

Preference Change and Behavioral Ethics: Can States Create Ethical People?

Yuval Feldman and Yotam Kaplan***

Law and economics scholarship suggests that, in appropriate cases, the law can improve people's behavior by changing their preferences. For example, the law can curb discriminatory hiring practices by providing employers with information that might change their discriminatory preference. Supposedly, if employers no longer prefer one class of employees to another, they will simply stop discriminating, with no need for further legal intervention. The current Article aims to add some depth to this familiar analysis by introducing the insights of behavioral ethics into the law and economics literature on preference change. Behavioral ethics research shows that wrongdoing often originates from semi-deliberative or non-deliberative cognitive processes. These findings suggest that the process of preference change through the use of the law is markedly more complicated and nuanced than previously appreciated. For instance, even if an employer's explicit discriminatory stance is changed, and the employer no longer consciously prefers one class of employees over another, discriminatory behavior might persist if it originates from semi-conscious, habitual, or non-deliberative decision-making mechanisms. Therefore, actual change in behavior might necessitate a close engagement with people's level of moral awareness. We discuss the institutional and normative implications of these insights and

* Mori Lazarof Professor of Legal Research, Bar-Ilan University Law School, PhD UC Berkeley.

** Assistant Professor, Bar-Ilan University Law School, S.J.D Harvard Law School. We wish to thank Kathryn Abrams, Robert Cooter, Alon Jasper, Maya Hay, Nir Kremerman, Saul Levmore, Alon Klement, Ariel Porat, Or Shua, Jasmin Wennersbusch, and participants in the 2020 UC Berkeley conference in honor of Robert Cooter for insightful comments and discussions. Onn Israeli provided excellent research assistance.

evaluate their significance for the attempt to improve preferences through the different functions of the legal system.

INTRODUCTION

The concept of preferences is central to contemporary economic thinking and to economic modeling of human behavior. As used by economists, preferences typically refer to the tastes, values, desires and wishes people have that would make them prefer one state of affairs to another.¹ Thus, if a person prefers the state of being healthy to the state of being sick, economists would say this person holds a preference for good health. To take another example, one can be said to have a preference for ethical behavior if they would prefer, all things being equal, to act ethically rather than unethically.² More generally, people can have preferences regarding types of food, personal goals, the company of others, certain types of behaviors and activities, financial statuses, the welfare of others, and so on. In conceptualizing preferences, economists traditionally assume that people's preferences are unchanging, exogenously determined, and are largely taken as a given.³ Thus, the traditional economic viewpoint is that preferences are stable,⁴ and do not change over time.⁵

More recently, economists have been increasingly willing to relax the assumption regarding the unchanging nature of preferences.⁶ This is of special interest to law and economics scholars: once it is assumed that preferences can change over time, this facilitates an important discussion regarding the possibility of using the law as a tool to change (and improve) preferences.⁷ Robert

-
- 1 TILL GRÜNE-YANOFF & SVEN OVE HANSSON, PREFERENCE CHANGE: APPROACHES FROM PHILOSOPHY, ECONOMICS AND PSYCHOLOGY 8 (2009); GARY S. BECKER, THE ECONOMIC APPROACH TO HUMAN BEHAVIOR 5 (1976).
 - 2 Robert Cooter, *Do Good Laws Make Good Citizens: An Economic Analysis of Internalized Norms*, 86 VA. L. REV. 1577, 1581 (2000).
 - 3 George J. Stigler & Gary S. Becker, *De Gustibus Non Est Disputandum*, 67 AM. ECON. REV. 76, 76 (1977) (“One does not argue about tastes for the same reason that one does not argue about the Rocky Mountains — both are there, and will be there next year, too, and are the same to all men.”).
 - 4 Steven E. Landsburg, *Taste Change in the United Kingdom, 1900-1955*, 89 J. POL. ECON. 92 (1981) (offering empirical support for the assumption regarding the stability of preference over time).
 - 5 *Id.*
 - 6 Ariel Porat, *Changing People's Preferences by the State and the Law*, 22 THEORETICAL INQUIRIES L. 215 (2021).
 - 7 Cooter, *supra* note 2. Cf. Jennifer Arlen & Lewis Kornhauser, *Does the Law Change Preferences?*, 22 THEORETICAL INQUIRIES L. 175 (2021) (arguing that

Cooter, as a pioneer of this new direction in law and economics scholarship, suggests that the possibility of changing people's preferences offers new venues for improving behavior and enhancing welfare.⁸ Cooter and others argue that this option is especially attractive, as improving preferences can lead people to improve their behavior with no need for further enforcement efforts.⁹ This is considered crucial, as widespread compliance with legal and social norms depends on people's ethical commitments more than on direct legal measures.¹⁰

We seek to contribute to the law and economics literature on preference change by considering its relation to the burgeoning scholarship on behavioral ethics and the law.¹¹ We point out that the law and economics literature on preference change still maintains a rather monolithic view of preferences. That is, this literature allows for preferences to change over time, but it still holds a strong implicit assumption regarding the internal consistency of preferences, assuming that people's preferences directly translate into their conduct and choices.¹² For instance, Cooter maintains that if people's preferences are changed and improved, this is synonymous with a positive change in behavior.¹³

We suggest that this assumption should be reevaluated in light of behavioral ethics findings. Works in cognitive and social psychology suggest that it is possible for a person to hold an explicit preference for one state of affairs, yet systematically make choices that seem to contradict this preference.¹⁴ In the context of ethical decision-making, it is possible (and even common) for a person to hold an explicit preference for ethical behavior, yet systematically behave unethically.¹⁵ The growing behavioral ethics literature connects this type of wrongdoing to implicit cognitive mechanisms and to semi-deliberative and

the law does not truly change preferences but merely provides actors with new information).

8 Cooter, *supra* note 2.

9 *Id.* at 1600.

10 Robert D. Cooter, *Three Effects of Social Norms on Law: Expression, Deterrence, and Internalization*, 79 OR. L. REV. 1 (2000); Gerard E. Lynch, *The Role of Criminal Law in Policing Corporate Misconduct*, 60 LAW & CONTEMP. PROBS. 23, 46 (1997).

11 YUVAL FELDMAN, *THE LAW OF GOOD PEOPLE: CHALLENGING STATES' ABILITY TO REGULATE HUMAN BEHAVIOR* 1 (2018).

12 Cooter, *supra* note 2.

13 *Id.* at 1581.

14 DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* 20-21 (2011).

15 FELDMAN, *supra* note 11, at 1 ("various psychological and social mechanisms . . . prevent people from recognizing their wrongdoing and encourage them to feel as if they are far more moral, unbiased, and law abiding than they actually are").

non-deliberative decision-making processes.¹⁶ *Semi-deliberative* wrongdoing occurs when potential perpetrators realize they are facing a moral dilemma, but fail to realize their moral deliberations are biased.¹⁷ For example, people may use their past good deeds to justify future wrongdoing, which can distort their ethical decision-making.¹⁸ In such cases, wrongdoing is semi-deliberative, as people fail to conduct a fully candid moral deliberation, and fail to realize what effects their judgement; their decision-making is biased, as moral licensing allows them to act in ways that diverge from their declared values and ethical commitments.¹⁹ *Non-deliberative* wrongdoing occurs when different biases prevent perpetrators from engaging in moral deliberations at all, and even from recognizing they are faced with a moral dilemma.²⁰ In such cases, wrongdoing is non-deliberative, as it originates from cognitive mechanisms that push individuals to ignore uncomfortable ethical questions.²¹ These insights also relate to the literature on situational wrongdoing, showing that minor situational alterations can bias people's ethical deliberations.²²

Carrying over these findings to the literature on preference change, we argue that preferences should not be viewed as monolithic in this framework, but as fragmented, as people behave in ways that indicate the existence of fractions or inconsistencies in their ethical preferences. Such behavioral inconsistencies present a significant challenge for the effort to improve preferences through

16 Jonathan Haidt, *The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814, 814-15 (2001) (arguing that moral reasoning is typically the result of quick, automatic evaluation and that rational justifications are only made after the fact).

