

Columbia Law School

## Scholarship Archive

---

Faculty Scholarship

Faculty Publications

---

2016

# When Extrinsic Incentives Displace Intrinsic Motivation: Designing Legal Carrots and Sticks to Confront the Challenge of Motivational Crowding-Out

Kristen Underhill  
*Columbia Law School*

Follow this and additional works at: [https://scholarship.law.columbia.edu/faculty\\_scholarship](https://scholarship.law.columbia.edu/faculty_scholarship)



Part of the [Banking and Finance Law Commons](#), [Environmental Law Commons](#), [Jurisprudence Commons](#), [Law and Economics Commons](#), and the [Law and Society Commons](#)

---

### Recommended Citation

Kristen Underhill, *When Extrinsic Incentives Displace Intrinsic Motivation: Designing Legal Carrots and Sticks to Confront the Challenge of Motivational Crowding-Out*, 33 YALE J. ON REG. 213 (2016).

Available at: [https://scholarship.law.columbia.edu/faculty\\_scholarship/2799](https://scholarship.law.columbia.edu/faculty_scholarship/2799)

This Article is brought to you for free and open access by the Faculty Publications at Scholarship Archive. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarship Archive. For more information, please contact [scholarshiparchive@law.columbia.edu](mailto:scholarshiparchive@law.columbia.edu).

# When Extrinsic Incentives Displace Intrinsic Motivation: Designing Legal Carrots and Sticks to Confront the Challenge of Motivational Crowding-Out

Kristen Underhill<sup>†</sup>

*The rise of “nudges” has inspired countless efforts to encourage individual choices that maximize personal and collective welfare, with a preference for less restrictive tools such as setting default options or reordering choice sets. As part of this trend, there has been renewed interest in the behavioral impacts of incentives—namely, rewards or penalties for shaping individual choices, including but not limited to financial incentives. Explicit incentives are pervasive in the law, including carrots offered by governments (for example, tax deductions for charitable contributions, rebates for recycling, sentence reductions for prisoners who complete drug rehabilitation programs, and incentives for criminal informants) and statutes or regulations that govern incentives offered by private parties (for example, workplace wellness programs, compensation for blood and organ donation, and pay-for-performance in executive compensation). But despite the intuitive appeal of incentives, legal commentators have expressed increasing alarm about a potential drawback: research in behavioral economics and psychology has come to show many ways in which the use of carrots and sticks may displace other motivations for good behavior, such as altruism, civic duty, or professionalism.*

*In legal scholarship, prevailing views of motivational crowding-out—the process by which incentives can interfere with “intrinsic” motivations for behavior—suggest that this phenomenon is an irremediable response to incentive-based policies. This Article examines a large but neglected body of empirical and theoretical literature on motivational crowding-out to show that these beliefs may be misguided. Motivational crowding-out is in fact a catch-all term for a diverse set of cognitive and behavioral processes that range from long-term changes in preferences, to the impairment of self-determination, to a complex set of signals that incentives can send to people about their abilities, social environment, values, and employers. Far from being inevitable, motivational crowding-out is responsive to changes in the way we design*

---

<sup>†</sup> Fellow in Law & Health, Yale Law School. J.D., Yale Law School (2011); D.Phil., Oxford University (2007). I am grateful to Ian Ayres, Abbe Gluck, David Grewal, Amy Kapczynski, Daniel Markovits, William M. Sage, participants in the Petrie-Flom 2014 Annual Conference, and attendees at the American Society of Law, Medicine & Ethics 2014 Health Law Professors Conference for discussions about motivational crowding-out effects and feedback on earlier drafts of this paper. All errors herein are my own.

*incentive-based policies. That is, once we understand the mechanisms of crowding-out, we can modify the incentive architecture to either minimize or amplify crowding-out effects. Remedies, however, must be tailored to the diverse causes of crowding-out, and the law has not yet recognized this problem. In light of deep anxieties about motivational crowding-out throughout the law, this Article proposes a taxonomy of crowding-out processes and introduces “incentive architecture:” the deliberate structuring of incentives to address crowding-out effects.*

Introduction.....	215
I. Motivational Crowding-Out .....	218
A. Defining Terms: Intrinsic Motivation, Extrinsic Motivation, and Incentives .....	219
B. Motivational Crowding-Out and the Separability Assumption.....	224
C. Crowding-Out Concerns and Incentive-Based Policies .....	229
1. Altruism Crowding-Out .....	230
2. Professionalism Crowding-Out.....	231
3. Civic Duty Crowding-Out.....	231
II. Heterogeneity in Motivational Crowding-Out Processes.....	232
A. Signaling Processes .....	233
1. Principal Signaling to Agents .....	233
2. Agent Signaling to Observers: Image-Spoiling .....	243
3. Agent Self-Signaling: Attribution and Self-Image-Spoiling....	245
B. Overjustification and Impaired Self-Determination.....	247
C. Endogenous Preference Adaptation .....	248
D. Learned Helplessness .....	251
III. Designing Incentive Architecture to Accommodate Motivational Crowding-Out .....	252
A. Program Development and Implementation .....	254
B. Framing Decisions.....	257
C. Incentive Definitions .....	259
1. Incentive Type .....	259
2. Incentive Size.....	261
3. Incentive Uncertainty and Specificity.....	262
4. Incentive Fungibility.....	263
5. Incentive Recipient .....	264
D. Contingency Decisions .....	266
E. Bundling Incentives with Information.....	270
F. Publicity Choices .....	272
G. Managing Heterogeneity.....	273
H. Discontinuing Incentive Programs .....	277
Conclusion .....	278

