with Fred and Henry and they decided to delegate the choice of appetizer to Liza, main course to Fred, and dessert to Henry. The domain of preference for each person would remain the meal, but the domain of choice for each would be the specified course. In this context, Liza (or Fred or Henry) might have *derived* preferences over appetizers (or main courses or desserts); such derived preferences exist only if the preferences over meals are separable in courses, an unlikely event. When preferences over meals are not separable, they nonetheless guide Liza’s choice but in a more complex way.

Alternatively, it could cause her to change the manner in which she integrates her existing criteria of evaluation into her all-things-considered preferences.

By contrast, an intervention could alter her meal choices, without altering her preferences, if it caused her to reassess the attributes of a particular meal choice given her existing preferences. For example, a person who values not eating highly sentient creatures could be induced to stop eating octopus once educated about octopus' intelligence and consciousness, without altering her underlying criteria of evaluation. Her choices would change but her underlying preferences would not.

B. The Common Structure of Rational Choice Models

Rational choice theory consists of models that share a common structure: An agent with given beliefs and preferences chooses, from the available options and in a fixed environment, the action that makes her best off (given her preferences).²⁸ These models may vary in each of these five elements: (1) the environment in which the agent acts;²⁹ (2) the set of options available to the agent;³⁰ (3) the beliefs ascribed to the agent;³¹ (4) the agent's all-things-considered preferences that are constructed from her criteria of evaluation; and (5) the decision-rule employed to identify the "best" option, given the agent's preferences.³²

28 The theory thus consists of a large number of models of different phenomena or a variety of models of a single phenomenon. The varied literature on the choice between settlement and litigation illustrates the variety of models analyzing an identical phenomenon.

29 The "environment" refers to all external factors affecting the costs and benefits of a particular choice, other than a change in the set of options available. The environment includes expected legal sanctions, market prices, and any costs from reputational damage.

30 For example, in our meal example the set of options would include all the meals that Liza may create from choices available for the appetizer, main course and dessert.

31 Agents rarely are fully informed about all the attributes of the available options. Instead, they form beliefs about them based on the available information. Thus, in assessing appetizers, Liza might rank the octopus higher than the beef tartar if she believes that cows are more sentient, but might alter her ranking upon learning about octopuses' intelligence.

32 In game theory, for example, each solution concept specifies a distinct account of rationality. An equilibrium in maximin strategies, for example, offers a different account of rationality than Nash equilibrium. Similarly, different refinements of Nash equilibrium elaborate the concept of rationality differently.

Applications of rational choice theory typically identify three of these five elements as levers that the law might manipulate to influence the agent's choices, including choices regarding whether to violate the law. First, law can, in several ways, change the environment in which the agent acts in a way that alters the attributes of the options available to the agent. It can directly alter the costs or benefits of making certain choices, for example, by imposing fines on those who take certain prohibited actions. It also can alter consequences indirectly by altering costs imposed by others on certain actions. For example, laws prohibiting specific conduct may cause people who engage in that conduct to suffer costs from reputational damage. It also can alter costs indirectly by altering the internal benefit or cost that a person experiences when making a decision. For example, the law might induce a person to experience enhanced well-being from selecting options that the law has identified as promoting social welfare (e.g., recycling).

Second, it can change the options available to the agent (i.e., her choice set), for example through the creation of enabling regimes like contracts or corporate law. For example, the law can affect the willingness of a company's owner-managers to engage in conduct that risks harm to others by introducing forms of organization that insulate owners from liability for tortious harms resulting from their activities.³³ These laws expand the options available to actors. In so doing, they also alter the environment, removing the threat of liability that otherwise would exist. Other bodies of law also have dramatic effects on the nature and structure of the agent's choices. Election law, for example, determines both the electorate and the balloting method that together determine the agent's options, though sometimes indirectly. The drawing of district lines, in single-member districts, determines the electorate in each district, which in turn will affect the set of candidates who declare for a seat. Similarly, a rule that requires voting for a "ticket" (say of president and vice president) rather than for each individual office structures the agent's choices differently.

The first two interventions in effect directly change attributes of the options presented to the agent. Under the third intervention, the law changes the agent's

33 For a discussion of how laws insulating companies from liability for their agents' torts reduces care-taking, see, for example, Lewis Kornhauser, *An Economic Analysis of the Choice Between Enterprise and Personal Liability for Accidents*, 70 CAL. L. REV. 1345 (1982); Henry Hansmann & Reinier Kraakman, *Toward Unlimited Shareholder Liability for Corporate Torts*, 100 YALE L.J. 1879 (1991); Jennifer Arlen & W. Bentley MacLeod, *Malpractice Liability for Physicians and Managed Care Organization*, 78 N.Y.U. L. Rev. 1929 (2003); Jennifer Arlen & W. Bentley MacLeod, *Beyond Master-Service: A Critique of Vicarious Liability*, in *EXPLORING TORT LAW* 111 (M. Stuart Madden ed., 2005).

beliefs about the attributes or consequences of the options available to her, without affecting the actual attributes of the options available.³⁴ When agents are not fully informed, the law can alter beliefs by providing information to the agent about the consequences of each option. The law can do so directly through signals provided about consequences by the law's enactment. For example, adoption of criminal prohibition on certain conduct can signal that the legislature concluded that the conduct imposes substantial and unacceptable harm on third parties or the actor. The law also can alter beliefs indirectly, by causing third parties to produce information. For example, adoption of a law requiring producers to disclose product risks can alter agents' expectations of the consequences of purchasing certain products. Interventions along any of these dimensions might alter the choices people make without altering their underlying preferences. They thus are all consistent with rational choice theory.

Rational choice models contain two additional elements that affect people's choices, however: the decision-rule and preferences. In contrast with the first three elements of the model, which are treated as potential targets of legal intervention — potential mechanisms for altering behavior — rational choice theory assumes that these two are not. Specifically, in rational choice models the choice of decision-rule or equilibrium concept (e.g., subgame perfect equilibrium versus trembling hand) is determined exogenously by the model's creator, as opposed to being derived endogenously. It thus is a construct that is not potentially subject to manipulation by the law. In addition, and more importantly for our purposes, rational choice models assume that agents' preferences — and their constitutive criteria of evaluation — are exogenously determined and stable over time. Preferences are thus not subject to manipulation by the law.³⁵

Consequently, legal changes to preferences, if possible, would represent a mechanism for altering behavior that rational choice theory typically eschews. Indeed, it would present a challenge to a foundational assumption of rational choice theory. It would also present a challenge for economic analyses of law predicated on rational choice, which are premised on the view that laws should maximize social welfare, as determined by people's preferences. Preference

34 A belief is a mental state that represents the state of the physical or social world. In models of choice under uncertainty, the agent's preferences over actions can be represented by a preference over outcomes and a set of beliefs about the likelihood that each event may occur. These degrees of belief express the agent's assessment of how likely each event is. In the sequel, we treat these probabilities as mental states.

35 See, e.g., Samuel Bowles, *Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions*, 36 J. ECON. LITERATURE 75 (1998).

maximization cannot necessarily be used as a metric to select between legal rules if the legal rules alter the basis of the preferences themselves³⁶ — for example, by altering their underlying criteria of evaluation.³⁷

To illustrate the differences and connections among these elements consider a simple example. Suppose a jurisdiction lowers the speed limit in school zones from 20 to 15 miles per hour, imposing a fine on convicted violators. This new law could alter people's behavior — their decision to drive over 15 mph — in a variety of ways. For example, the threatened sanctions change the environment by raising the cost of driving over 15 mph. The law also could alter people's beliefs about the riskiness of driving over 15 mph in a school zone. This change in the environment and beliefs could induce an agent to reduce her speed in the school zone. This change in behavior would occur even though her preferences (her criteria of evaluation) remain unchanged.

C. Classical and General Deterrence Theory

The claim that the law shapes preferences is often presented as a challenge to deterrence theory. To evaluate this claim it is important to distinguish between different understandings of deterrence theory — specifically classical deterrence and what we shall call general deterrence — and also preference-shaping.