17 Ann E. Tenbrunsel & David M. Messick, *Ethical Fading: The Role of Self-Deception in Unethical Behavior*, 17 SOC. JUST. RES. 223, 228 (2004).

18 Anna C. Merritt, Daniel A. Effron & Benoît Monin, *Moral Self-Licensing: When Being Good Frees Us to Be Bad*, 4 SOC. & PERSONALITY PSYCHOL. COMPASS 344, 344 (2010).

19 *Id.*; Emily Pronin, Thomas Gilovich & Lee Ross, *Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others*, 111 PSYCHOL. REV. 781, 781-82 (2004).

20 Emily E. Balcetis & David Dunning, *See What You Want to See: Motivational Influences on Visual Perception*, 91 J. PERSONALITY & SOC. PSYCHOL. 612 (2006); Gerd Gigerenzer, *The Bias in Behavioral Economics*, 5 REV. BEHAV. ECON. 303 (2018).

21 *Id.*

22 MAX H. BAZERMAN & ANN E. TENBRUNSEL, *BLIND SPOTS: WHY WE FAIL TO DO WHAT'S RIGHT AND WHAT TO DO ABOUT IT* 1-3 (2011) (explaining the concept of ethical blind spots, situations in which ethical deliberation is hindered and unethicity therefore proliferates).

the use of the law. If people's preferences indeed lack internal consistency, the relationship between preferences and behaviors is more complex than is currently assumed in the law and economics preference-change literature. In particular, if people hold a conscious preference to behave in a socially desirable way, this still does not ensure such behavior will indeed occur.²³ Therefore, improving a person's expressed belief or preference, or encouraging a "taste for fairness,"²⁴ will not necessarily result in a positive change in behavior.

We suggest a possible reorientation of the effort to improve preferences through the use of the law. That is, improving the actual (often non-deliberative) ethicality of people's behavior may require constant maintenance of moral awareness, rather than a discrete intervention designed to alter expressed attitudes. The challenge of improving behavior is especially great since implicit attitudes and habits are often more stable, and more difficult to change, than explicit preferences. We develop this point to present a critique of the preference-change endeavor, and suggest that the law may in fact be ill-equipped to produce long-lasting improvements in people's internal drive to behave ethically and legally. As an alternative, we suggest that the main role of the law should be to improve ethical awareness indirectly, through changing conventional social norms, institutions and organizations.

The Article proceeds as follows. Part I focuses on the existing law and economics literature on preference change. It explains the basic assumptions this literature utilizes, its goals, and its importance. It shows that the ability to change preferences holds great promise for law and policymaking, and that it challenges the traditional understanding of law enforcement. Part II offers a critique of the law and economics literature on preference change by introducing several insights originating in the behavioral ethics literature. Specifically, it shows that changing explicit ethical preferences may not be a relevant remedy, since wrongdoing often originates from implicit rather than explicit decision-making processes. Part III completes the discussion by exploring the possibility of using the law to generate long-lasting improvements in ethical awareness. It argues that, in line with behavioral ethics findings,

23 Traditional preference change literature recognizes external constraints on people's preferences. That is, it might be that a person holds a preference for helping others, but she is not able to act on that preference since she does not possess the necessary financial means. Our argument in this Article adds another layer, that of internal constraints, to the factors limiting people's ability to realize their preferences. That is, we argue that even if a person has a preference for helping others, and that person is able to realize this preference as a matter of external constraints, it might still be the case that cognitive biases interrupt and limit that person's ability to realize her explicit preference.

24 Cooter, *supra* note 2, at 1579.

that improving behavior would necessitate not only a change in explicit ethical preferences, but also an improvement in people's capacity for candid ethical deliberations. We conclude this Part by raising doubts regarding the ability of the law to generate such a change independently; as an alternative, we suggest that the law might be able to improve ethical deliberations when operating in conjunction with other societal mechanisms.

I. THE LAW AND ECONOMICS LITERATURE ON PREFERENCE CHANGE

The scholarly tradition of law and economics aims to improve behavior in order to increase total social welfare. The standard method within the rich law and economics literature is for the state to introduce appropriate sanctions or rewards in order to improve incentives and therefore behavior and social outcomes.²⁵ To illustrate, consider the example of an employer screening CVs of prospective employees while holding a discriminatory preference against a specific group of people. To induce the employer to stop discriminating, the state can introduce sanctions that will make discriminating not worthwhile for the employer, despite her preference against one group of employees. Thus, the employer's preference is taken as a given, and sanctions are tailored around it in order to improve behaviors and achieve desirable social goals.²⁶ Note that the employer's discriminatory preference remains unchanged; the employer must simply balance this preference against her preference to avoid legal sanction.

As an alternative to this traditional approach, scholars are increasingly willing to consider the possibility that the state might attempt to change and improve preferences.²⁷ Robert Cooter was the first law and economics scholar to systematically study this option, under the heading of "internal motivation."²⁸

25 WERNER Z. HIRSCH, *LAW AND ECONOMICS: AN INTRODUCTORY ANALYSIS* 1 (2d ed. 1988) ("[L]aws are authoritative directives that impose costs and benefits on participants in a transaction and in the process alter incentives.").

26 Steven Shavell, *Law Versus Morality as Regulators of Conduct*, 4 *AM. L. & ECON. REV.* 227, 227 (2002) ("It is evident that both law and morality serve to channel our behavior. Law accomplishes this primarily through the threat of sanctions if we disobey legal rules.").

27 For an early version of this idea, see J.T. Romans, *Moral Suasion as an Instrument of Economic Policy*, 56 *AM. ECON. REV.* 1220 (1966).

28 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); Cooter, *supra* note 2. For a discussion of some of the main forms of internalization, see Yuval Feldman & Robert MacCoun, *Some Well-Aged Wines for the 'New Norm' Bottles, Implications of Social Psychology to Law and Economics*, in *THE LAW*

For instance, in the example described above, the state might act to alter the employer's discriminatory preference, or cause her to "internalize" a more egalitarian norm. If this is accomplished, the employer will stop discriminating, not due to the fear of sanction, but simply because she no longer holds a preference for employees of one group over another. Importantly, if the employer's preference is changed, no further enforcement action is needed to prevent discrimination. Under a deterrence approach, employers will refrain from discriminating only if they think they are likely to be caught and sanctioned. Conversely, if employers' discriminatory preference is changed, they will cease discriminating regardless of any such calculus. Cooter emphasizes the crucial importance of this advantage, as general compliance with the law depends on internalization rather than on deterrence.²⁹ It is simply unrealistic to expect all people to behave legally all the time out of fear of a direct state sanction.³⁰ Therefore, for society to function, most people have to obey the law for reasons of conscience and conviction, and not out of fear of punishment.³¹

Internal motivation as a source of compliance with the law has been studied intensively,³² with tax evasion being a classic case study.³³ Richard Schwartz

AND ECONOMICS OF IRRATIONAL BEHAVIOR 358 (Francesco Parisi & Vernon Smith eds., 2005).