## Introduction

Imagine that you have newly taken charge of a local government, say as mayor of a large city, and you must confront a range of social, environmental, and health problems. Rates of recycling are waning, pollution is high, residents are evading local taxes, public schools report poor student attendance, obesity and smoking are up, and low rates of flu vaccination pose a public health danger. You may also believe that there is room for improvement in the productivity of city employees, and you would love to secure support for building a new wastewater treatment plant. Despite extensive public education campaigns by your predecessors, these issues have seen little to no improvement in prior years. You are, of course, wary of imposing outright mandates for many of these problems: even if the city council supported new ordinances addressing such behaviors (for example, mandatory recycling or flu vaccination), the backlash would be politically costly, and you would be assured an expensive and potentially unsuccessful courtroom battle. So instead, you design a sophisticated and progressive system of incentives (blessed, perhaps, by deep city coffers). You will pay rebates for recycling; offer free public transit fares on days with elevated air pollution; provide tax credits for prompt tax payment; offer students cash for perfect attendance; make small payments to families who receive flu vaccines, lose weight, or stop smoking; offer merit-based bonuses to city employees; and promise additional tax credits to the neighborhood where you plan to build the plant. Your colleagues in city council agree, and you implement your incentive programs, eagerly anticipating improvements.

But what if your plans backfire? What if, contrary to your expectations, your incentives actually *increase* tax noncompliance, smoking, weight gain, truancy, and car use, while *decreasing* recycling, flu vaccination, civil servant productivity, and support for your development project? Or what if you see temporary improvements—but when budget cuts put an end to your carrots, the problems return more acute than they were before? Worse yet, what if other good behaviors that were left unincentivized (for example, voting and energy conservation) now begin to decline? All of these effects are contemplated impacts of the displacement of “intrinsic” motivations by “extrinsic” carrots or sticks, a phenomenon known as motivational crowding-out.<sup>1</sup>

Each of the foregoing examples is drawn from the vast literature on motivational crowding-out, which considers how incentives may have counterintuitive and counterproductive effects on human behavior.<sup>2</sup> In many

---

1. See Yuval Feldman & Orly Lobel, *The Incentives Matrix: The Comparative Effectiveness of Rewards, Liabilities, Duties, and Protections for Reporting Illegality*, 88 TEX. L. REV. 1151, 1178-79 (2010).

2. For some of the most recent and comprehensive overviews of crowding-out to date, see BRUNO S. FREY, *NOT JUST FOR THE MONEY* (1997); Yochai Benkler, *Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production*, 114 YALE L.J. 273 (2004);

ways, law has entered an era of nudges, deprioritizing mandates and command-and-control regulation in favor of choice architecture and subtler ways of achieving behavioral goals. Rewards are widely viewed as a less forceful means of influencing behavior, compared to mandates or penalties.<sup>3</sup> Given this shift, the potential drawbacks of carrots and sticks have gained increasing attention, with concerns about motivational crowding-out taking center stage.<sup>4</sup> Crowding-out has animated commentary and development regarding many incentive-based legal rules, including whether paying organ donors<sup>5</sup> or Good Samaritans<sup>6</sup> crowds out altruism, whether pay-for-performance crowds out professionalism and intrinsic motives for good work,<sup>7</sup> whether incentives for whistleblowers crowd out their moral motivations,<sup>8</sup> whether complete contracts—those that cause agents to internalize every benefit and every cost of their actions—crowd out mutual trust,<sup>9</sup> whether compensating jurors crowds out civic duty,<sup>10</sup> and whether heightened tax penalties may reduce moral or social reasons for compliance.<sup>11</sup>

The concept of motivational crowding-out derives from research conducted in psychology and economics in the 1970s, beginning in seminal fields such as payment for blood donation and educational achievement, and expanding over the decades to encompass a broad range of incentive programs. Legal scholars have discussed motivational crowding-out in every area that attempts to shape human behavior, in fields as diverse as health law, employment law, tax, torts, contracts, criminal law, intellectual and real property law, and education. Across these and other areas, crowding-out has met with a variety of receptions, ranging

---

Samuel Bowles & Sandra Polanía-Reyes, *Economic Incentives and Social Preferences: Substitutes or Complements?*, 50 J. ECON. LITERATURE 368 (2012); 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); Feldman & Lobel, *supra* note 1; Bruno S. Frey & Reto Jegen, *Motivation Crowding Theory: A Survey of Empirical Evidence*, 5 J. ECON. SURVS. 589 (2001); Marianne Promberger & Theresa M. Marteau, *When Do Financial Incentives Reduce Intrinsic Motivation? Comparing Behaviors Studied in Psychological and Economic Literatures*, 32 HEALTH PSYCHOL. 950 (2013).

3. See, e.g., RUTH W. GRANT, *STRINGS ATTACHED: UNTANGLING THE ETHICS OF INCENTIVES* 41 (2012).

4. See, e.g., Emad H. Atiq, Note, *Why Motives Matter: Reframing the Crowding Out Effect of Legal Incentives*, 123 YALE L.J. 1070, 1081 (2014) (describing crowding-out concerns in conjunction with several incentive-based policies).

5. See David Horton, *Indescendibility*, 102 CALIF. L. REV. 543, 574-75 (2014).

6. See William M. Landes & Richard A. Posner, *Salvors, Finders, Good Samaritans, and Other Rescuers: An Economic Study of Law and Altruism*, 7 J. LEGAL STUD. 83, 98-100 (1978).

7. See Lynn A. Stout, *Killing Conscience: The Unintended Behavioral Consequences of "Pay for Performance"*, 39 J. CORP. L. 525, 552-53 (2014).

8. See David Freeman Engstrom, *Whither Whistleblowing? Bounty Regimes, Regulatory Context, and the Challenge of Optimal Design*, 15 THEORETICAL INQUIRIES L. 605, 616 (2014); Feldman & Lobel, *supra* note 1.

9. See Iris Bohnet et al., *More Order with Less Law: On Contract Enforcement, Trust, and Crowding*, 95 AM. POL. SCI. REV. 131 (2001).