Classical deterrence theory is a narrow subset of deterrence theory that derives from a seminal paper by Gary Becker.³⁸ In that paper, as in the subsequent literature, the agent is assumed to have narrowly self-interested preferences; she does not care about others or about norm compliance. These models also assume that criminal laws only affect agents' welfare through one channel: the imposition of government-imposed sanctions. Thus, in these models, the cost to an agent of engaging in misconduct depends entirely on the level of sanctions imposed by the government and the probability of enforcement. Civil and criminal law thus influence behavior by altering the individual's incentives to select particular options solely by altering the consequences of those choices in a particular way — through the threat that certain choices will

36 *E.g.*, Dau-Schmidt, *supra* note 6, at 16 (1990) (the normative validity of Pareto optimality and social welfare are undermined if laws alter preferences); Sunstein, *supra* note 6, at 1145.

37 *See supra* note 20 and *infra* note 44 (discussing the possibility that the domain of preference and the domain of choice are not identical).

38 Becker, *supra* note 3.

lead to government-imposed sanctions (e.g., financial sanctions or imprisonment) or liability to private parties.³⁹

All claims that expressive law alters preferences represent a challenge to classical deterrence theory. The most general understanding of these claims is that the law can alter behavior by altering the social meaning of the prohibited conduct, deterring by leveraging people's other-regarding or prosocial preferences. The stronger version of these claims is that the law actually changes what people want in a fundamental way. Both of these claims are inconsistent with classical deterrence theory, which assumes that people do not care about others' welfare or ethics, and have stable preferences.

Yet classical deterrence theory only explores a subset of the ways that the law can deter under rational choice theory. General deterrence theory permits consideration of the manifold ways in which the law can cause people to refrain from illegal conduct through the channels recognized by rational choice theory.⁴⁰ Under rational choice theory people can have other-regarding and social preferences. Thus, law can deter, consistent with deterrence theory, by leveraging people's concern for others or their place in society. Expressive law theories thus are potentially consistent with both general deterrence theory and rational choice theory, as we will show. Only the strong claim — that the law shapes preferences in the sense of altering what people fundamentally value — is inconsistent with deterrence theory in general, as well as rational choice theory, since both rest on the foundational assumption that preferences are stable.

Accordingly, the claim that the law shapes preferences only represents a challenge to both rational choice theory and general deterrence theory, as some have claimed,⁴¹ if the strong claim is correct that the law changes what people value. To evaluate this claim, in the next Part we explicate what it would mean for the law to change preferences in this way. Then in Part III, we identify the three channels through which the preference-shaping literature

39 Becker, *supra* note 3; e.g., A. Mitchell Polinsky & Steven Shavell, *The Optimal Use of Fines and Imprisonment*, 24 J. PUB. ECON. 89 (1984). Scholars mining this vein also recognize that the law may alter behavior by changing the agent's set of options. For example, the law could reduce the choice set by setting an "infinite" price (e.g., immediate death) on an option. Alternatively, it could expand their options by transforming a non-waivable right into a waivable one or an inalienable right into an alienable one. Recognizing this important role for law does not alter the central conclusions of our analysis about preference-shaping claims.

40 See *supra* Section I.B.

41 See *supra* note 9.

claims that the law shapes preferences and show that none of them involves changing people's preferences.

II. AN EXPANDED VIEW OF THE STRUCTURE OF RATIONAL CHOICE MODELS

In this Part, we explicate the concept of preferences within the formal structure of rational choice models in order to enable us to elucidate the distinction between legal interventions that change preferences, in the sense of changing what people want, and interventions that alter choices by changing people's environment, beliefs, or options, without altering underlying preferences.⁴² This formal structure is both abstract and very simple. Yet it reveals the suppleness of rational choice models and enables us to distinguish between interventions that change preferences and those that alter choices by changing people's perception of the attributes of (the attractiveness of) a particular choice, while leaving their preferences unchanged.

A. Formal Model of Preferences

Formally, a *preference* is simply a *linear order* R over some *domain* D .⁴³ The domain D is simply the set of elements "ordered" by the preference; in the standard interpretation, the elements constitute what the agent cares about. Using the terminology developed in Part I, the domain D is the set of options available to the agent.⁴⁴ The domain of preference thus varies with the problem

42 We study "preference" rather than "utility" as we think "preference" is the more basic concept. An important element of preference theory consists of representation theorems that identify the conditions under which a preference is representable by a utility function (or by expected utility). See ANDREU MAS-COLELL ET AL., MICROECONOMIC THEORY, ch. 3C (1995).

43 Technically we should think of a preference as a pair (D, R) . This pairing plays an important role in our discussion below.

44 In the standard model, the analyst assumes that the agent's domain of preference and her domain of choice are identical. Yet the domain D of objects over which the agent has preferences can and often does deviate from the domain C of actions from which the agent chooses. For example, consumers often care about consumption bundles evaluated as a unit. The set of bundles is the domain of preference, D . For example, a person seeking to lose weight may evaluate food consumption over a day or a week, instead of over a single meal. Consumers, however, often construct consumption bundles commodity by commodity over time. Each consumer constructs her own consumption bundle by sequential choice

studied. In consumer theory, for example, the domain of preferences is the “consumption bundle.” In a theory of legislative elections, by contrast, the domain of preferences might be legislative programs. If we return to Liza from Part I, the domain D would be the set of all possible meals available from the restaurant’s menu.

A linear order R is a binary relation between any pair of options, x and y in D , such that the ordering is complete (i.e., is defined for every pair of elements in D), antisymmetric (if xRy and yRx then $x=y$), and transitive (i.e., if xRy and yRz then xRz).⁴⁵ Hausman includes a fourth condition in classical

of its elements or components. Each element of the consumption bundle thus constitutes a distinct domain of choice. For example, a person ordering lunch at a restaurant cannot simultaneously select her lunch and the rest of her meals for the week. Similarly, citizens do not vote on legislative programs directly; they elect members to a legislature that will enact a legislative program. The domain of choice for each voter thus consists of the set of candidates seeking election. When the domain of preference and the domain of choice differ, the analysis of preferences can become complex. If the agent has well-defined preferences over her domain of choice, maximization of her derived preferences entails maximizing her fundamental preferences as well. By contrast, if her fundamental preferences are not separable, no derived preference will exist, so she must act to maximize her fundamental preferences. In addition, some changes to her choice set or to her understanding of the attributes of objects within her domain of choice may entail the lack of a well-defined derived preference over her (new) choice set. (Of course, either of these changes might lead to well-defined preferences over the choice set when previously there were none). Benoît & Kornhauser, *supra* note 20, illustrate the complexity of the modeling exercise. Some models assume that the voters rank candidates on the basis of how closely the candidate’s legislative program is to the voter’s. This ranking is indeed a preference but not a preference that derives from the voter’s ranking of legislative programs. Such a derived preference only exists if the preferences over legislative programs are separable in candidates, which is unusual in the space of all possible preferences over legislative programs. On this point, see Jean-Pierre Benoît & Lewis Kornhauser, *On the Separability of Assembly Preferences*, 16 *SOC. CHOICE & WELFARE* 429 (1999). Notice that, when no derived preferences over the domain of choice exist, the agent must rely on a more sophisticated conception of rationality (than choose the maximal, feasible element in her choice set) if her fundamental preferences are to guide her behavior. Consequently, her behavior may superficially appear to violate standard assumptions on some non-derivative ranking of her domain of choice.

45 This definition includes many things that are not preferences in the economic interpretation at all. Let R be the relation “higher than” and D the set of mountains on earth. Then, formally, the pair (R, D) is a preference.

preference theory: context independence;⁴⁶ but we suggest below that context dependence can emerge endogenously from the richer formal structure we outline below.

An individual constructs her preferences from a set *E* of evaluative criteria.⁴⁷ Each criterion ranks the objects in the domain of preferences. The fundamental preference represents the integration of the rankings of all evaluative criteria into an all-criteria-considered ranking.⁴⁸ For example, an individual's evaluative criteria could include life expectancy, vitality, the financial welfare of progeny, personal comfort, self-perception as being a moral person, and standing with others.