29 Cooter, *supra* note 2, at 1589.

30 Lynch, *supra* note 10, at 46.

31 *Id.*

32 Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765 (1996) (studying the effect of internalized social norms on compliance and cooperation in the context of contract law); Robert D. Cooter, *Punitive Damages, Social Norms, and Economic Analysis*, 60 LAW & CONTEMP. PROBS. 73 (1997); Robert D. Cooter, *Structural Adjudication and the New Law Merchant: A Model of Decentralized Law*, 14 INT'L REV. L. & ECON. 215 (1994); Melvin A. Eisenberg, *Corporate Law and Social Norms*, 99 COLUM. L. REV. 1253 (1999) (studying internal motivation as a source of compliance in the context of corporate law); Steven Hetcher, *Creating Safe Social Norms in a Dangerous World*, 73 S. CAL. L. REV. 1 (1999) (studying internal motivation as a source of compliance in tort law); Eric A. Posner, *Family Law and Social Norms*, in *THE FALL AND RISE OF FREEDOM OF CONTRACT* 256 (F.H. Buckley ed., 1999) (studying internalization in family law); Eric A. Posner, *Symbols, Signals, and Social Norms in Politics and the Law*, 27 J. LEGAL STUD. 765 (1998); Elizabeth S. Scott & Robert E. Scott, *A Contract Theory of Marriage*, in *THE FALL AND RISE OF FREEDOM OF CONTRACT* 256 (F.H. Buckley ed., 1999); Lior Jacob Strahilevitz, *How Changes in Property Regimes Influence Social Norms: Commodifying California's Carpool Lanes*, 75 IND. L.J. 1231 (2000) (offering internalization as a source of compliance with traffic regulation).

33 Eric Posner, *Law and Social Norms: The Case of Tax Compliance*, 86 VA. L.

and Sonya Orleans provide a famous example, demonstrating in an experiment that convincing people to pay taxes is more effective than threatening them with punishment in case they refuse to do so.³⁴ Similar results have been presented in multiple contexts, where the limits of deterrence and extrinsic motivation were compared empirically to the force of internal motivation to comply with the law.³⁵ One of the most popular lines of research within this literature is related to the view that deterrence cannot account for the level of compliance we observe in society.³⁶ Scholars therefore offer that compliance originates not only with deterrence, but also with intrinsic factors such as duty, legitimacy, and a sense of moral commitment.³⁷ Many of the studies

REV. 1781 (2000); James Andreoni et al., *Tax Compliance*, 36 J. ECON. LIT. 818, 820-21 (1998); Michael J. Graetz et al., *The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement*, 2 J.L. ECON. & ORG. 1 (1986).

- 34 Richard D. Schwarz & Sonya Orleans, *On Legal Sanctions*, 34 U. CHI. L. REV. 274 (1967); for a critique of Schwarz & Orleans, see LAWRENCE M. FRIEDMAN & STEWART MACAULAY, *LAW AND THE BEHAVIORAL SCIENCES* 324 (2d ed. 1977). For a more controlled experiment which comes to the opposite conclusions, see Charles R. Tittle & Alan R. Rowe, *Moral Appeal, Sanction Threat, and Deviance: An Experimental Test*, 20 SOC. PROBS. 488 (1973). Robert Kagan has shown that the most important factor predicting tax compliance is the ability of the authorities to monitor income. See Robert A. Kagan, *On the Visibility of Income Tax Law Violations*, in 2 TAXPAYER COMPLIANCE 76 (Jeffery A. Roth & John T. Scholz eds., 1989).
- 35 Matthew Silberman, *Toward a Theory of Criminal Deterrence*, 41 AM. SOC. REV. 442 (1976); FRANCLIN E. ZIMRING & GORDON J. HAWKINGS, *DETERRENCE: THE LEGAL THREAT IN CRIME CONTROL* (1973).
- 36 Raymond Paternoster & LeeAnn Iovanni, *The Deterrent Effect of Perceived Severity: A Reexamination*, 64 SOC. FORCES 751, 768 (1986) (arguing that both severity and certainty of punishment could not account for any effect of delinquent behavior). Others were more modest in their arguments, but still demonstrated through a review of many empirical studies that deterrence could not fully account for compliance; John Braitwaite & Toni Makkai, *Testing an Expected Utility Model of Corporate Deterrence*, 25 LAW & SOC'Y REV. 7 (1991). Others have suggested that the problem was not motivational but rather cognitive, as people are not really cognizant of the law on the books. See Paul H. Robinson & John M. Darley, *Does Criminal Law Deter? A Behavioral Science Investigation*, 24 OXFORD J. LEGAL STUD. 173 (2004).
- 37 Tom R. Tyler & Jeffrey Fagan, *Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in their Communities?*, 6 OHIO ST. J. CRIM. L. 231 (2008); Michael Wenzel, *The Impact of Outcome Orientation and Justice Concerns on Tax Compliance: The Role of Taxpayers' Identity*, 87 J. APPLIED PSYCHOL. 629 (2002); James L. Gibson, *The Legitimacy of the U.S. Supreme*

within this tradition focus not only on the limits of deterrence, but also on the comparison between the different mechanisms, deterrence and internal motivation, in their relative effectiveness in changing people's behavior.³⁸ Studies have also highlighted the advantages of internalization in generating desirable outcomes that can go beyond the strict requirements of the law.³⁹

The advantages of internal motivation as a basis for widespread compliance raise an important question regarding the possibility of using the law to change harmful preferences for more beneficial ones. Cooter suggests that the law can contribute to preference change indirectly, by piggybacking on conventional morality, social norms, and intimate social dynamics. The starting point for Cooter's argument is that preferences can change due to the existence of high-order and low-order preferences.⁴⁰ High-order preferences represent primary goals, while low-order preferences represent intermediate ones. For instance, a person might have a high-order preference for good health, which in turn leads to a low-order preference for drinking water over drinking juice. The existence of high- and low-order preferences facilitates preference change, since people may want to change their low-order preferences to maximize their high-order preferences.⁴¹

First, the distinction between high- and low-order preferences can allow preference change through the expressive function of the law.⁴² For instance, on the assumption that an employer has a high-order preference for ethical conduct, the law might change that employer's low-order discriminatory preference by presenting information according to which discrimination is considered widely immoral. Once such information is provided, and in accordance with the employer's high-order preference for morality, she might change her low-order discriminatory preference.⁴³ Cooter's arguments regarding the expressive

Court in a Polarized Polity, 4 J. EMPIRICAL LEGAL STUD. 507 (2007). For an empirical demonstration of the limits of traditional economic models in the context of legal compliance, see Yuval Feldman & Doron Teichman, *Are all Legal Probabilities Created Equal?*, 84 N.Y.U. L. REV. 840 (2009).

38 Schwarz & Orleans, *supra* note 34.

39 Yuval Feldman, *The Complexity of Disentangling Intrinsic and Extrinsic Compliance Motivations: Theoretical and Empirical Insights from the Behavioral Analysis of Law*, 35 WASH. U. J.L. & POL'Y 11 (2011) (emotions like love can drive individuals to altruistic behaviors beyond what is required by law).

40 Cooter, *supra* note 2.

41 *Id.* at 1596.

42 Robert Cooter, *Expressive Law and Economics*, 27 J. LEGAL STUD. 585 (1998).

43 Richard McAdams made important contributions to this literature as well: Richard H. McAdams, *The Origin, Development, and Regulation of Norms*,

function of the law are supported in multiple studies.⁴⁴ For example, Nigel Walker and Michael Argyle have demonstrated empirically that the law can change people's views about what society deems morally wrong.⁴⁵

Cooter suggests a second, more instrumental mechanism, by which the law might utilize high-order preferences to change undesirable low-order preferences.⁴⁶ Cooter posits that people can improve their social standing by being perceived as virtuous. Thus, to satisfy high-order preferences for elevated social status, people might change their low-order ethical preferences, in order to appear moral to others,⁴⁷ in a process Cooter terms "Pareto self-improvement."⁴⁸ Of course, this mechanism will only truly improve ethicality if people cannot falsely present themselves as virtuous, or at least if this is costly to accomplish.⁴⁹ Thus, Cooter argues that if the best way to *appear* moral is to actually *be* moral, then people will want to become moral and change their ethical preferences.⁵⁰ Therefore, close social groups, in which people's true morality is observable, are instrumental in improving ethical preferences.⁵¹ Piggybacking on these social mechanisms, the state can then improve compliance with the law by enacting laws that have moral flavors. That is, if people care to appear moral to others, and if the law is equated (at least somewhat) with morality, people will wish to appear lawful and will change their preferences accordingly.⁵²

96 MICH. L. REV. 338 (1997); Richard H. McAdams, *A Focal Point Theory of Expressive Law*, 86 VA. L. REV. 1649 (2000).

44 For an excellent longitudinal study on internalization of legal rules and moral values, see ELLEN S. COHN & SUSAN O. WHITE, *LEGAL SOCIALIZATION: A STUDY OF NORMS AND RULES* (1990).