10. See Evan R. Seamone, *A Refreshing Jury Cola: Fulfilling the Duty to Compensate Jurors Adequately*, 5 N.Y.U. J. LEGIS. & PUB. POL'Y 289, 379 (2002).

11. See, e.g., Yair Listokin & David M. Schizer, *I Like to Pay Taxes: Taxpayer Support for Government Spending and the Efficiency of the Tax System*, 66 TAX L. REV. 179, 203-04 (2013).

from outright dismissal, to cautious concern, to acceptance as a justification for barring incentives altogether. Legal scholarship is increasingly committed to many views of this behavioral phenomenon, including the view that it does not exist. But in order to engage productively with crowding-out—to dismiss it, to evaluate it, or to embrace it—we need a careful and nuanced account of the phenomenon and its manifestations in the law. In the absence of this account, concerns about crowding-out have been “muddled.”<sup>12</sup> the term “crowding-out” has been used to encompass a range of different effects, and there has been no systematic inquiry into either crowding-out processes or the incentive designs that may remedy these problems.

Unlike prior legal scholarship, which focuses on *whether* crowding-out occurs, I answer the more complex questions of what crowding out *is*, and whether it might be remediable. This is the first legal scholarship to interrogate the different mechanisms by which motivational crowding-out occurs. My examination of the empirical literature in psychology, economics, and public health identifies four separate, empirically-supported mechanisms by which incentives (i.e., carrots, sticks, or mandates) may diminish intrinsic motivations such as altruism, professionalism, or civic duty. Evidence for these processes includes field and laboratory experiments that impose rewards or penalties for performance of a task, then measure indicators of intrinsic motivation (for example, overall behavior, free time spent on a task, willingness to spend resources, or self-reported motivation). Where incentives backfire or produce relative declines in other indicators of intrinsic motivation, the findings are consistent with crowding-out. I will discuss four major crowding-out processes: signaling effects, impaired self-determination, endogenous preference changes, and “learned helplessness.” By overlooking heterogeneity across crowding-out mechanisms, legal scholars have allowed crowding-out to become a mysterious malaise of incentive-based legal rules, and we continue to neglect opportunities to evaluate and alleviate the potentially counterproductive effects of incentive schemes.

This is also the first Article to link crowding-out to the way incentives are designed—what I will call “incentive architecture.” I introduce the term as distinct from “choice architecture” because incentives differ from nudges generally. Nudges influence decisions without changing the character of those choices, and as such, they exclude subsidies, taxes, and fines.<sup>13</sup> Classic examples of nudges include reordering options or setting default choices. Because default rules are unavoidable, choice architecture—namely, the setting of entitlements,

---

12. Mark Schlesinger, *Crowding Out: Multiple Manifestations, Muddled Meanings*, 37 J. HEALTH POL. POL'Y & L. 851 (2012).

13. See Cass Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REG. 413, 417 (2015) (defining “nudges” to exclude interventions that “impose significant material incentives,” including taxes and fines). Of course, the effort involved in choosing against a default rule, such as opting *in* to organ donation or *out* of default savings plans, is a cost. But choice architecture scholars would not consider this to be a “material incentive.”

orders, and defaults<sup>14</sup>—is inevitable. But unlike nudges, incentives alter the costs and benefits of available options. They impose separable costs and benefits on agents, apart from the consequences of the choice itself. For example, giving rebates for recycling bottles changes the decision to recycle; now recyclers receive money, which is separable from any other benefits they may accrue from the choice. By adding separable rewards or penalties, incentives interfere with the conscious decision-making process, rather than capitalizing on inertia or cognitive biases. Automatically enrolling employees in a savings plan is a choice architecture decision; matching their savings up to a 5% contribution is an incentive architecture decision. Unlike choice architecture, the use of incentives is not inevitable.<sup>15</sup> Where crowding-out takes place, tailoring incentives to the specific mechanisms driving the behavior is crucial, because a remedy for one mechanism may exacerbate another mechanism (for example, increasing an incentive may reduce crowding-out if agents view incentives as a sign of task value, but it could worsen crowding-out due to framing effects). Although tailored solutions may be difficult to implement, crowding-out is a diverse phenomenon that requires a nuanced set of solutions.

My analysis proceeds in three Parts. Part I introduces the general phenomenon of motivational crowding-out and the empirical evidence supporting these effects. In Part II, I provide an in-depth look at the different processes that give rise to crowding-out. Each is supported by evidence from economics, psychology, and empirical legal research, and each presents different implications for intervention. In Part III, I identify how the design of incentives may influence motivational crowding-out effects. This Part serves as an introduction to incentive architecture, and it considers a series of decisions that governments and private principals must make when structuring and implementing incentive plans. Throughout this Part, I emphasize that remedies for crowding-out must be tailored to the mechanisms by which crowding-out occurs.

## I. Motivational Crowding-Out

Motivational crowding-out looms large in discussions of law and public policy design, particularly in the last decade of behavioral law and economics research. This Part will demarcate intrinsic, extrinsic, and reputational motivations; describe motivational crowding-out effects; and illustrate several areas in private law and public policy where crowding-out effects are relevant.<sup>16</sup>

---

14. *Id.*

15. An alternate view here may be to consider the *lack* of an incentive to be an incentive architecture decision; if we take this perspective, incentive architecture is also inevitable. The key point here, however, is that while there must always be defaults and entitlements, separable incentives are deliberate and not inevitable—the default is always the absence of an incentive.

16. This Article is primarily concerned with motivational crowding-out. But regulators, legal scholars, and courts have also used crowding-out terminology in other circumstances, generally

Throughout the Article, “agents” are individuals engaging in a task of interest, while “principals” are governments, employers, or others attempting to influence agents’ choices using incentives.<sup>17</sup> “Motivation” will denote a reason for an individual choice, considered as both a categorical variable (multiple reasons may exist) and a continuous variable (the strength of an individual reason for behavior is a quantity of motivation). Multiple motivations can drive the same behavior—for example, the decision to pay taxes may reflect not only a fear of being caught evading, but also a genuine dedication to civic duty, the desire to conform with social norms that discourage evasion, and intrinsic preferences for obeying the law. The relative balance of these motivations, however, is a key concern for crowding-out theory.