46 See HAUSMAN, *supra* note 22.

47 We thus agree with HAUSMAN, *supra* note 22, that the best interpretation of a preference ranking is as an all-things-considered ranking of the elements in the domain of preference.

48 A loose description of this process of integration might say that the agent places a "weight" on each of these criteria to construct the fundamental preference. But this description suggests that the agent's all-things-considered preference is a weighted average of the individual criteria, which implies separability. Separability is a property of only a small subset of possible rankings. Economists often speak of "utility functions" and "utility" rather than "preferences." In that alternative terminology, each criterion of evaluation may (but need not) refer to an argument in the utility function. In this context, saying that the agent places a "weight" on each argument suggests that the all-things-considered utility function is simply a weighted sum of the arguments in the function. But many common utility functions do not have this form. Consider, for example, state-dependent utility functions often used in situations where the decision-maker could remain healthy or instead suffer a life-changing permanent injury, such as quadriplegia. Her all-things-considered utility is not a weighted sum of her wealth and her health status. Typically, the analyst assumes that for every wealth *w*, the agent's "utility" is lower when she is in the quadriplegic state than it is in the fully healthy state. See, e.g., Arlen, *supra* note 15. Finally note that theorists have investigated the conditions under which an agent's preference can be represented by a continuous utility function. Not all preferences can be so represented.

For an example in which the arguments in the utility function do not correspond to criteria of evaluation, consider our example of Liza choosing a meal. One might define a utility function with arguments of appetizer, main course, and dessert; but these would not be her criteria of evaluation. Her criteria of evaluation would likely include more fundamental attributes such as taste, cost, calories, cholesterol, vitamin content, and, perhaps, impact on the environment.

Preference, of course, is only one element of a model. A fully specified model identifies (for each agent) both her choice (or strategy) set,⁴⁹ and the perceived attributes of each option in the choice set, which, as explicated, will depend on the agent's criteria of evaluation, her beliefs, and the environment in which she acts.⁵⁰

In order to rank the options in her domain of preference, the agent must identify and evaluate each option with respect to the attributes of that option that are relevant to her criteria of evaluation. Thus, two people with fundamentally different criteria of evaluation (e.g., financial reward versus beneficial impact on the environment) will typically assess different attributes of the same option.⁵¹

The attributes of each option depend on the physical and social nature of the option, and the environment of the choice. An agent generally has beliefs about the attributes of the object of choice that are subject to revision. Changes in beliefs about the attributes of course do not change the actual attributes or the object of choice, but they may alter her choice by changing her assessment of the attributes of the object of choice.

The attributes also depend on the environment. The relevant environment is all facets of the world (outside the agent) — including the law — that can affect the agent's assessment of the available options given the criteria of evaluation from which she constructs her preferences. The environment can affect the attributes directly by altering the consequences to an agent of a particular option or set of options, as when a law is promulgated that renders one option illegal. Or it can affect the agent's beliefs about the attributes of an option without actually affecting the actual characteristics of the option. The solution to the model identifies the “best” action of each agent (or what “rationality” requires) given her preferences, the options available to her, the environment, and her beliefs.⁵²

49 See *supra* notes 29-32 and accompanying text.

50 See Section I.B. (identifying the factors that affect agents' choices in rational choice models).

51 Kelvin Lancaster developed this insight in the context of consumer theory. Kelvin J. Lancaster, *A New Approach to Consumer Behavior*, 74 J. POL. ECON. 132 (1966); KELVIN J. LANCASTER, *CONSUMER DEMAND: A NEW APPROACH* (1971). An individual, for example, chooses among breakfast cereals on the basis of various attributes: nutritional content, taste, texture, and price. Each of these attributes, of course, differs from the breakfast cereal itself.

52 See Section I.B. (identifying the factors that affect agents' choices in rational choice models). In a single-person decision problem, the agent's “best” action identifies the action that provides an outcome in the domain of preference that is maximal among the outcomes that can feasibly be achieved. In strategic

This framework thus elucidates two distinctions that are important to the analysis of the preference-shaping role of law. The first is the distinction between preferences⁵³ and the evaluative criteria that underlie or are employed to construct them. The second distinguishes between the objects of choice and the attributes of these objects of choice. The first distinction provides a deeper account of the agent's preferences and allows the agent to respond to a legal rule through reasons other than the expected sanction. The second distinction shows how law can alter behavior, without changing preferences, whenever it changes an agent's assessment or valuation of the attributes of an option in her domain of choice. The law, for example, might change either the environment of choice in ways that alter an option's costs or benefits, an agent's beliefs about the consequences of the conduct, or the options available.

An elaboration of the school zone example may clarify these distinctions and their significance. Consider two different traffic control regimes in a city. In one regime U, traffic is unregulated. In regime S, traffic outside school zones is unregulated, but, within school zones, a limit of 20 miles per hour is imposed. An agent, Liza, must travel through the city on a specified route that passes through a school zone. Her domain of choice is her speed at each point on her route. Liza has preferences over this domain of choice that derive from her criteria of evaluation as applied to this domain of choice.⁵⁴ She may have various criteria of evaluation, including her total travel time, the cost of travel, her own safety, the physical integrity of her car, and her risk of harming others. Her speed implicates each of these. She also may care about complying with traffic regulations.

Now consider Liza's decision in Regime U, which does not regulate speed in the school zone. She ranks each option by applying each of her criteria of evaluation to it and then integrates these criterial rankings into her all-things-considered preference. She then selects the choice that she ranks highest. Assume that she selects 30 mph in a school zone based, in part, on her belief that it does not present a material risk of harm to others.

Suppose that now the jurisdiction institutes Regime S. The new legal regime alters her assessment of the attributes of all driving speeds over 20 mph in a school zone. Most obviously, driving over 20 mph now entails higher costs

situations, each solution concept identifies a conception of rational action that characterizes the agent's "best" outcome.

53 *See supra* note 44 (discussing the concepts of domain of preferences versus the domain of choice).

54 As we shall discuss in the next Part, the direct application of the criteria of evaluation to the options in the domain of choice occurs because her domain of choice is also her domain of preference.

as a result of the potential fine. In addition, the law increases Liza's expected costs of driving over 20 mph if it leads her to conclude either that such conduct presents a higher risk of harming children than previously believed, or will cause her to suffer social opprobrium should she be observed driving over 20 mph in a school zone.

The change from regime U to regime S may thus have altered Liza's choice/behavior by changing the perceived net benefit to her of driving over 20 mph, even though her underlying criteria of evaluation (including her preference for speed) and her method of integrating them are unchanged. Only the (perceived) attributes of her options changed and these changes occasioned a change in her ranking. We discuss these underlying mechanisms further in subsection C.

B. Attributing Preferences

In the prior discussion, the agent's domain of choice and criteria of evaluation were exogenously determined. Yet economists developing models must instead attribute preferences to each agent modeled. Economists do so by identifying the criteria of evaluation likely to be particularly relevant to people in the domain of choice of interest to the researcher. For example, economic models of tort law focus on agents' preferences over (health, wealth) pairs.⁵⁵ In models of divorce bargaining, parents are assumed to have preferences over (time-with-the-child, wealth) pairs.⁵⁶ Models of adjudication typically attribute judges' preferences over dispositions, policies, or both.⁵⁷ In each instance, the model attributes preferences to agents over a domain that corresponds to things that the agent plausibly cares about in the phenomenon under study. Each dimension corresponds to a criterion of evaluation that fits the situation modeled.