45 Nigel Walker & Michael Argyle, *Does the Law Affect Moral Judgments?*, 4 BRIT. J. CRIMINOLOGY 570 (1964).

46 Cooter, *supra* note 2, at 1581.

47 *Id.* at 1594.

48 *Id.* at 1581.

49 For a critical discussion of this assumption, see Robert, E. Scott, *The Limits of Behavioral Theories of Law and Social Norms*, 86 VA. L. REV. 1603 (2000).

50 Cooter, *supra* note 2, at 1597.

51 *Id.*

52 *Id.* at 1598. Similar arguments are advanced by Paul Robinson and John Darley, who argue that legal rules are more effective when they converge with conventional morality: Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453 (1997). See also TOM R. TYLER, *WHY PEOPLE OBEY THE LAW* 65 (1990); Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661 (1998); Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943, 964-73 (1995); Lawrence Lessig, *Social Meaning and Social Norms*, 144 U. PA. L. REV. 2181 (1996).

These mechanisms Cooter suggests are based on a rational choice model of decision-making.⁵³ Under Cooter's view, people change their preferences as a logical response to new information or new social opportunities.⁵⁴ According to such a model, once preferences change and an individual acquires a stronger "taste for fairness" or a preference for moral conduct, the assumption is that this taste immediately translates to behavior.⁵⁵ These assumptions are the target of the critique we offer below.

II. PREFERENCE CHANGE: A BEHAVIORAL ETHICS CRITIQUE

The literature on preference change described above relaxes the assumption that preferences stay stable over time. It largely maintains, however, an implicit assumption that preferences are monolithic, and are directly reflected in behavior. In other words, the literature on preference change assumes that if a person is made to prefer one outcome to another, this person will necessarily choose that preferred outcome; in fact, this is the very definition of a revealed preference.⁵⁶

While this assumption is of course a useful modeling technique, it is a stark oversimplification and ignores any behavioral complexity or decision-making mechanism. The cognitive processes by which an individual chooses one set of affairs over another have been shown to be complex and fragmented, rather than straightforward and monolithic. People have conscious preferences, wishes and values, but also implicit tendencies, habits and biases. People can be expressly committed to a set of moral values (or preferences) but behave in ways that seem to directly contradict those commitments.⁵⁷

We therefore seek to introduce a richer model of decision-making into the law and economics literature on preference change, by highlighting cognitive biases and behavioral decision-making mechanisms. This move offers a more nuanced view of the preference change project: if and when wrongdoing does not originate from explicit preferences, it is unclear that using the law to change explicit preferences can effectively improve behavior. Therefore, we must consider the ability of the law to change habits and implicit decision-making processes, and not only explicit attitudes or values. In other words, some work needs to be done in order to better define how "preferences" are used by law

53 Cooter, *supra* note 28, at 903.

54 Cooter, *supra* note 2, at 1581.

55 *Id.* at 1587-89.

56 *Id.*

57 Mary C. Kern & Dolly Chugh, *Bounded Ethicality: The Perils of Loss Framing*, 20 PSYCHOL. SCI. 378, 381-83 (2009).

and economics scholars in the context of the preference-change literature, and how these preferences stand in relation to psychological concepts such as implicit attitudes or intuitions.

To illustrate these general claims, let us return to the example of discriminatory hiring practices described above. The literature on preference change assumes that such negative antisocial behaviors can be prevented if the preferences of the relevant actors are altered. Thus, if an employer holds a racist preference, she might discriminate against people of one ethnicity in making hiring decisions; to solve this problem, the literature on preference change would suggest that the law should intervene to facilitate a process of “Pareto self-improvement,” by which the employer will change her explicit discriminatory preference for a more egalitarian one.⁵⁸ That is, if the law indicates that an egalitarian approach is more virtuous, the employer, wishing to enjoy the status benefits of a virtuous person, might truly internalize an egalitarian worldview and stop discriminating.

Behavioral insight greatly complicates this picture, suggesting that even if norms are “internalized,” there is no guarantee that the change in the employer’s expressed beliefs will translate into a change in behavior; rather, this will depend on the decision-making processes that generate discriminatory outcomes. Such processes may be deliberative, semi-deliberative, or non-deliberative.⁵⁹ First, prior to her change in preferences the employer may have knowingly decided to discriminate, believing this to be the best course of action for her business. We consider this to be deliberative wrongdoing if the employer is fully aware of all the relevant factors and nevertheless chooses to discriminate through a calculated process. This is most typically the assumption in the law and economics literature on preference change, in which the moral agent makes an explicit choice between an ethical and an unethical act. If wrongdoing is fully deliberative and originates from an express preference, changing that preference can indeed be effective.

However, the employer can also discriminate based on a semi-deliberative process. For instance, the employer may deliberate and choose people of one ethnicity only, but convince herself, through a biased decision-making process, that she is not discriminating at all.⁶⁰ That is, the employer may deliberate, but fail to do so candidly and objectively. Studies suggest that this commonly occurs, as people’s biases lead them to think of themselves as virtuous and

58 Porat, *supra* note 6, at 244-45.

59 Haidt, *supra* note 16, at 814-15.

60 Yuval Feldman & Yotam Kaplan, *Behavioral Ethics as Compliance*, in *THE CAMBRIDGE HANDBOOK OF COMPLIANCE* 50 (D. Daniel Sokol & Benjamin Van Rooij eds., 2021).

objective, and to systematically prefer interpretations that present them as such.⁶¹ Therefore, even if the employer thinks discrimination is wrong, she will simply prefer the interpretation that she chooses employees based on merit, and not based on a discriminatory preference. Due to objectivity biases, the employer will also not be aware of the biased nature of her choice.⁶² When this is the case, changing the employer's explicit discriminatory preference will not necessarily improve behavior, as the employer's semi-deliberative choices will simply bypass her new and improved preference.

Finally, the employer may discriminate due to completely non-deliberative choice processes.⁶³ Discrimination is non-deliberative if, in screening thousands of CVs, the employer unwittingly chooses more employees of one group without even realizing she is doing so and without any conscious deliberative process or awareness of the ethical dilemma.⁶⁴ In such cases, the employer discriminates without meaning to do so (as is often the case), due to implicit rather than explicit racial biases.⁶⁵ When this is the case, it is not clear that changing explicit preferences and expressed values will solve the problem, as discrimination originates from unconscious habits. Below, we generalize on this example and offer evidence for the prevalence of implicit wrongdoing, to suggest a critique of the preference-change argument in its current form.

A. The Prevalence of Implicit Wrongdoing

Research in behavioral ethics shows that wrongdoing or noncompliance is often the result of implicit, rather than explicit, cognitive processes.⁶⁶ That

61 Dolly Chugh, Max H. Bazerman & Mahzarin R. Banaji, *Bounded Ethicality as a Psychological Barrier to Recognizing Conflicts of Interest*, in *CONFLICTS OF INTEREST: CHALLENGES AND SOLUTIONS IN BUSINESS, LAW, MEDICINE, AND PUBLIC POLICY* 74, 74 (Don. A. Moore et al. eds., 2005) (explaining that people view themselves as more objective than others and are therefore unable to see themselves as corrupt).