This Part will begin by presenting the difference between intrinsic and extrinsic motivation. The central fear of crowding-out is that the use of incentives will strengthen extrinsic motivations at the expense of intrinsic motivations, so defining these categories is an essential foundation. The second section of this Part will define motivational crowding-out and discuss a brief history of research on the interaction between incentives and motivation. I will then summarize the empirical evidence that supports this phenomenon. I will conclude this Part with several illustrations of how motivational crowding-out is relevant to the law, providing examples from health law, employment, torts, and other areas.

#### *A. Defining Terms: Intrinsic Motivation, Extrinsic Motivation, and Incentives*

Crowding-out theory rests on the distinction between two types of motivation: intrinsic and extrinsic. Commentators differ in the extent to which they include reputational or “image” motivations as a separate third category; I will do so below.

Conventional economics considers material self-interest to be our primary motivation.<sup>18</sup> But evidence supports the influence of innumerable other

---

referring to incursions (often governmental incursions) into areas of limited scope, such as markets or geographical areas. Regulations governing the State Children Health Insurance Program, for instance, require states to implement measures that minimize “crowding out” of private and employer-sponsored plans. See, e.g., 42 C.F.R. § 457.805 (2014). Similar concerns have arisen in the case of government-offered flood insurance. See Richard A. Epstein, *Hayekian Socialism*, 58 MD. L. REV. 271, 296 (1999); Michael Murray, Note, *The Law of Describing Accidents: A New Proposal for Determining the Number of Occurrences in Insurance*, 118 YALE L.J. 1484, 1522 (2009). Courts have used “crowding-out” language to comment that invasive biological organisms can displace native species from their habitats, see *Michigan v. U.S. Army Corps of Eng’rs*, 667 F.3d 765 (7th Cir. 2011), and to suggest that frivolous claims may displace meritorious litigation from judicial dockets, see *O’Loughlin v. Doe*, 920 F.2d 614, 618 (9th Cir. 1990). I propose that these variations be labeled “market share crowding-out” to distinguish them from “motivational crowding-out.”

17. I follow Gerrit De Geest and Giuseppe Dari-Mattiacci in characterizing lawmakers/incentivizers as principals and citizens/incentivized persons as agents. See, e.g., Gerrit De Geest & Giuseppe Dari-Mattiacci, *The Rise of Carrots and the Decline of Sticks*, 80 U. CHI. L. REV. 341, 357 (2013). My use of these terms throughout this Article, however, does not require a formal principal-agent relationship as defined by agency law.

18. See Ernst Fehr & Urs Fischbacher, *Why Social Preferences Matter—The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives*, 112 ECON. J. C1, C1 (2002).



motivations, including reciprocity,<sup>19</sup> inequity aversion,<sup>20</sup> altruism,<sup>21</sup> spite,<sup>22</sup> envy,<sup>23</sup> curiosity,<sup>24</sup> autonomy,<sup>25</sup> professionalism,<sup>26</sup> civic duty,<sup>27</sup> morality,<sup>28</sup> love,<sup>29</sup> dedication to an organizational mission,<sup>30</sup> intellectual passion,<sup>31</sup> intrinsic preference for obeying law,<sup>32</sup> the desire to be a faithful fiduciary,<sup>33</sup> and amusement<sup>34</sup> (to list only a few). Self-interest and other motivations may co-exist. Lawyers are paid for their advocacy, but they may also seek to fulfill intrinsic motivations such as deep concern for their clients, pride, genuine belief in professional standards, and the satisfaction of intellectual curiosity. Despite the long-standing premise that agents are “knaves,” economics research increasingly accommodates motivations beyond economic self-interest, and indeed it must do so to explain phenomena like crowding-out.<sup>35</sup>

Each of the motivations listed above falls under the umbrella of intrinsic motivation. Intrinsic motivation is “the inherent satisfaction[]” of a task, while extrinsic motivation is the desire for a “separable outcome,”<sup>36</sup> such as a reward

---

19. See Dan M. Kahan, *Trust, Collective Action, and Law*, 81 B.U. L. REV. 333 (2001).

20. See Fehr & Fischbacher, *supra* note 18, at C3.

21. *Id.*

22. *Id.*

23. *Id.* at C4.

24. See Richard M. Ryan & Edward L. Deci, *Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions*, 25 CONTEMP. EDUC. PSYCHOL. 54, 56 (2000).

25. See, e.g., Frank Eijkenaar et al., *Effects of Pay for Performance in Health Care: A Systematic Review of Systematic Reviews*, 2 HEALTH POL’Y 115, 125 (2013) (describing the impact of pay-for-performance programs on physicians’ perceived professionalism and autonomy).

26. *Id.*

27. See Yuval Feldman & Oren Perez, *Motivating Environmental Action in a Pluralistic Regulatory Environment: An Experimental Study of Framing, Crowding Out, and Institutional Effects in the Context of Recycling Policies*, 46 LAW & SOC’Y REV. 405, 411 (2012).

28. See Nina Walton, *Crowding Theory and Executive Compensation*, 13 THEORETICAL INQUIRIES L. 429, 435 (2012) (noting “social preferences” that motivate behavior).

29. See Feldman, *supra* note 2, at 12.

30. See, e.g., Timothy Besley & Maitreesh Ghatak, *Competition and Incentives with Motivated Agents*, 95 AM. ECON. REV. 616, 630 (2005) (highlighting nonprofits, public bureaucracies, and education providers as examples where agents are involved in “mission-oriented production” with a non-pecuniary interest in performance); Tore Ellingsen & Magnus Johannesson, *Pride and Prejudice: The Human Side of Incentive Theory*, 98 AM. ECON. REV. 990, 992 (2008) (noting the difference between profit-maximizing and mission-oriented motivations).