Researchers face challenges attributing preferences when the domain of choice of interest to the researcher does not naturally correspond with either the ultimate domain of choice of primary concern to the agent or her criteria of evaluation. Consider the analysis of legislative elections. Researchers studying voters' choices among candidates might be tempted to assume that

55 See, e.g., Arlen, *supra* note 15. The standard model considered only pecuniary loss.

56 See Robert Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950 (1979).

57 See Charles Cameron & Lewis Kornhauser, *Chapter 3: What Do Judges Want? How to Model Judicial Preferences*, NYU Law and Economics Research Paper (June 2, 2017), <https://ssrn.com/abstract=2979419>.

the voter cares about, and has preferences, over candidates. But voters often do not care primarily about candidates but instead care about the policies that the legislature will enact. To the extent that this is the case, then the objects of choice are the legislative programs that the voters care about and the relevant preferences are the voters' preferences over legislative programs. Ideally, we could derive the voter's preferences over candidates directly from her preferences over legislative programs. However, the connection to the voter's underlying concerns about policy often will not be transparent.⁵⁸

When, as above, the agent's domain of preference differs from her domain of choice, we use the term "fundamental preference" to refer to the preference over the domain of preference, and "derived preference," if it exists, to refer to derived preference over the domain of choice.⁵⁹ This terminology does not mean that the criteria of evaluation attributed to her fundamental preferences are fundamental to the agent's personality or being. Consider the election example. We could try to determine the criteria against which the voter assesses each legislative program. She may care, for instance, about how the legislative program affects her income, the income of others, the environment, and many other things. If we started with these more primordial criteria and the voter's integration of them into an all-things-considered preference, we might find that the agent's preference over legislative programs derives from some more basic preference over life plans or something else. This tactic would thus likely prove analytically intractable.

The essential insight for our analysis is that changing preferences requires either changing the criteria of evaluation or the manner of their integration. While theorists have considerable leeway to specify what those criteria should be, theorists should attend to whether the criteria specified are plausibly the criterion of evaluation that people access when assessing, evaluating, and ranking the choice options before them.

C. What Would It Mean for the Law to Shape Preferences?

We now can ask what it would mean for an intervention (such as a change in the law) to change a person's *preferences*. And does a law that changes the choices that people make, through a channel other than sanctions and enforcement, necessarily entail a change in preferences? The framework set out

58 In many situations, a voter's preferences for candidates cannot be derived from her preferences over policies. Indeed, in most instances, no well-defined preference over candidates derives from the voter's preferences over legislative programs. For a fuller discussion, see Benoît & Kornhauser, *supra* note 20.

59 We can do this only when derived preferences over the domain of choice exist.

above provides a simple answer to the first question: an intervention changes preferences if, but only if, it alters either the agent's criteria of evaluation or the way in which her criteria of evaluation are integrated into the agent's all-criteria-considered preference ranking.⁶⁰ The law cannot be said to alter preferences simply because it alters people's choices.

All of the claims that the law shapes preferences are predicated on the view that the law can alter people's willingness to engage in conduct that the law seeks to discourage through its ability to express that the conduct is socially undesirable — that it harms others, violates social norms or is unethical. In the colloquial sense of the word, these interventions can be said to alter people's "preference" or demand for undertaking the harmful action. Yet the framework outlined above highlights that preference, in the colloquial sense of preference for a specific choice, is not the same thing as preference as in a technical, economic sense. It is important to distinguish between the agent's choice (her preference for a specific action) and the criteria of evaluation that produce the ranking (or preference) that governs her decisions. Only changes to her criteria of evaluation or the way they are integrated genuinely alter an agent's values and represent the challenge to both rational choice theory and general deterrence theory embraced by proponents of the strong claim for preference-shaping. In turn, only a change in fundamental preferences potentially raises the concern of legislatures undermining individual autonomy by altering people's fundamental wants and desires.⁶¹

Thus, in order to evaluate the claims that the law shapes preferences by changing what people value and who they are, we must assess whether any of these theories of how the law shapes preferences entail altering either agents' set of criteria of evaluation or the way they are integrated into a person's all-things-considered preferences. For example, a legal intervention could properly be said to alter people's preferences if it led someone who previously placed no value on the welfare of others to treat the welfare of others as a positive criterion of evaluation.⁶² The law also could be said to change preferences if it caused an individual who previously ranked financial returns over the safety of others to now rank the latter more highly than the former.

60 This formulation is loose because it presupposes that the agent's fundamental preferences are separable. They may not be; but they may change nonetheless because of the way in which the criteria are integrated.

61 See text accompanying notes 6-8; but see Porat, *supra* note 6 (challenging the claim that it is normatively problematic for the state to interfere with fundamental preferences).

62 Similarly, an antidiscrimination law could be said to alter preferences if the adoption of the law caused people who previously disliked people of other races to no longer include race as a negative attribute in their criterion of evaluation.

By contrast, a legal intervention does not alter preferences if it simply alters an agent's willingness to select a particular option through the channels identified by rational choice theory, without affecting her criteria of evaluation or how they are integrated. The law could do this by changing the environment to increase the cost of a particular choice. The law can do so, consistent with classical deterrence theory, through the threat of legal sanctions. But it also can do so, as discussed below, by changing the social meaning of the prohibited conduct, thereby causing people to anticipate negative social repercussions should they undertake the conduct. Alternatively, the law could alter people's beliefs, causing them to view the conduct as more harmful. All of these channels involve the law altering a person's decision to undertake prohibited conduct without altering her preferences.

III. DOES THE CRIMINAL OR CIVIL LAW CHANGE FUNDAMENTAL PREFERENCES?

The prior Part identified what must be established in order to conclude that the law changes preferences. In this Part, we evaluate the claims made by the preference-shaping literature about how the law changes preferences to determine whether these entail changing people's criteria of evaluation (or the way they are integrated) or instead only affect people's perceptions of the attributes of the options the law prohibits.

We focus on three core claims about the pathways that the preference-shaping literature has identified as the channels through which the expressive features of the criminal or civil law affect behavior. First, promulgation or enforcement of a law can alter people's beliefs about the harm caused by the conduct to others or to the actor (the social harm mechanism).⁶³ Second, the law can establish or enhance the salience of a social or moral norm against the conduct, thereby potentially providing an internal motivation for people to eschew undesired conduct (the social norm mechanism).⁶⁴ Third, the

63 *E.g.*, Kahan, *supra* note 8; Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943 (1995) [hereinafter Lessig, *Regulation of Social Meaning*]; Lawrence Lessig, *Social Meaning and Social Norms*, 144 U. PA. L. REV. 2181 (1996) [hereinafter Lessig, *Social Meaning and Social Norms*].

64 Cass R. Sunstein, *On the Expressive Function of Law*, 144 U. PA. L. REV. 2021 (1996); Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903 (1996); *see* Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 MICH. L. REV. 338 (1997) [hereinafter McAdams, *Origin, Development and Regulation of Norms*]; Richard H. McAdams, *A Focal Point Theory of Expressive Law*, 86 VA. L. REV. 1649 (2001). Thus, in this view

law could cause people who are not socially motivated to become better versions of themselves, engaging in what Robert Cooter calls “Pareto self-improvement” to favor law-abiding behavior in order to obtain the greater opportunities available to law-abiding and norm-conforming citizens (the self-improvement mechanism).⁶⁵

We show that none of these channels involve any structural change in people’s preferences. Specifically, none involve changes to people’s criteria of evaluation or their ranking. Instead, through each channel, criminal and civil law cause people to change their assessment of the attributes associated with a now illegal choice by expressing that the prohibited conduct is either harmful to others or contrary to a social norm. This can lead people to conclude that the prohibited conduct has previously unassessed negative attributes to the extent that their criteria of evaluation include concern for the welfare of others, for conforming to social norms, for maintaining a self-identity as being a moral person, or for reputation with others. Law thereby changes their choices without altering their underlying preferences

A. Altering Behavior and the Social Harm Mechanism

Civil and criminal law can alter people’s choices by influencing their beliefs about the attributes of one or more actions⁶⁶ available to them. When citizens

deterrence through people’s inclination to comply with the law depends more on people’s general ethical and social preferences than on direct legal measures, such as enforcement. *See Lynch, supra* note 9.

65 Cooter, *Good Citizens*, *supra* note 6; *see* Cooter, *Models of Morality*, *supra* note 6; Cooter, *Expressive Law*, *supra* note 6; Robert Cooter, *Three Effects of Social Norms on Law: Expression, Deterrence, and Internalization*, 9 OR. L. REV. 1 (2000) [hereinafter Cooter, *Three Effects*]. Thus, Cooter claims that the law does not simply induce socially inclined people to alter their behavior by establishing a norm against certain harmful conduct, but it can motivate other less socially inclined people to engage in “Pareto self-improvement” to change their preferences to achieve greater social welfare. Cooter, *Good Citizens*, *supra* note 6, at 1581.