62 Pronin, Gilovich & Ross, *supra* note 19, at 781-82.

63 FELDMAN, *supra* note 11, at ch. 10; Marianne Bertrand, Dolly Chugh & Sendhil Mullainathan, *Implicit Discrimination*, 95 *AM. ECON. REV.* 94 (2005).

64 Research shows that most discriminatory decisions in the employment field are made implicitly rather than explicitly; Linda Hamilton Krieger, *The Content of our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 *STAN. L. REV.* 1161, 1164 (1995); Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 *CALIF. L. REV.* 997, 1027-30 (2006).

65 Krieger, *supra* note 64, at 1164; Krieger & Fiske, *supra* note 64, at 1027.

66 Jennifer J. Kish-Gephart, David A. Harrison & Linda Klebe Trevino, *Bad*

is, many instances of wrongdoing originate with semi-deliberative or non-deliberative choice, when individuals do not choose to behave “badly” but arrive at a socially harmful choice while convincing themselves they are doing no wrong, or while different biases cause them to ignore or underappreciate the harmfulness of their own actions.⁶⁷ If indeed wrongdoing can be the product of implicit attitudes, automatic reasoning or non-deliberative choice, this suggests a major challenge to the law’s ability to increase ethical behavior by changing explicit preferences.

We argue that it cannot be simply assumed that changing explicit preferences will increase compliance. Rather, the effect of preference change on behavior is complex and context-dependent. Consider again the problem of discrimination in the labor market. In this context, research shows that implicit wrongdoing is prevalent, as discrimination is often “structural” or implicit.⁶⁸ That is, even people who do not have an explicit preference for employees from one race or ethnicity over employees of another race or ethnicity, will very often unconsciously discriminate and screen CVs in a way that is consistent with a discriminatory preference.⁶⁹ Behavioral research shows that discrimination often originates from biases that will help people justify their actions in order to preserve an ethical self-image.⁷⁰ That is, the employer, when screening CVs, will easily (and typically) convince herself that she is choosing people of a specific ethnicity simply because those people happen to be more qualified, and not because she holds a discriminatory preference.⁷¹ Only rarely, will a person that considers herself a “good person” openly admit, even to herself, that she is driven by an immoral racist motivation.⁷² This means that discriminatory outcomes are commonly created through subconscious habits and implicit judgements, and not necessarily through an explicit or deliberative preference for one group of employees over another.

Similar claims can be made in the context of corruption. Research shows that people find it difficult to identify their own conflicts of interests.⁷³ For

Apples, Bad Cases, and Bad Barrels: Meta-Analytic Evidence About Sources of Unethical Decisions at Work, 95 J. APPLIED PSYCHOL. 1 (2010).

67 FELDMAN, *supra* note 11; Feldman & Kaplan, *supra* note 60.

68 FELDMAN, *supra* note 11, at ch. 10; Bertrand, Chugh & Mullainathan, *supra* note 63.

69 Krieger, *supra* note 64, at 1164; Krieger & Fiske, *supra* note 64, at 1027.

70 Feldman & Kaplan, *supra* note 60.

71 Don A. Moore & George Loewenstein, *Self-Interest, Automaticity, and the Psychology of Conflict of Interest*, 17 SOC. JUST. RES. 189, 189 (2004).

72 Chugh, Bazerman & Banaji, *supra* note 61; Pronin, Gilovich & Ross, *supra* note 19, at 781-82.

73 Chugh, Bazerman & Banaji, *supra* note 61.

example, a politician can easily fail to recognize that a relationship with a donor is affecting her policies.⁷⁴ When this is the case, even if the politician has a preference against corruption, she might nevertheless end up behaving in a corrupt way; this is not because she prefers to be corrupt, but because she fails to understand the forces that have affected her own judgment. Therefore, also in the context of corruption, the effectiveness of preference change is questionable.

Similar conclusions pertain also in the context of “ordinary unethicity,”⁷⁵ or day-to-day unethical behaviors that are easily perceived as effectively harmless. Ordinary unethicity can be seen in practices such as minor employee theft, or “wardrobing”: buying an item, using it, and then returning it for a full refund.⁷⁶ In such cases, people are less likely to deliberate, and “smoking guns” are rarely present. Such cases are also of lesser interest to traditional enforcement bodies and usually leave sufficient ethical grey areas⁷⁷ for people to be able to come up with various justifications for their wrongdoing.⁷⁸ Change of explicit preferences therefore seems less effective also in this broad category of wrongdoing. More generally, the prevalence of implicit wrongdoing means that preference change is not a panacea; further work is required to identify those specific contexts in which it might be beneficial.

B. The Persistence of Implicit Attitudes

If indeed wrongdoing often originates from implicit attitudes and habits rather than conscious preferences, this presents a significant challenge to the law’s ability to change behaviors. While the logical extension of our argument is that we need to change people’s implicit decision-making mechanisms,

74 *Id.*

75 Francesca Gino, *Understanding Ordinary Unethical Behavior: Why People Who Value Morality Act Immorally*, 3 *CURRENT OP. BEHAV. SCI.* 107, 107-08 (2015).

76 Wardrobing in fact costs retailers \$16 billion a year; Nina Mazar, On Amir & Dan Ariely, *The Dishonesty of Honest People: A Theory of Self-Concept Maintenance*, 45 *J. MARKETING RES.* 633, 633 (2008).

77 Jason Dana, Roberto A. Weber & Jason Xi Kuang, *Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness*, 33 *ECON. THEORY* 67 (2007).

78 Shaul Shalvi et al., *Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior*, 115 *ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES* 181 (2011); Mazar, Amir & Ariely, *supra* note 76, at 633 (offering the theory of self-concept maintenance, according to which “people behave dishonestly enough to profit but honestly enough to delude themselves of their own integrity”).

current research suggests that to be a major challenge.⁷⁹ The reason for this is that implicit attitudes are often much more stable, and much more difficult to change, than explicit attitudes.⁸⁰ That is, it might be possible to convince an employer that hiring women is just as effective (if not more so) than hiring man, simply by presenting data regarding women's high work performance and work ethics.⁸¹ In the language of the preference-change literature, we would say that if an employer has a discriminatory preference against women, it is easy enough to change that preference by presenting information that disproves it and exposes it as based on erroneous factual assumptions.⁸² However, it might be much more difficult to change the employer's *implicit* discriminatory hiring patterns.⁸³ Since these patterns are mostly driven by non-deliberative and semi-deliberative decision-making processes, they may stay stable even after the employers' explicit preference is changed when they learn new information.⁸⁴

The stability of implicit wrongdoing relates to the biased nature of ethical deliberations. People tend to justify their misconduct, and to interpret their choices as ethical. If preferences are "improved," people's increased moral sensitivity might simply act to strengthen those mechanisms responsible for the self-denial of wrongdoing. Thus, if employers are convinced that discrimination is morally wrong, this will activate those cognitive biases that help them maintain a positive self-image, in order to help them avoid the dissonance of any case in which their behavior might have been discriminatory.⁸⁵ Many biases can easily bypass legal instruments.⁸⁶ For instance, when the legal norm is unclear, or when it is unclear what action best fits the requirement of the law, psychological processes such as motivated reasoning⁸⁷ and self-

79 William A. Cunningham, Kristopher J. Preacher & Mahzarin R. Banaji, *Implicit Attitude Measures: Consistency, Stability, and Convergent Validity*, 12 PSYCHOL. SCI. 163, 167 (2001).

80 *Id.*

81 Porat, *supra* note 6.

82 *Id.* at 235-36.

83 Bertram Gawronski & Fritz Strack, *On the Propositional Nature of Cognitive Consistency: Dissonance Changes Explicit, but not Implicit Attitudes*, 40 J. EXPERIMENTAL SOC. PSYCHOL. 535 (2004).