31. See Richard S. Saver, *Is It Really All About the Money? Reconsidering Non-Financial Interests in Medical Research*, 40 J.L. MED. & ETHICS 467, 468 (2012).

32. See Robert Cooter, *The Intrinsic Value of Obeying a Law: Economic Analysis of the Internal Viewpoint*, 75 FORDHAM L. REV. 1275, 1275-79 (2006).

33. See also Jason F. Shogren, *Money Pumps & Nudges*, 37 J. AGRIC. & RESOURCE ECON. 349, 355 (2012) (listing examples of motivations that may be crowded out by incentives).

34. *Id.*

35. See Robert Cooter, *Who Gets on Top in Democracy? Elections as Filters*, 10 SUP. CT. ECON. REV. 127, 129-30 (2003).

36. See GRANT, *supra* note 3, at 113; ALFIE KOHN, PUNISHED BY REWARDS 68 (1999); Benkler, *supra* note 2, at 323-24; Christopher Buccafusco et al., *Experimental Tests of Intellectual Property Laws’ Creativity Thresholds*, 92 TEX. L. REV. 1921, 1935 (2014) (describing intrinsic and extrinsic motivation); Feldman, *supra* note 2 (same); Feldman & Lobel, *supra* note 1, at 1178 (reviewing intrinsic and extrinsic motivation); Ryan & Deci, *supra* note 24, at 55-56.

or avoidance of a penalty.<sup>37</sup> Extrinsic motivations are generally associated with material rewards, although this category is broad enough to include other external influences such as mandates.<sup>38</sup> Many commentators have described the difficulty of separating intrinsic and extrinsic motivation;<sup>39</sup> for example, motivations such as the desire to increase self-esteem or avoid guilt may appear on either side of this line.<sup>40</sup> Literature on crowding-out often reflects a preference for intrinsic motivations as more laudable than their “impoverished” extrinsic counterparts.<sup>41</sup> This is to be expected, as fears about motivational crowding-out presume that intrinsic motivations are worth preserving. Commentators alarmed by crowding-out invoke “altruism, reciprocity, and . . . civic virtues,”<sup>42</sup> and using incentives that displace these values may draw us closer to Hume’s view that “every man ought to be supposed . . . to have no other end, in all his actions, than private interest.”<sup>43</sup> But it is important to avoid overlooking intrinsic motivations that may be less laudable, such as envy, spite, anger, revenge, or discriminatory intent. Although intrinsic motivations may not be uniformly noble, however, they may be instrumentally beneficial. Intrinsic motivation has been associated with increased satisfaction and learning,<sup>44</sup> lower costs for principals attempting

---

37. Some consider intrinsic motivations to denote any motivations “that were present before the introduction of external rewards.” See Antoine Beretti et al., *Using Money to Motivate Both “Saints” and “Sinners”: A Field Experiment on Motivational Crowding-Out*, 66 KYKLOS 63, 66 (2013). This temporal definition, however, neglects any extrinsic motivations that predate incentive schemes.

38. See Feldman, *supra* note 2, at 13 (noting that scholars in this field now consider extrinsic motivators “to include not only monetary rewards, but also various legal, verbal, social, and organizational mechanisms which attempt to cause people to engage in socially desirable behaviors”). Most research on motivational crowding-out, however, has focused on the effect of financial rewards. See *id.* at 23.

39. See, e.g., Friedel Bolle & Philipp E. Otto, *A Price Is a Signal: On Intrinsic Motivation, Crowding-Out, and Crowding-In*, 63 KYKLOS 9 (2010); Feldman, *supra* note 2; Rebecca Hollander-Blumoff, *Intrinsic and Extrinsic Compliance Motivations: Comment on Feldman*, 35 WASH. U. J.L. & POL’Y 53, 55-59 (2011); Ryan & Deci, *supra* note 24, at 60-62.

40. Ryan & Deci, *supra* note 24, at 62 (classifying both as “extrinsic”).

41. *Id.* at 55; see also IAN AYRES, *THE \$500 DIET* 38-43 (2010); Samuel Bowles, *Policies Designed for Self-Interested Citizens May Undermine “The Moral Sentiments”: Evidence from Economic Experiments*, 320 SCIENCE 1605, 1606 (2008); William M. Sage, *Some Principles Require Principals: Why Banning “Conflicts of Interest” Won’t Solve Incentive Problems in Biomedical Research*, 85 TEX. L. REV. 1413, 1427 (2007) (noting an “undercurrent” of “moral judgment” in discussions about financial incentives for researchers, in which commentators “seem to be looking for ‘pure-hearted professionals’ whose *raison d’être* is to pursue the ideals of science”).

42. Samuel Bowles & Sung-Ha Hwang, *Social Preferences and Public Economics: Mechanism Design When Social Preferences Depend on Incentives*, 92 J. PUB. ECON. 1811, 1812 (2008).

43. *Id.* at 2 (citing DAVID HUME, *ESSAYS: THE MORAL, POLITICAL AND LITERARY* 117 (T.H. Grose & T.H. Green eds., 1964) (1742)); see also Bruno S. Frey, *A Constitution for Knave Crowds Out Civic Virtues*, 443 ECON. J. 1043 (1997) (using Hume’s language to suggest that legislating on the presumption that individuals will behave in purely self-interested ways will indeed encourage self-interested behavior); Adrian Vermeule, *Hume’s Second-Best Constitutionalism*, 70 U. CHI. L. REV. 421, 424-26 (2003) (suggesting that rules designed on the basis of “the knavery principle” may displace public-spirit).