Similar arguments are made by Lessig. Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661 (1998) [hereinafter Lessig, *New Chicago*]; Lessig, *The Regulation of Social Meaning*, *supra* note 63; Lessig, *Social Meaning and Social Norms*, *supra* note 63.

66 In the context of criminal law, the domain of choice is the set of actions available to the agent. We can thus abandon the abstract description “object of choice” for the less abstract term “action.” *See supra* notes 20 and 44 (discussing how the domain of preference can deviate from the domain of choice).

trust the legislature to enact laws in the social interest, for example, a statute that subjects previously unregulated conduct to a criminal sanction may cause people to conclude that expected harms to themselves or others associated with the prohibited actions are greater than they otherwise would have expected.⁶⁷ Traffic regulations that set speed limits may alter people's beliefs about the set of speeds it is safe to drive in that location.⁶⁸ Environmental laws that prohibit depositing car batteries in household trash or pouring car oil down storm drains signal that these actions cause more harm to the environment than people might otherwise have thought. Laws prohibiting bribery of foreign officials or sexual harassment may lead people to conclude that such activities are more harmful than they would have otherwise believed, given their prevalence prior to the prohibition.⁶⁹

Laws that cause people to conclude that prohibited conduct is especially harmful to others can alter people's decisions to engage in the disfavored conduct. Yet they do so by altering people's beliefs about the attributes or character of the illegal action. After the law's enactment, when people assess the action it has an additional negative attribute: being harmful to others. This belief can lead people whose criteria of evaluation includes concern for others' welfare to eschew the illegal conduct. This change in behavior, however, occurs without the law changing people's underlying preferences. Laws that alter beliefs about harmfulness do not affect the internal, structural processes that determine whether people's criteria of evaluation include concern for the safety of others (or the weight given to that concern). Thus, they do not alter preferences themselves. Accordingly, laws that change beliefs about harmfulness can alter the behavior of those whose preferences already placed weight on avoidance of harm to others. By contrast, a person whose criteria of evaluation do not include a concern for others will not alter her choices in

67 Bilz & Nadler, *supra* note 6 (discussing how lack of trust of government authorities can alter the impact of laws).

68 On a mountain road, a traffic sign with an arrow bent at right angles and the numeral 10 under it signals that continuing around the upcoming curve at speeds in excess of 10 mph is unsafe. Typically traffic regulations also change the environment in which the agent acts by imposing a sanction on speeds in excess of the posted limit.

69 In addition, the law can influence people's beliefs about what other people will do, thereby facilitating coordination or cooperation. *See, e.g.*, Cooter, *Good Citizens*, *supra* note 6, at 1586-88 (smoking bans in public facilities or workplaces and pooper-scooper laws show how the legal rule almost instantly shifts society from one equilibrium to another). This is likely to alter behavior to the extent the new equilibrium is viewed as a social or moral norm, and thus is discussed in Section III.C.

response to the effect of the law on her beliefs about the law's harmfulness; information about harmfulness to others is not relevant to her decision.⁷⁰

To be more precise, to change fundamental preferences, a law would need to cause people to alter either their criteria of evaluation or how their existing criteria are integrated into their all-things-considered preference. In our example, the law would need to cause someone who never cared about other people's welfares to include the welfare of others as a criterion of evaluation that affects their choices. Yet the claim that the law alters beliefs about the social harm caused by certain actions does not involve a claim that the law causes people who are indifferent to the welfare of others to care about others. Rather, it asserts that the law can alter an individual's beliefs about the social harm an action may cause, leading them to view the action more negatively.

To illustrate why this distinction between changing beliefs about the attributes of a specific choice and changing fundamental preferences matters, consider how a law that causes people to believe that a given act, say corruption, harms people in society would affect the behavior of two different potential wrongdoers. Consider Hui who is considering whether to engage in bribery that would enable her to obtain a contract that would enhance her job prospects. Hui has preferences that integrate both narrowly egocentric criteria of evaluation (including lifetime consumption, career opportunities, and wealth), and more socially-regarding criteria of evaluation, including moral virtue, not harming others, and good citizenship. Thus, Hui's choices will be predicated, in part, on concern for the welfare of others.⁷¹ In order to focus solely on the expressive role of law, assume that Hui believes that she will not get caught if she bribes. Although Hui is not concerned about being sanctioned, the law's expressive features may nonetheless dissuade her from enhancing her career through bribery if (1) the law convinces Hui that bribery is socially harmful and (2) she places sufficient weight on avoidance of social harm that her "optimal" choice is to eschew bribery. She now chooses not to bribe because she now sees

70 This Part considers how a law that changes beliefs about the risk of harm can alter behavior directly by interacting with people's criteria of evaluation not to harm others. Information about the harm produced by conduct also could affect behavior if social or moral norms disfavor conduct that harms others. This possible channel is considered in the next Part which discusses social and moral norms.

71 As we saw, rational decision-making when the domain of preference differs from the domain of choice can be very complex, as often the agent's preferences will not be separable in the "unit" of choice. In the face of non-separability, the agent must have remarkable foresight and undertake difficult calculations to determine an action. But her fundamental preferences still guide her deliberations.

bribery as socially harmful and her preexisting preferences (which include a concern for others) dictate that she eschew the bribe.

By contrast, consider Janice, who has preferences that derive from criteria of evaluation that conform to the classical understanding of *homo economicus*: she values only her own material well-being and her career opportunities, and places no value at all on either the welfare of others or being a moral person. Assume that she is faced with the same choice as Hui following enactment of the same law. The law may alter her beliefs about whether her actions could harm others, but these new beliefs will not alter her assessment of the decision to bribe because her criteria of evaluation do not include harm to others. Thus, the law's effect on her beliefs will not deter her from using bribery to gain career opportunities.

Thus, we see that laws that express society's belief that conduct is socially harmful can alter a decision to engage in socially harmful conduct, but not by changing people's underlying preferences. The law's ability to change beliefs only affects those people whose preexisting preferences include, among their criteria of evaluation, the welfare of others or a concern for behaving ethically. Merely leading people to believe that others will be harmed does not suffice to transform people who do not care about others into those who do.

B. Changing Behavior through the Social Norm Mechanism

Laws also can alter behavior by changing the social meaning given to certain actions.⁷² The constitution of social meaning determines how agents understand or interpret an action. Legal regulation may take many forms. Each form expresses a different judgment about the ethical nature or social value of the regulated conduct. Laws that impose a tax on particular conduct can express that the conduct is harmful, but do not in and of themselves express society's condemnation of the conduct. By contrast, civil or criminal laws that prohibit and sanction conduct express society's view that a good citizen should not engage in the conduct at all. Those who do would be violating a social norm.⁷³ Criminal law (particularly common-law crimes), which often derive from or comport with basic intuitions or concepts of morality and ethics, can express that it would be unethical or immoral to engage in the conduct.

⁷² See Lessig, *Regulation of Social Meaning*, *supra* note 63.

⁷³ For a more detailed discussion of legal form and normativity, see Kornhauser, *supra* note 5, at § 4.

The law's expression that prohibited actions are immoral may influence behavior because people often have strong moral impulses.⁷⁴ People with an aversion to being immoral, and who value their own good opinion of themselves, may thus be influenced not to undertake the illegal action.⁷⁵ Alternatively, the law might be perceived as establishing or enhancing a social norm by expressing society's views that particular conduct is inconsistent with the social norms or expectations⁷⁶ — norms that are sufficiently material to enforce through criminal sanctions.⁷⁷ This expression could lead people to comply with the law who are averse to disappointing the expectations of others,⁷⁸ or who evaluate actions in significant part on their perceived fairness.