84 Cunningham, Preacher & Banaji, *supra* note 79, at 167.

85 Mazar, Amir & Ariely, *supra* note 76, at 633.

86 Yuval Feldman & Yotam Kaplan, *Big Data and Bounded Ethicality*, 29 CORNELL J.L. & PUB. POL'Y 39 (2019); Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480, 480 (1990).

87 Kunda, *supra* note 86 ("[t]here is considerable evidence that people are more likely to arrive at conclusions that they want to arrive at, but their ability to do

deception⁸⁸ will allow people to promote their self-interest while maintaining the perception that they are behaving in line with their ethical preferences and legal obligations.

To overcome such difficulties, the law would need to recognize ex-ante what justifications people might implicitly use in order to excuse divergence from their newfound and improved preferences. Self-interest is often more intuitive than duty, and can change the way people interpret dilemmas and situations.⁸⁹ Therefore, in order for the law to affect the way people behave, it must also address self-driven interpretations of reality. These are significant challenges and it is unclear whether the law can accomplish such tasks. Accounting for the complexity of cognitive processes might therefore lead us to question whether the effort to use the law to change people's preferences is worthwhile.

C. Situational Wrongdoing & Ethical Blind Spots

Another challenge to the preference-change endeavor relates to the situational nature of wrongdoing. Research in behavioral ethics shows that wrongdoing is often situational, meaning that in many cases wrongdoing is best predicted based on situational, rather than personal, factors.⁹⁰ In some situations, to which we refer as ethical blind spots,⁹¹ a great majority of individuals, including “good people” who value morality, will tend to act unethically.⁹² For example, such ethical blind spots are often characterized by the existence of ambiguous legal or ethical norms,⁹³ harms to unidentified victims,⁹⁴ or joint decision-

so is constrained by their ability to construct seemingly reasonable justifications for these conclusions.”).

88 FELDMAN, *supra* note 11.

89 Moore & Loewenstein, *supra* note 71; Steven J. Spencer et al., *Automatic Activation of Stereotypes: The Role of Self-Image Threat*, 24 PERSONALITY & SOC. PSYCHOL. BULL. 1139 (1998).

90 DAN ARIELY & SIMON JONES, THE (HONEST) TRUTH ABOUT DISHONESTY: HOW WE LIE TO EVERYONE, ESPECIALLY OURSELVES 104 (2012).

91 BAZERMAN & TENBRUNSEL, *supra* note 22, at 1-3.

92 FELDMAN, *supra* note 11, at 1.

93 Yuval Feldman & Henry E. Smith, *Behavioral Equity*, 170 J. INST. & THEORETICAL ECON. 137, 141 (2014). For empirical evidence for the effect of legal ambiguity, see Constantine Boussalis, Yuval Feldman & Henry E. Smith, *Experimental Analysis of the Effect of Standards on Compliance and Performance*, 12 REG. & GOVERNANCE 277, 288 (2018).

94 Amitai Amir, Tehila Kogut & Yoella Bereby-Meyer, *Careful Cheating: People Cheat Groups Rather Than Individuals*, 7 FRONTIERS PSYCHOL. 371, 371 (2016).

making;⁹⁵ similarly, wrongdoing has been shown to be more common when perpetrators work to promote the interest of others and not just their own interests, or when the profits from wrongdoing are shared.

If indeed the antecedents of wrongdoing are often more situational than personal, this presents a significant challenge to the preference-change project. That is, if it is situational factors more than people's preferences that lead to wrongdoing, changing preferences is not likely to be an effective way to curb wrongdoing. This is especially true in the cases where research suggests that situational factors can create ethical traps where the majority or at least a substantial proportion of the population is likely to behave unethically. In such situations, and especially with regard to milder forms of unethical behavior, it seems less reasonable to believe that wrongdoing originates from people's ethical preferences, and therefore changing those preferences is not likely to be an effective way to curb misconduct.

Of course, the analysis presented above is not meant to suggest that personal factors are never important in predicting unethicality or that situational factors are the only significant antecedent of unethical conduct. Thus, different personality scales, developed by researchers, can be used in an attempt to predict unethicality based on personal factors. For instance, Shane Frederick's CRT (cognitive reflective test) can be used to predict the likelihood of implicit wrongdoing.⁹⁶ Similarly, scales measuring the propensity to morally disengage,⁹⁷

95 Ori Weisel & Shaul Shalvi, *The Collaborative Roots of Corruption*, 112 PROC. NAT'L ACAD. SCI. 10651 (2015); Yuval Feldman, Adi Libson & Gideon Parchomovsky, *Corporate Law for Good People*, 115 NW. U. L. REV. 1125 (2021).

96 Shane Frederick, *Cognitive Reflection and Decision Making*, 19 J. ECON. PERSP. 25 (2005); Maggie E. Toplak, Richard F. West & Keith E. Stanovich, *The Cognitive Reflection Test As A Predictor Of Performance On Heuristics-And-Biases Tasks*, 39 MEMORY & COGNITION 1275 (2011). This scale rates people based on the likelihood that they will use system 2 thinking to overcome system 1. The main focus of the studies done on the basis of this scale is related to the findings connecting people's CRT grades and various other behavioral measures; Joseph M. Paxton, Leo Ungar & Joshua D. Greene, *Reflection and Reasoning in Moral Judgment*, 36 COGNITIVE SCI. 163 (2012).

97 Lisa L. Shu, Francesca Gino & Max H. Bazerman, *Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting*, 37 PERSONALITY & SOC. PSYCHOL. BULL. 330, 344 (2011); Albert Bandura, *Moral Disengagement in the Perpetration of Inhumanities*, 3 PERSONALITY & SOC. PSYCHOL. REV. 193, 204 (1999).

moral identity,⁹⁸ or moral firmness⁹⁹ are also highly relevant for predicting unethicity based on personal factors. Our point here is not to deny the importance of all these tools, but to highlight the fact that situational factors, alongside personal ones, are strong predictors of unethical behavior.¹⁰⁰ This in itself suffices to present a significant challenge to the current literature on preference change.

III. A BEHAVIORAL ETHICS APPROACH TO PREFERENCE CHANGE

In light of the critique outlined in Part II, this Part explores the possibility of using the law to change ethical awareness. More specifically, we focus on the possibility of using the law to produce a long-term change in people's internal ethical commitments, in a way that will also be expressed in their actual conduct. This task proves to be a challenge for the law, and perhaps beyond the capabilities of a modern legal system in a liberal state. The reason for this is that a true improvement in people's ethical awareness would require a specific type of legal intervention, which modern legal systems seem to purposely refrain from. We therefore propose that the main use of the law to change preferences may be indirect, and require the support of other social institutions.

A. From Preferences to Awareness

The critique we offer in Part II points to a potential correction of the preference-change endeavor. Namely, if wrongdoing is often implicit rather than explicit and can be attributed to situational rather than personal antecedents, it might be more impactful to target people's levels of awareness instead of their explicit preferences, wishes and values. Thus, the law might seek to make people more reflective and deliberative when they make decisions in ethically sensitive contexts. Along similar lines, it may be preferable to try to improve people's implicit attitudes and habits. All this, however, is not easy to do. This

98 Karl Aquino et al., *Testing A Social-Cognitive Model of Moral Behavior: The Interactive Influence of Situations and Moral Identity Centrality*, 97 J. PERSONALITY & SOC. PSYCHOL. 123, 138-39 (2009) (showing that people's likelihood of doing harm, even implicitly, could be different across different situations based on their level of moral identity).

99 Shaul Shalvi & David Leiser, *Moral Firmness*, 93 J. ECON. BEHAV. & ORG. 400, 400-01 (2013) (connecting people's likelihood of engaging in misconducts with their ability to exploit ambiguity).