44. See Bruno S. Frey, *On the Relationship Between Intrinsic and Extrinsic Work Motivation*, 15 INT’L J. INDUS. ORG. 427, 435-36 (1997); Bruce J. Winick, *Harnessing the Power of the Bet: Wagering with the Government as a Mechanism for Social and Individual Change*, 45 U. MIAMI L. REV. 737, 1761 (1991).

to secure good performance in all activities (not just the incentivized tasks);<sup>45</sup> and improvements in self-monitoring, problem-solving, and cognitive processing.<sup>46</sup>

Publicly observable behaviors present a special motivational case. Scholars such as Roland Bénabou and Jean Tirole<sup>47</sup> and Dan Ariely and colleagues<sup>48</sup> divide motivations for prosocial, public behavior into three categories: intrinsic motivation, extrinsic motivation, and “image” or reputational motivation—the desire to shape others’ perceptions of oneself.<sup>49</sup> Reputational motivations are not purely intrinsic: social standing, including one’s opinion of oneself, is separable from the task itself. Such motivations are also not purely extrinsic, as they do not immediately advance material self-interest. Although a good reputation or self-image may ultimately produce material gain, an interim step is needed to transform goodwill into gain (e.g., using social connections to obtain a job or contract). Both extrinsic and reputational motivations are doubtless instrumental.<sup>50</sup> For this analysis, however, it is not necessary to resolve definitively whether reputational motivations are a distinct third category or a subset of extrinsic motivation. It is sufficient to note that incentives may interact with reputational motivations, just as they can interact with intrinsic motivations.

Identifying the precise type, number, and strength of motivations explaining any particular action is daunting and perhaps impossible, even for individual agents. The mix of motivations for a given action is “unobservable,”<sup>51</sup> multiple types of motivation may coexist,<sup>52</sup> and the balance will change across people and time.<sup>53</sup> For example, several studies have identified cultural<sup>54</sup> and gender-based<sup>55</sup> differences in intrinsic motivation. Such distinctions also yield differences in

45. See Winick, *supra* note 44.

46. See Stephanie Stern, *Encouraging Conservation on Private Lands: A Behavioral Analysis of Financial Incentives*, 48 ARIZ. L. REV. 541, 564 (2006).

47. Roland Bénabou & Jean Tirole, *Incentives and Prosocial Behavior*, 96 AM. ECON. REV. 1652, 1654 (2006).

48. Dan Ariely et al., *Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially*, 2009 AM ECON. REV. 544, 544-45 (2009).

49. *Id.* (using image motivation to explain why “people will act more pro-socially in the public sphere than in private settings,” such as by contributing named gifts to charities); see also Roberto Censolo et al., *Electromyographic Activity of Hand Muscles in a Motor Coordination Game: Effect of Incentive Scheme and Its Relation with Social Capital*, 6 PUB. LIBR. SCI. ONE e17372, 2 (2011) (noting that “individuals are intrinsically motivated to social relations . . . [as] distinct from the explicit goal that actually prompts [a] behavior”). Others have classified reputational concerns as *intrinsic* motivations. See, e.g., Beretti et al., *supra* note 37, at 64.

50. See YOCHAI BENKLER, *THE WEALTH OF NETWORKS* 95 (2006); Benkler, *supra* note 2, at 279.

51. Bénabou & Tirole, *supra* note 47, at 1654.

52. See Hollander-Blumoff, *supra* note 39, at 56.

53. See Beretti et al., *supra* note 37, at 63-64.

54. See Robert Eisenberger et al., *Effects of Reward on Intrinsic Motivation—Negative, Neutral, and Positive: Comment on Deci, Koestner, and Ryan (1999)*, 125 PSYCHOL. BULL. 677, 678 (1999).

55. See Rachel Croson & Uri Gneezy, *Gender Differences in Preferences*, 47 J. ECON. LITERATURE 448 (2009).

crowding-out effects. This variation poses a problem for remedying crowding-out in policy and law,<sup>56</sup> given that universal policies may interact with different motivations in various segments of the population.<sup>57</sup>

To take blood donation as an example, one study has found that incentives for blood donation may crowd out willingness to donate among women, but not among men.<sup>58</sup> One explanation for this effect is that women are more motivated by prosocial concerns; another may be that women are more motivated by the reputational benefits gained from uncompensated altruistic behavior.<sup>59</sup> A European survey found that, although residents of many countries disapprove of financial incentives for blood donation, there is cultural variation in acceptance of non-monetary incentives, with highly positive views in Austria and negative views in Italy and Sweden.<sup>60</sup> Cross-cultural variations in motivations and the acceptability of incentives are an important topic for research on crowding-out, but empirical studies are just beginning to consider these factors. There has been little, if any, attempt to account for these differences in the design of incentive-based policies to date.

Building on the definitions above, “incentives” are penalties or rewards intended to shape agents’ behavior by supplying extrinsic motivation. Ruth Grant has provided a helpful definition: an incentive is

(1) an extrinsic benefit or a bonus that is neither the natural or automatic consequence of inaction nor a deserved reward or compensation; (2) a discrete prompt expected to elicit a particular response; and (3) an offer intentionally designed to alter the status quo by motivating a person to choose differently than he or she would be likely to choose in its absence.<sup>61</sup>

Incentives must be known to the agent in advance of his choice; they differ, that is, from windfalls.<sup>62</sup>

---

56. See Beretti et al., *supra* note 37, at 63–64 (“[I]ndividuals may differ in their degree of compliance with some particular kind of intrinsic motivation, such as altruism, . . . [and] several logics of intrinsic motivations—not only altruism but also shame, guilt, self-esteem, reputation concern, etc.—may be at work to varying degrees in different individuals. Neglecting this heterogeneity might distort our understanding of the crowding-out phenomenon at the aggregate level.”).