Thus, the expressive feature of the law can change behavior. Yet, as we show below, it does so without changing preferences. When the law changes the social meaning of an action — for example, by criminalizing previously unregulated conduct — it changes the attributes of the action. To the extent that preexisting criteria of evaluation include concern for either the morality of her actions or how others interpret her actions — i.e., to the extent to which such considerations are among the criteria of evaluation — the law may change the agent's ranking of actions and hence her behavior. But it will not change

74 Considerable evidence supports this proposition. *See, e.g.*, Diana C. Robertson et al., *Business Ethics: The Promise of Neuroscience*, 144 J. BUS. ETHICS 679 (2017). *See also* Donald Langevoort, *Behavioral Ethics, Behavioral Compliance*, in THE RESEARCH HANDBOOK ON CORPORATE CRIME AND FINANCIAL MISDEALING ch. 11 (Jennifer Arlen ed., 2018). *See generally* Donald Langevoort, *Culture of Compliance*, 54 AMER. CRIM. L. REV. 933, 946 (2017).

75 Paul H. Robinson & John M. Darley, *Utility of Desert*, 91 NW. U. L. REV. 453 (1997); Janice Nadler, *Flouting the Law*, 83 TX. L. REV. 1399 (2005). *See also* Bilz & Nadler, *supra* note 6, at 245; Porat, *supra* note 6, at 219.

76 *See, e.g.*, Cooter, *supra* note 5; Richard H. McAdams, *An Attitudinal Theory of Expressive Law*, 79 OR. L. REV. 339 (2000) [hereinafter McAdams, *Attitudinal Theory*]; Sunstein, *Social Norms*, *supra* note 64.

77 The seriousness of norm violation might be communicated not only through the labelling but also through the size of the sanction. In studies of social norms, this “pricing” effect may have paradoxical effects. The imposition of a fine for noncompliance with a norm, it is argued, may “crowd out” the agent's moral motivation; in the terms above, it reduces the cost of noncompliance or, phrased differently, it reduces the guilt or shame that arises from noncompliance. The choice of sanction for noncompliance might thus interact with the decision to label noncompliance as criminal. The choice of an “inappropriate” sanction might undermine rather than enhance the deterrent effect created by the labelling.

78 *See, e.g.*, McAdams, *Origin, Development and Regulation of Norms*, *supra* note 64.

preferences. A person with purely egocentric and materialistic preferences will not suddenly care about being moral or conforming to social norms.

To explain how the law alters behavior through this channel, we consider insights from psychology on how moral or social norms influence behavior.

1. How Expressed Moral or Social Norms Influence Behavior: Role of Guilt and Shame

Although some scholars assert that people have a “taste” for complying with social or ethical norms — without specifying its behavioral micro foundations — a rich and growing literature in experimental economics and psychology has helped to identify specific psychological channels through which these concerns influence behavior.

Psychology has identified two important motivations that lead people to alter their behavior in order to avoid violating either a moral or social norm.⁷⁹ Each relies on a distinct criterion of evaluation. The first is predicated on the value the agent places on her self-perception as moral, fair, or a good citizen. The second values her perception by others as moral, fair, or compliant with social norms.⁸⁰ Decisions contrary to each criterion are associated with a distinct negative emotion, respectively guilt and shame.⁸¹ Each emotion in turn may be understood as the psychological mechanism that implements this criterion of evaluation in the decision process. The aversion to guilt or

79 The precise psychological mechanisms through which the establishment or enhancement of an ethical or social norm leads people to be averse to the illegal action often is not well specified. Yet a dominant explanation involves the role of guilt and shame. See Ernst Fehr & Ivo Schurtenberger, *Normative Foundations of Human Cooperation*, 2 *NATURE HUM. BEHAV.* 458, 463 (2018). Guilt and shame are not the only psychological attitudes that play a role in norm-compliance. There is also esteem. On esteem, see McAdams, *Origin, Development and Regulation of Norms*, *supra* note 64, and McAdams, *Attitudinal Theory*, *supra* note 76.

80 Economists sometimes model each criterion as an argument in the agent’s utility function, a representation of her all-things-considered preference. See *supra* note 48.

81 Guilt is the negative emotion that people experience when they disappoint their own expectations of themselves: when they do not conform to the behavioral standards that comport with their own expectations, for example by acting unethically, unfairly, or contrary to a social norm. Shame is the negative emotion that people experience when they disappoint the expectations of others. People experience shame if they act unethically, immorally, or contrary to an established norm to the extent that they expect that others expected them to comply.

shame thus can motivate people to avoid taking an action that they anticipate will cause them to experience these emotions.⁸²

2. *Do Expressed Morals and Norms Shape Preferences?*

Statutes or common-law rules that succeed in either rendering salient or establishing that prohibited conduct is either immoral or contrary to a social norm thus can induce people to avoid that behavior if, at the moment of choice, they are aware of the implications of their choice, and anticipate that they will experience guilt or shame if they engage in the prohibited conduct.⁸³ This understanding of how law shapes behavior does not involve the alteration of an agent's preferences. Instead, it relies on the agent's preexisting preferences (that reflect her aversion to guilt or shame) to alter the agent's decision to engage in the legally prohibited conduct. In the simplest account, the law alters the ethical or social meaning of the prohibited action because it either causes people to believe that the conduct, previously considered acceptable, is immoral (or contrary to social expectations), or strengthens the ethical or social norm against conduct that previously was considered ethically or socially ambiguous.⁸⁴ This mechanism can alter the consequences for the agent of engaging in the prohibited conduct if the law causes the person to anticipate guilt or shame if they engage in the conduct as a result of having preexisting criteria of evaluation that value both behaving ethically and conforming to social norms.

Thus, laws that operate to change the ethical or social meaning attributed to specific actions do alter behavior, but they do so by altering the perceived attributes of the object of choice (the prohibited conduct) and not by changing agents' criteria of evaluation or the integration thereof. Indeed, the foundational claim of this literature is that the law can change the social meaning of a given action — causing people to perceive it to be moral or not. This is a

82 For example, the agent at the time of decision must recognize that her action is potentially guilt- or shame-inducing. See Jennifer Arlen & Lewis A. Kornhauser, *Battle for our Souls: How Organizations Undermine the Expressive Force of Law* (draft on file with the authors).

83 People do not always anticipate experiencing guilt or shame over causing conduct that is unethical, unfair, or violative of a social norm, as shown in Part IV. Of particular importance, they do not anticipate experiencing such emotions if they do not perceive themselves to be responsible for causing the violative conduct.

84 E.g., Cooter, *supra* note 6; Lessig, *The Regulation of Social Meaning*, *supra* note 63; Lessig, *Social Meaning and Social Norms*, *supra* note 63; McAdams, *Origin, Development, and Regulation of Norms*, *supra* note 64; McAdams, *A Focal Point Theory of Expressive Law*, *supra* note 64.

claim that the law is acting on people's *perceptions* about the action itself — rendering actions immoral or contrary to social norms that previously would not have been so perceived. This intervention to establish a norm does not, however, operate to change what people care about — to change their criteria of evaluation. By revealing that an action is immoral, the law does not also cause immoral and antisocial people to suddenly care about being ethical and other-regarding.

To illustrate, consider laws against bribery of foreign government officials. Until relatively recently, across the world bribery of officials was both commonplace and considered by many to be an acceptable business practice.⁸⁵ The widespread adoption across the globe of anti-bribery laws, coupled with international attention to enforcement, expressed an international ethical norm against bribery.⁸⁶ This prohibition could cause people to change their beliefs about the normative attributes of the act of bribery — viewing it as immoral or contrary to social expectations, instead of as an acceptable business practice.

Whether this change in the attributes ascribed to such payments reduces an individual's willingness to bribe will depend on her preferences prior to and after the law. Specifically, it depends on whether her criteria of evaluation place a value on either ethical behavior or conformity to social norms and, if so, the way in which these criteria are integrated to construct her all-things-considered preferences. By establishing that bribery is immoral or violative of a social norm, the law will cause people to alter their behavior only if they have preexisting preferences that value ethical conduct or conformity to social norms. The law's expression that an action is immoral or contrary to social norms will not alter the behavior of those people whose preexisting preferences do not treat morality and social norms as criteria that are material to their choices. People who care only about the financial consequences of their conduct — with no regard for morality or their place in society — will not refrain from bribery simply because the law has revealed that the conduct is immoral or contrary to social norms.⁸⁷ This is because the channel of

85 Indeed, bribery was so accepted that multinationals, such as Siemens, had safes with cash in local offices around the world in order to give their employees ready access to bribery payments. See KEVIN E. DAVIS, *BETWEEN IMPUNITY AND IMPERIALISM: THE REGULATION OF TRANSNATIONAL BRIBERY* (2019).