100 Aquino et al., *supra* note 98, at 138-39.

analysis calls for a reorientation of the preference-change literature, in light of behavioral ethics findings. Thus, in the case mentioned above of discrimination by the employer, the main goal may not be to change the employer's explicit preference for hiring people of a specific ethnicity, as this explicit preference may not be the source of the problem. Instead, the appropriate solution, re-conceptualizing the idea of "preference," might be to change the employer's implicit decision-making processes. While this might still fit under some concept of "preference change," it is quite different from what the current literature calls for. For instance, the means for changing preferences will not be exposing the employer to new information, but rather making sure that the employer is more deliberative in her decision-making process, through the use of mandatory training programs or ethical reminders. This may require a completely different set of regulatory interventions, depending on the types of cognitive biases and implicit decision-making mechanisms that contribute to the employer's wrongdoing.¹⁰¹

B. Changing Ethical Awareness

Once the black-box concept of "preference" is unpacked, we can recognize that "preferences" are in fact compound elements, which include explicit ethical positions as well as deliberative and non-deliberative choice mechanisms. The question then arises regarding the possibility of changing implicit judgement and improving ethical awareness. Very generally, it seems that people's ethical awareness can be changed through systematic interventions. For example, Max Bazerman and Ann Tenbrunsel have suggested various techniques that might be used in order to improve self-awareness to wrongdoing.¹⁰² One way to do this is to increase people's awareness to their own ethical blindness and to the fact that they might encounter situations in which they will make ethically questionable choices. Similarly, research in behavioral ethics demonstrates the possibility of ethical training, meaning that targeted interventions can change people's implicit attitudes.¹⁰³ Studies demonstrate the possibility of training

101 This analysis is in line with recent proposals to improve behavior through the use of moral emotions such as compassion, guilt or empathy: Hila Keren, *Guilt-Free Markets-Unconscionability, Conscience, and Emotions*, 2016 B.Y.U. L. REV. 427 (2016); Mathieu Cassotti et al., *Positive Emotional Context Eliminates the Framing Effect in Decision-Making*, 12 EMOTION 926 (2012). This line of research suggests that legal actors should improve compliance by evoking or suppressing such emotions, which might better affect decision-making processes.

102 BAZERMAN & TENBRUNSEL, *supra* note 22.

103 Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 PERSONALITY & SOC. PSYCHOL. REV. 242 (2002); Margo J. Monteith et al., *Putting*

people to mitigate the impact of implicit racial biases on their decisions.¹⁰⁴ Some experiments showed improvement to be possible and long lasting,¹⁰⁵ while other studies found only short-term modifications in behavior.¹⁰⁶

The possibility that ethical awareness can be changed and improved is, however, only a first step in terms of the present inquiry. Thus, even if ethical decision-making can generally be improved, we must still consider the role of the law and the state in this context. More specifically, state intervention in such a capacity raises considerable concern, as it is not clear that governments can and should be in the business of bettering their citizens' ethical deliberations. Of course, one could argue that changing decision-making processes is actually less intrusive than changing preferences, as this only helps people make decisions in a more deliberative manner, thus supposedly enhancing their independent will rather than limiting it. We believe, however, that such arguments may downplay the intense level of engagement required to change decision-making processes.

C. Ethical Nudges versus Changing Ethical Preferences

We wish to distinguish between two different modes of state action in the context of improving ethical deliberations. First, the law can be used to improve ethical awareness *locally*. That is, the law can provide targeted interventions, or "ethical nudges," that might guide individuals in making more candid ethical deliberations in specific cases, when those "nudges" are introduced. While we believe this can in many cases be effective, we are hesitant to call this "preference change," as it does not seem that any stable alteration in individuals' attitudes is being discussed here. Second, the law can be used to improve ethical deliberations *globally* (rather than locally). This involves using the law to produce a more permanent improvement in people's ethical capabilities. We tentatively suggest that this form of ethical preference change, which is close to reeducation, might be more problematic.

the Brakes on Prejudice: On the Development and Operation of Cues for Control, 83 J. PERSONALITY & SOC. PSYCHOL. 1029 (2002).

104 Patricia G. Devine et al., *Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention*, 48 J. EXPERIMENTAL SOC. PSYCHOL. 1267 (2012).

105 *Id.*; Curtis E. Phillips et al., *Mind the Gap: Increasing Associations Between the Self and Blacks with Approach Behaviors*, 100 J. PERSONALITY & SOC. PSYCHOL. 197 (2011).

106 Danielle E. Warren, Joseph P. Gaspar & William S. Laufer, *Is Formal Ethics Training Merely Cosmetic? A Study of Ethics Training and Ethical Organizational Culture*, 24 BUS. ETHICS Q. 85 (2014) (finding both short-term and long-term improvement in bank employees' intent to behave ethically).

Elsewhere, we have highlighted the possibility of using ethical nudges in order to improve ethical deliberations and ethical decision-making in specific cases.¹⁰⁷ Ethical nudges are any legal interventions designed to improve awareness and help people avoid the cognitive biases that typically distort their ethical deliberations.¹⁰⁸ We argue that ethical nudges must be targeted and tailored in order to be effective, meaning that they must be operated at specific times, when ethical biases are shown to be operative, in order to alert potential wrongdoers and help them engage in a more candid process of ethical decision-making.¹⁰⁹ Research shows that such nudges, if deployed appropriately, can be effective in improving ethical deliberations.¹¹⁰ For instance, individuals made to sign moral statements or declare their commitment to an ethical code of conduct can be nudged to make more fully deliberative ethical choices.¹¹¹ Such mechanisms can prove useful, for instance, in the discriminating employer case discussed above.

We distinguish such local, targeted, ethical nudges from long-lasting preference change. Ethical nudges are tailored interventions designed to improve ethical deliberations in specific cases and to remove ethical blind spots, and not to improve internal compliance mechanisms in the more general sense of producing individuals that are independently and consistently better able to behave ethically. In other words, ethical nudges, as the term suggests, call for targeted interventions in crucial junctures of ethical decision-making, in order to “nudge” people towards a more ethical choice. Conversely, an ethical preference change would require the legal system to help people become better moral agents on their own, without being “nudged” in specific instances.

This distinction is important. While ethical nudges can be effective legal tools to improve behavior,¹¹² we are more skeptical regarding the possibility of using the law to facilitate a long-lasting improvement in people’s independent ethical awareness. This might be difficult to achieve in most legal and regulatory frameworks, or beyond the current scope and abilities of a modern legal system in a liberal state.

107 Feldman & Kaplan, *supra* note 86.

108 Feldman & Kaplan, *supra* note 60.

109 Feldman & Kaplan, *supra* note 86.

110 FELDMAN, *supra* note 11, at 199.

111 *Id.*; Eyal Peer & Yuval Feldman, *Honesty Pledges for the Behaviorally-based Regulation of Dishonesty*, 28 J. EUR. PUB. POL’Y 761 (2021).

112 FELDMAN, *supra* note 11, at 199; Shu, Gino & Bazerman, *supra* note 97.

D. Should the State Improve Ethicality?

If it is possible, as a practical matter, to change moral awareness and long-term ethical capabilities, a more normative question follows. Namely, is it legitimate to use the law for this purpose, and if so in what way? Behavioral ethics research presents a challenge to the project of changing preferences through use of the law. The reason for this is that the legal mechanisms necessary to implement this change in awareness are more robust than the literature on preference change currently assumes. Since ethical conduct is affected not just by explicit preferences but also (often mainly) by semi-deliberative and non-deliberative cognitive processes, and since these are much more difficult to change, the intensity of the legal intervention required in order to truly change ethical tendencies may be far greater than what scholars currently assume.