57. See *id.* at 64.

58. See Nicola Lacetera & Mario Macis, *Do All Material Incentives for Pro-Social Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations*, 31 J. ECON. PSYCHOL. 738 (2010) (finding a strong aversion to cash payments among women, compared to men); Carl Mellström & Magnus Johannesson, *Crowding out in Blood Donation: Was Titmuss Right?*, 6 J. EUR. ECON. ASS’N 845 (2008) (finding that financial incentives may reduce donations by women, but not men); *c.f.* Joan Costa-Font et al., *Not All Incentives Wash out the Warm Glow: the Case of Blood Donation Revisited* 16 (Centre for Econ. Performance Paper No. 1157, 2012), [http://eprints.lse.ac.uk/47679/1/\\_libfile\\_REPOSITORY\\_Content\\_Centre\\_for\\_Economic\\_Performance\\_Discussion\\_papers\\_dp1157.pdf](http://eprints.lse.ac.uk/47679/1/_libfile_REPOSITORY_Content_Centre_for_Economic_Performance_Discussion_papers_dp1157.pdf) (finding that men are more accepting of monetary incentives for donating blood).

59. See Mellström & Johannesson, *supra* note 58.

60. See Costa-Font et al., *supra* note 58, at 16.

61. GRANT, *supra* note 3, at 43.

62. See Edward L. Deci et al., *A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation*, 125 PSYCHOL. BULL. 627, 628 (1999).

Incentives need not be financial; verbal praise or censure, public recognition or shaming, criminal penalties, grades, and promotions are a few examples. Although the crowding-out literature often focuses on positive incentives—carrots that improve the agent’s position relative to her *ex ante* status—incentives also include sticks<sup>63</sup> and mandates. Entire regulatory regimes have been classified as incentives, such as the design of national constitutions,<sup>64</sup> tax law,<sup>65</sup> the tort system,<sup>66</sup> contract enforcement regimes,<sup>67</sup> and workplace regulations.<sup>68</sup> Incentives are agnostic to agents’ other motivations, and they are particularly attractive when complete contracts are impossible due to incomplete information or public policy reasons for nonenforcement.<sup>69</sup>

### *B. Motivational Crowding-Out and the Separability Assumption*

At one time, “the relationship between rewards and motivation was considered to be one of the clearest links in the social sciences,”<sup>70</sup> growing out of psychological research on conditioning behavior<sup>71</sup> and early uses of incentive pay for worker performance in the early 1900s.<sup>72</sup> The traditional view is known as “the separability assumption”<sup>73</sup>—the idea that intrinsic motives and incentives are independent or “additive” in their influence on behavior.<sup>74</sup> Research as early

63. Some incentives may walk a fine line between “negative” and “positive,” such as reductions in an expected criminal sentence, insurance premium, or tax refund. See, e.g., Brian Galle, *The Tragedy of the Carrots: Economics and Politics in the Choice of Price Instruments*, 64 STAN. L. REV. 797, 804-05 (2012) (noting that when a stick is a reduction of an expected benefit, the distinction between carrot and stick is murky). Carrots and sticks may be reframed so that they are synonymous; “carrots can be rewritten as sticks and vice versa.” De Geest & Dari-Mattiacci, *supra* note 17, at 349. There are, however, important differences between the two approaches, including the fact that sticks “can make citizens worse off while enforcement with carrots cannot.” *Id.* at 364.

64. See Frey, *supra* note 43; Vermeule, *supra* note 43, at 425 (“Designing institutions for knaves may itself beget knaves.”).

65. See Frey, *supra* note 43, at 1050; Sagit Leviner, *A New Era of Tax Enforcement: From “Big Stick” to Responsive Regulation*, 42 U. MICH. J.L. REFORM 381, 402-03 (2009).

66. See Hollander-Blumoff, *supra* note 39.

67. See Bohnet et al., *supra* note 9.

68. See, e.g., Bruno S. Frey, *Shirking or Work Morale?: The Impact of Regulating*, 37 EUR. ECON. REV. 1523 (1993).

69. See Bowles, *supra* note 41, at 1605-06.

70. Amy Rummel & Richard Feinberg, *Cognitive Evaluation Theory: A Meta-Analytic Review of the Literature*, 16 SOC. BEHAV. & PERSONALITY 147, 147 (1988) (citing early crowding-out research). For an overview of other early studies, see Deci et al., *supra* note 62, at 627 (citing research in the education field from 1971 onward).

71. Deci et al., *supra* note 62.

72. See, e.g., GRANT, *supra* note 3, at 14-30 (describing the history of socially-engineered incentives in labor management, socialist economics, and behavioral psychology); MICHAEL J. SANDEL, WHAT MONEY CAN’T BUY 86-87 (2012); Edward L. Deci, *The Effects of Contingent and Noncontingent Rewards and Controls on Intrinsic Motivation*, 8 ORG. BEHAV. & HUM. PERFORMANCE 217 (1972).

73. See Bowles, *supra* note 41, at 1605; Alfie Kohn, *Why Incentive Plans Cannot Work*, HARV. BUS. REV., Sept.-Oct. 1993, at 2, 7.

74. See, e.g., SANDEL, *supra* note 72, at 122; Walton, *supra* note 28, at 436 (“Agency theory accepts that individuals may have social preferences; however, . . . it assumes that the effect of incentives is strictly an additive to underlying social preferences . . .”).

as 1968, however, suggested that rewards had diverse and potentially detrimental effects.<sup>75</sup> Separate strands of research in psychology and economics<sup>76</sup> have investigated this phenomenon since the 1970s,<sup>77</sup> particularly incited by Richard Titmuss's seminal 1971 argument against payment for blood donation,<sup>78</sup> Kenneth Arrow's famous rejection of Titmuss's theory,<sup>79</sup> Edward Deci's studies of motivation,<sup>80</sup> and Mark Lepper's study of incentives for education.<sup>81</sup> Both disciplines now cast doubt on the separability assumption,<sup>82</sup> and incentives are now expected to have two separate effects: (1) the "disciplining effect" or "relative price effect," which refers to the extent to which an incentive supplies extrinsic motivation,<sup>83</sup> and (2) crowding-out (or crowding-in), which refers to the way in which incentives interact with intrinsic motivations.