86 See OECD, *Convention on Combating Bribery of Foreign Public Officials in International Business Transactions*, Dec. 17, 1997, 37 I.L.M. 1.

87 To explicate our point, let's return to Hui and Janice. Assume that Janice's fundamental preferences are predicated on criteria of evaluation that do not include morality, ethics, the welfare of others, or any concern for being a good member of society. Hui's fundamental preferences, by contrast, do include a criterion of evaluation that values being ethical. Hui thus expects to be worse

influence at the foundation of the claim that the law can express moral or social norms relies on the law's ability to alter people's understandings of the social or moral meaning of bribery. It does not entail a claim that the law's expression of morality can, in effect, turn coal into gold, causing egocentric and immoral people to suddenly become moral. People's preferences — and their nature — remain unchanged.

Observe that we can reach a similar conclusion using the language of psychology. Criminal law, in this story, does not alter people's fundamental views about the importance of morality or conforming to society. Instead, criminal law affects behavior by causing people to anticipate that they will experience guilt or shame from engaging in specific conduct because it is now legally prohibited. This change in the personal and social meaning of the action will cause people to alter their conduct if, and to the extent that, their basic nature leads them to experience guilt or shame over ethical violations. Those who do not experience guilt or shame from immoral or antisocial conduct will not be induced to alter their behavior through this channel.

Laws that establish social norms can, of course, affect the behavior of people who do not care about morality or norm compliance. A purely egocentric agent, such as Janice, could be induced to comply if she expects violation of the law to reduce her opportunities — to cause her, in other words, to incur costs from reputational damage.⁸⁸ This pathway opens up the possibility that

off — and to suffer internal costs in the form of guilt — if she violates a law that establishes a moral norm, even if she does not expect to be caught. We now consider whether either of them would be willing to bribe a foreign official in order to obtain a lucrative, career-enhancing contract following enactment of a law prohibiting bribery. In order to focus on the law's expressive role, assume that both Janice and Hui reasonably believe that they will not get caught, and that the law establishes a norm that bribing public officials is immoral and unethical. This law will cause both Hui and Janice to view the conduct differently — to conclude that it is immoral. Their changed understanding of the conduct will cause Hui to change her behavior, but not Janice, because the law alters a characteristic (morality) that is material to Hui's decision-making but not to Janice's. Janice will continue to focus on whether bribery enables her to maximize her welfare as defined by her narrow financial self-interest.

Of course, the law nevertheless could cause Janice to adjust her behavior, even if she does not expect the government to sanction her, if she faces a material risk that her misconduct will be detected and that detection will cause her to suffer costs from reputational damage. *See infra* note 88.

88 A wrongdoer can suffer costs from reputational damage if discovery of their misconduct provides information to their counterparties that she presents an above average risk of harming their counterparties in the future, thereby leading their

expressive laws can affect people who do not directly care about morality or norms. Yet it does not do so by changing their preferences. Here again the law affects behavior by changing the perceived consequences to someone of engaging in prohibited conduct. She now perceives it as an action that could provide short-term returns but at a considerable long-term cost should others detect her conduct and place sufficient negative weight on her actions that they are no longer willing to deal with her on favorable terms.⁸⁹

C. The Self-Improvement Mechanism

In a series of influential articles, Robert Cooter builds on the claim that the law can alter the social meaning of an action to offer the most detailed and explicit account of how law may change “preferences.”⁹⁰ The argument has several steps. First, Cooter assumes that most people in society care about the ethics and morality of their counterparties. Thus, a given actor — let’s call her Liza — choosing among contracting partners would prefer a contracting partner Freddy to a contracting partner Henry if she had reason to believe that Freddy has a better internalized set of moral norms, such as honesty and integrity, than Henry. Thus, using the framework outlined in Part III, the set of contracting partners constitutes Liza’s domain of choice. Thus, all else equal, people who are perceived to have better internalized norms have larger opportunity sets.⁹¹

counterparties to be less willing to deal with them. *E.g.*, Jonathan M. Karpoff & John R. Lott, Jr., *The Reputational Penalty Firms Bear from Committing Criminal Fraud*, 36 J.L. & ECON. 757 (1993); see Cindy R. Alexander & Jennifer Arlen, *Does Conviction Matter? The Reputational and Collateral Effects of Corporate Crime*, in RESEARCH HANDBOOK ON CORPORATE CRIME AND FINANCIAL MISDEALING ch. 11 (Jennifer Arlen ed., 2018) (discussing when corporate violators are likely to incur costs from reputational damage); W. Bentley MacLeod, *Reputations, Relationships and Contract Enforcement*, 45 J. ECON. LIT. 595 (2007) (discussing the role of reputation in the contractual context).

- 89 As discussed in Section III.C, Robert Cooter relies on this reputational sanction as the foundation of his concept of Pareto self-improvement.
- 90 Cooter, *Good Citizens*, *supra* note 6. See also Cooter, *Models of Morality*, *supra* note 6; Cooter, *Expressive Law*, *supra* note 6; Cooter, *Three Effects*, *supra* note 65. These articles also note the effect of law on both incentives and beliefs. Cooter emphasizes the role of the law in equilibrium selection where the law affects individual beliefs about what actions others will adopt.
- 91 Liza may value ethical behavior because ethical behavior is one of her criteria of evaluation. Alternatively, she could care only about the expected value of her business relationships. In this case, she would care about her counterparties’

Cooter then considers an agent who did not previously share others' preference for ethical or norm-conforming behavior. Cooter argues that this agent may be induced by the law's expressive features to conform to the law, even when he otherwise would not be so inclined, if doing so would enable him to signal to counterparties such as Liza that he is an ethical person. Thus, if the law induces people to conclude that others will obey the law, and will treat it as a social norm, then even those who did not previously value norm compliance might comply in order to obtain the opportunities associated with being perceived as an ethical or norm-compliant person.⁹²

This argument shows how the law can change the behavior of people who do not themselves experience guilt or shame when violating a social or moral norm. Yet it does not establish that the law changes an agent's preferences. It establishes only that an agent whose criteria of evaluation place a value on material welfare understands the value of her reputation. She benefits when others believe she has integrity whether she has it or not.

Cooter takes the analysis an additional step, however. He argues that the law can cause self-interested people to not only comply with the law, but also to change their internal moral values in order to obtain the opportunities associated with being perceived as a law-compliant person.⁹³ People, he claims, can engage in Pareto self-improvement. This additional step appears to be a claim that the law can alter people's criteria of evaluation, causing them to care about norm-compliance when they previously did not.⁹⁴

Yet Cooter's explanation of why people would be motivated to change preferences, instead of simply changing their conduct, arises from a strong assumption in his model: that people cannot induce others to believe they are ethical, moral and norm-compliant unless in fact they are. In this unlikely situation, the only way for people to gain opportunities is to alter who they are.⁹⁵

In Cooter's analysis, however, each actor's future opportunities do not depend on his actual preferences; they depend rather on people's *perceptions*

legal violations when past misconduct serves as evidence that the counterparty is likely to harm her in the future. *See supra* note 88. In this case, Liza would likely avoid contracting with unlawful counterparties who offer opportunities similar to others. By contrast, if unethical counterparties offer her superior business opportunities, with benefits that outweigh the perceived risks, then Liza might well be willing to deal with less ethical counterparties, depending on the relative weights given to financial returns.

92 Cooter, *Good Citizens*, *supra* note 6, at 1590.

93 *Id.* at 1581.

94 *Id.* at 1593.