A change in ethical preferences, in the sense of improving people's long-term ability to conduct candid ethical deliberations, may be beyond the scope of possibilities for a modern legal system. At least, it does not seem to us that the liberal state is actively pursuing such a goal. In this, modern legal systems differ from earlier legal regimes, which emphasized a much closer connection between law and morality and included mechanisms for improving moral awareness. For example, Jewish law offers multiple concepts that are designed to improve ethical awareness and has an explicit goal of improving moral deliberations and helping practitioners overcome negative habits and biased ethical thinking. Thus, religious symbols and items of clothing, as well as the habit of thinking and picturing the concept of the deity, are meant to help the practitioner avoid temptations and improve ethical conduct.¹¹³ These are central themes of religious practice, which emphasizes strong behavioral elements designed to constantly remind people of their religious and ethical duties and to create habits of ethical awareness through the use of rituals and repetition,¹¹⁴ ethical symbolism and a strong ethos of compliance.¹¹⁵ For instance, prayer beads — rosaries in Christianity and *misbahahs* in Islam¹¹⁶ — are used as physical reminders of religious duties, constantly held, touched and looked at. The fact that ancient legal systems devoted this sort of energy to maintaining and improving ethical commitments indicates both

113 Sreedhari D. Desai & Maryam Kouchaki, *Moral Symbols: A Necklace of Garlic against Unethical Requests*, 60 ACAD. MGMT. J. 7 (2017).

114 Jacob Milgrom, *The Biblical Diet Laws as an Ethical System: Food and Faith*, 17 INTERPRETATION 288 (1963).

115 Leonard Weller et al., *Religiosity and Authoritarianism*, 95 J. SOC. PSYCHOL. 11 (1975).

116 LOIS SHERR DUBIN, *THE HISTORY OF BEADS: FROM 100,000 B.C. TO THE PRESENT* 79-92 (2009).

the crucial importance of this activity for improving behavior, as well as the deep insight these early systems possessed into human psychology. These systems, however, critically differ from modern liberal legal systems, which seem to be deliberately uninterested in this type of endeavor.

In the name of moral neutrality,¹¹⁷ liberal states are hesitant to go into the business of creating ethical symbols and ceremonies for their citizens. States and legal institutions typically refrain from engaging directly with people's moral habits, and, operating at a safe distance from people's everyday lives, seem ill equipped to do so. Of course, the modern state could, in theory, choose to engage more closely with people's moral habits; yet this may involve a complete overhaul of the way the state interacts with its citizens. State efforts to improve moral capabilities (as opposed to ethical deliberation in specific cases) entails constant and deep engagement with the moral habits of the populace, akin to those practiced in religious legal systems. Such endeavors, if pursued, therefore require fundamental changes to the levels of freedom the state seeks to allow its citizens, and to the central ethos of moral neutrality.¹¹⁸

E. The Indirect Role of the Law

Our analysis up to this point suggests a limited role for the state in operating directly to affect a long-lasting improvement in people's ethical capabilities. As explained above, the connection between preferences and behavior is much more tentative than is currently assumed in the law and economics literature on preference change. This means that a true change in outcomes would require an intensive change in implicit habits, a change that the law might be ill-suited to generate independently. It therefore may be the case that the more appropriate use of the law is to change preferences indirectly, by changing institutions and social norms that, in turn, can more successfully affect people's internal decision-making mechanisms.

Organizations, such as schools and workplaces, can be more effective than the law and the state in inducing ethical awareness and in changing people's implicit attitudes. Such organizations offer intense social frameworks, in which people spend a significant amount of time in close proximity to guiding rules and supervisory authorities. Such organizations are also allowed to engage

117 Edward A. Harris, *Fighting Philosophical Anarchism with Fairness: The Moral Claims of Law in the Liberal State*, 91 COLUM. L. REV. 919, 942 (1991); Peter De Marneffe, *Liberalism, Liberty, and Neutrality*, 19 PHIL. & PUB. AFF. 253, 253-274 (1999); Will Kymlicka, *Liberal Individualism and Liberal Neutrality*, 99 ETHICS 883 (1989).

118 Marneffe, *supra* note 117; Kymlicka, *supra* note 117.

in practices of habit-formation that we might not tolerate if states were to engage in them. This means that the law can change ethical preferences more effectively not by trying to engage with people's awareness directly, but by creating requirements that change the relevant organizations and incentivize those organizations, in turn, to engage directly with people's preferences and awareness. Thus, the law might sanction organizations when they discriminate, in the hope that they will then act to improve ethical awareness among their decision-makers. In such cases, the law acts using simple traditional deterrence mechanisms, without any direct attention to considerations relating to people's moral awareness; it simply leaves these responsibilities to more suitable institutional actors, operating in closer proximity with people's day-to-day decision-making processes.

This argument helps bring together the literature on preference change with the research on social enforcement.¹¹⁹ The law can generate internal motivation when operating in tandem with social norms as an enforcement mechanism.¹²⁰ Social norms are often vague, or slow to form and change;¹²¹ the law can signal the moral significance of a particular behavior, and then rely on social enforcement mechanisms to kick in and improve compliance.¹²² As social networks are more intimate than legal ones, they are more likely to be able to affect awareness and engage successfully with people's ethical decision-making.

These conclusions bring our analysis back to Cooter's original argument, highlighting the need to use families and close-knit communities in order to improve moral commitments,¹²³ though we reach this conclusion for different reasons. Cooter argues that people can verify the true moral character of those close to them; something legal institutions are unable to do. Therefore, Cooter argues, rational decision-makers will have an incentive to change their moral character in order to improve their standing with their peers.¹²⁴ Conversely, we emphasize ethical biases and the possibility that ethical decision-making

119 Ernst Fehr & Simon Gächter, *Fairness and Retaliation: The Economics of Reciprocity*, 14 J. ECON. PERSP. 159 (2000); Ernst Fehr & Simon Gächter, *Altruistic Punishment in Humans*, 415 NATURE 137 (2002); Ernst Fehr, Urs Fischbacher & Simon Gächter, *Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms*, 13 HUM. NATURE 1 (2002).

120 Richard H. McAdams, *The Origin, Development and Regulation of Norms*, 96 MICH. L. REV. 338 (1997).

121 Cass R. Sunstein, *On the Expressive Function of Law*, 144 U. PA. L. REV. 2021 (1996).

122 Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903 (1996).

123 Cooter, *supra* note 2, at 1581.

124 *Id.*

greatly depends on levels of moral awareness. Therefore, improving ethical decision-making would require deep engagement with people's habits and subconscious attitudes. This type of involvement in people's lives is more likely to be attainable through close-knit social communities rather than through formal legal institutions.

CONCLUSION

The classic economic model assumes that preferences are stable over time and internally consistent. The law and economics literature on preference change relaxes the assumption regarding the consistency of preferences over time. This is crucially important for law and economics scholars, as it offers the possibility of using the law to improve behavior by changing harmful or undesirable preferences. This literature still retains the assumption regarding the internal consistency of preferences, and assumes also that preferences directly translate into actions and choices. We challenge this assumption by introducing insights from behavioral ethics research into the literature on preference change. In particular, we show that behavior, especially in cases of ethical dilemmas, depends on complex cognitive processes, and not on a direct translation from explicit preferences to conduct. Ethical decision-making is often biased, especially when legal and ethical standards are ambiguous and no "smoking guns" are present. Thus, people often ignore their own wrongdoing, or find ways to justify it in order to avoid dissonance and maintain a positive self-image as moral individuals. This means that changing explicit preferences will not necessarily improve behavior, since wrongdoing often originates from implicit decision-making processes. We further show that in order to improve compliance, the law needs to change ethical awareness and improve people's moral deliberations rather than simply change "preferences." We suggest that this task of improving ethical awareness may be best accomplished through legal actions that are supported or supplemented by nonlegal societal mechanisms.