95 *Id.* at 1594 (assuming that preferences are translucent — that revealed preferences reveal actual preferences).

about whether others are ethical or moral. These expectations are based on people's actions. But the preferences revealed — or inferable — from a person's actions need not align with their actual fundamental preferences when an actor might have two reasons for making a choice: either strong underlying morality or a strong desire to *appear* moral in order to reap future financial rewards. Given this, the law could cause people in a certain domain of conduct to engage in "Pareto self-improvement" with respect to the morality of their *choices* — so that they can appear moral to counterparties — without actually changing their criteria of evaluation (or the integration thereof), which could explicitly govern their conduct in the aspects of their business life that they do not expect counterparties to learn about.⁹⁶

In other words, Janice in our example could be motivated to obey all laws whose violation she anticipates could be detected and that cause harms that her counterparties care about purely to create beliefs that she is a moral person, without ever changing the criteria of evaluation that constitute her preferences. She could obtain the desired increase in opportunities resulting from creating the perception that she is moral by obeying those laws that (1) counterparties care about, when (2) there is a material risk that noncompliance would be detected if she did not have any genuine moral or social concerns.

Thus, again, this account of how the law changes preferences in fact entails that the law alter choices without altering people's criteria of evaluation (or their integration process). Instead, the law alters the benefits to each agent of making the choice to be legally compliant when such conduct is potentially detectable. The law on this account does not lead people who did not previously treat ethics, norms and morality as material criteria of evaluation to decide suddenly that they care about such things — even in situations where they could violate the law without any impact on their future opportunities.

IV. DISCUSSION AND CONCLUSION

Thus, the existing theories about how the criminal and civil law can alter behavior by expressing society's view reveal that the law does so by expressing that the conduct is socially harmful or immoral, and thereby altering people's

96 Oren Bar-Gill and Chaim Fershtman offer a related theory of how the law can change preferences. In their theory, the legal rules that favor fair contracting enable people with preferences for fairness to fare better than others. Over time, those people will come to dominate the market. Yet this theory does not involve any individual changing his or her preferences. Moreover, they also assume that people's preferences are observable and thus "unfair" people cannot simply mimic "fair" people. See Bar-Gill & Fershtman, *supra* note 6, at 338 n.7, 343-46.

willingness to engage in the undesirable conduct. Yet under each of these theories the law alters behavior by changing people's perception of the attributes of the regulated conduct. The channels of influence proposed do not induce changes to people's fundamental preferences. The law does not change either people's underlying criteria of evaluation under these claims or how those criteria are integrated into an all-things-considered preference.⁹⁷

This conclusion makes it evident that rather than being a direct challenge to rational choice theory, the preference-shaping literature is consistent with, and operates to enrich, rational choice theory. The three channels of influence identified by the preference-shaping literature expand the mechanisms through which the law can exert influence beyond the narrow set identified by classical deterrence theory. Yet each of the mechanisms has been identified by rational choice theory as a potential channel through which the state (here through the law) could alter behavior: by changing the perceived or actual consequences of an action by changing either people's beliefs about the harm caused, the internal costs to them of the action (e.g., as a result of guilt or shame), or the externally imposed consequences of the action, as when the conduct is likely to lead the actor to suffer costs from reputational damage. None of these channels entails use of the mechanism that rational choice explicitly rules out: direct alteration of people's criteria of evaluation and thus of their fundamental preferences. As a result, contrary to the claims of some,⁹⁸ embracing the expressive power of law does not undermine the normative power of economic efficiency concepts, since under these expressive law theories preferences (and thus fundamental welfare rankings) remain stable even after the law's adoption.

Our analysis also reveals that the expressive law theories are entirely consistent with deterrence theory. Moreover, they also are complementary to, and mutually dependent on, the government sanctions examined by classical

97 A similar conclusion may apply to many of the ways in which the law can alter behavior under behavioral law and economics models. For example, to the extent that the endowment effect is predicated on regret aversion, then a legal change that shifts initial endowments alters an agent's willingness to trade by changing her anticipated regret over a decision to trade. In so doing, it alters the consequences to her of a decision to trade. It does not alter her preferences. This is not to say that legal interventions never alter preferences once we expand our analysis to incorporate deviations from rational choice. We simply observe that many interventions that rely on behavioral deviations from rational choice that may appear to alter preferences do not do so once the underlying cause of the behavioral deviation (such as regret or guilt) is identified.

98 See citations in *supra* note 7.

deterrence theory.⁹⁹ Thus, contrary to the claims of some,¹⁰⁰ embracing the expressive power of law does not justify reduced reliance on enforcement and sanctions against those violations that do occur. Indeed, the opposite may be true. In order to alter people's beliefs about either social harm or social and ethical norms, civil and criminal laws generally need to be enforced and perpetrators sanctioned. This is the case for several reasons. First, society cannot credibly express opprobrium by adopting a law that is never enforced. Actual enforcement, and the sanctions imposed, are part of how society expresses its view that the conduct is socially harmful.¹⁰¹ In addition, establishment of a social norm requires that people expect others to comply with the law. Sanctions are needed to achieve this goal when either many people do not have other-regarding preferences or otherwise ethical people would violate the law, absent the threat of sanctions, because the material gains from violating the law swamp the militating effect of guilt or shame.

The conclusion that the law's expression of social meaning alters behavior by increasing the costs of the prohibited conduct, instead of by changing people's fundamental wants and desires, implies that society can seek to exploit the criminal and civil law's expressive features without any risk of interfering with people's individuality and autonomy.¹⁰² The law's expressions through these mechanisms leave people's preferences untouched. It may make certain conduct more costly — and thereby deter it — but it does not interfere with people's ability to wish that they could engage in the conduct were it less costly.

Our argument that the preference-shaping claims examined in this Article do not support the conclusion that civil and criminal liability for harmful conduct changes preferences does not imply that the law cannot result in a change in people's preferences through other channels.¹⁰³ For example, the adoption *and* subsequent enforcement of a law could alter the environment within which people live, leading them to have a different set of experiences as they are growing up and forming their preferences. Thus, for example, laws prohibiting discrimination in employment, housing and schools —

99 See Kahan, *supra* note 8 (discussing the dependence of criminal law norm creation on sanctions and enforcement).

100 *E.g.*, Bilz & Nadler, *supra* note 6; Dau-Schmidt, *supra* note 6; Sunstein, *supra* note 6.

101 *See id.*

102 *See* citations in *supra* note 6.

103 *E.g.*, Kaplow & Shavell, *supra* note 12, at 1335 & n.912; *see also* MICHAEL J. TREBILCOCK, *THE LIMITS OF FREEDOM OF CONTRACT* 196 (1997) (discussing evidence that immigration to Canada reduced residents' discriminatory attitudes by altering their experiences).

if effectively enforced — could cause people to have more contact, and develop friendships, with people from different races, thereby leading them not to share their parents' negative attitudes about people of different races. Yet even if law may change preferences in this way, it can do so only if it is enforced sufficiently to alter people's environment. Moreover, this effect is not immediate. Such changes take time and may be more possible during a person's formative years than later. To understand law's role, if any, in such transformations requires a theory of how our preferences are formed. Such a theory would include an understanding of moral development and of the stages in life at which fundamental changes in our basic criteria of evaluation are possible or likely. Legal change, combined with other social changes, possibly may alter the social environment so that those born after the adoption of the law develop different preferences than those whose character was formed before the legal change.¹⁰⁴ But the process of getting there will inevitably require sanctions and enforcement. In the short- to medium run, achieving the change in the environment may require the state to leverage all the tools offered by rational choice theory, including both the threat of sanctions and the expressive effects of law.

104 People's natures appear to be most affected by environmental factors they experience earlier in life. It is far from clear that a changed social environment will have the same effect on shaping the preferences of older people. There is reason to believe, for example, that civil rights legislation and enforcement decades ago resulted in today's young people growing up in a more diverse and more tolerant environment than their parents. This environmental difference could have led more of them to care about eliminating race, sex or sexual preference discrimination. But any such change did not result directly from the law but rather from the influence of the law on the circumstances in which people were raised. And such changes do not necessarily reflect a change in our most fundamental preferences — e.g., people's tendency to disfavor those they view to be "other" — but instead result from the influence of the law, and in turn social change, on inducing a subset of society to adopt a different view of who constitutes people who are "other."