Motivational crowding-out occurs when introducing an incentive for a task provokes a loss of intrinsic motivation.<sup>84</sup> Synonyms include the "hidden cost of reward,"<sup>85</sup> the "corruption effect,"<sup>86</sup> the "undermining effect,"<sup>87</sup> the

75. See Rummel & Feinberg, *supra* note 70, at 147.

76. The contemporary psychology literature is most defined by the work of Edward Deci, while much of the recent research on crowding-out in economics is by Bruno Frey. See BENKLER, *supra* note 50, at 93-94 (summarizing current literature).

77. See, e.g., Benkler, *supra* note 2, at 321-23; Frey & Jegen, *supra* note 2, at 589-90; Uri Gneezy et al., *When and Why Incentives (Don't) Work to Modify Behavior*, 25 J. ECON. PERSP. 191, 193 (2011); Uri Gneezy, *The W Effect of Incentives*, 21-22 (U. Chi. Working Paper, 2003).

78. RICHARD TITMUSS, *THE GIFT RELATIONSHIP* (1971); see also BENKLER, *supra* note 50, at 93 (describing Titmuss's work); Walton, *supra* note 28, at 437 (citing Titmuss's work as "the first to raise the theoretical possibility of crowding out").

79. See SANDEL, *supra* note 72, at 125-26 (recounting the Titmuss-Arrow disagreement); Kenneth J. Arrow, *Gifts and Exchanges*, 1 PHIL. & PUB. AFF. 343 (1972).

80. See Edward L. Deci, *Effects of Externally Mediated Rewards on Intrinsic Motivation*, 18 J. PERSONALITY & SOC. PSYCHOL. 105 (1971).

81. See Mark R. Lepper et al., *Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the "Overjustification" Hypothesis*, 28 J. PERSONALITY & SOC. PSYCHOL. 129 (1973); see also Evan R. Seamone, *A Refreshing Jury Cola: Fulfilling the Duty to Compensate Jurors Adequately*, 5 N.Y.U. J. LEGIS. & PUB. POL'Y 289, 379 n.359 (2002) (describing how Lepper's work "is often noted in the psychology literature as the prototype for intrinsic motivation").

82. For a survey of the crowding-out literature, see Bowles, *supra* note 41, at 1606; Deci et al., *supra* note 62; Frey & Jegen, *supra* note 2; Promberger & Marteau, *supra* note 2, at 950-52 (discussing economic and psychological literature on crowding-out).

83. Frey, *supra* note 44, at 429; see also Frey, *supra* note 43, at 1046; Frey & Jegen, *supra* note 2, at 593; Gneezy et al., *supra* note 77, at 192 (discussing the "standard direct price effect").

84. Some have suggested that intrinsic motivations may crowd one another out; for example, the motivation to satisfy "visceral" needs such as hunger or pain may trump other motivations. See, e.g., Adam Benforado et al., *Broken Scales: Obesity and Justice in America*, 53 EMORY L.J. 1645, 1683 (2004); Jon Hanson & David Yosifon, *The Situational Character: A Critical Realist Perspective on the Human Animal*, 93 GEO. L.J. 1, 121-22 (2004). This dynamic, however, is outside the scope of my analysis.

85. See, e.g., Frey, *supra* note 43, at 1044; Bruno S. Frey & Felix Oberholzer-Gee, *The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out*, 87 AM. ECON. REV. 746, 746 (1997).

86. See Shogren, *supra* note 33, at 355 (listing synonyms for this effect).

87. See Christopher P. Cerasoli et al., *Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis*, 140 PSYCHOL. BULL. 980 (2014).

“overjustification effect,”<sup>88</sup> and “the price of a price.”<sup>89</sup> The reverse of crowding-out, of course, is crowding-in, in which the incentive not only exerts a relative price effect on behavior but also simultaneously *augments* intrinsic motivation.<sup>90</sup> Given the normative preference for intrinsic motivation, however, crowding-in is rarely cited as a cause for concern about incentive-based policies.<sup>91</sup> Where crowding-out occurs, the net impact of an incentive depends on the relative strength of the disciplining effect compared to the crowding-out effect. Where the disciplining effect is stronger, an incentive can have a net behavioral benefit, although this benefit will be lessened by crowding-out effects.

To date, the crowding-out literature has not distinguished between absolute and relative reductions in intrinsic motivation. I shall attempt to do so. When an incentive causes an *absolute* reduction, intrinsic motivation is lower than it would have been without the incentive. This is the classic concern about crowding-out effects; if this effect is severe, crowding-out may yield net worse behavior.<sup>92</sup> But crowding-out should also encompass a *relative* reduction of intrinsic compared to extrinsic motivation, whereby the absolute measure of intrinsic motivation may be unchanged, but extrinsic motivation is now the driving force behind the agent’s choices.<sup>93</sup> Although relative losses of intrinsic motivation may not undermine incentive effects, they may have important impacts on performance quality.<sup>94</sup>

Several prerequisites are needed for either type of crowding-out to occur. First, intrinsic motivation must exist for the task, and crowding-out concerns may be more applicable when intrinsic motivation is high.<sup>95</sup> Some have interpreted this to mean that crowding-out can *only* occur when incentives are applied to “interesting tasks,”<sup>96</sup> but intrinsic motivations may also drive boring or

88. See Bénabou & Tirole, *supra* note 47, at 1654; Deci et al., *supra* note 62, at 630.

89. Maarten C.W. Janssen & Ewa Mendys-Kamphorst, *The Price of a Price: On the Crowding Out and In of Social Norms*, 55 J. ECON. BEHAV. & ORG. 377 (2004). When rewards are offered by companies seeking to reward consumer behavior, crowding-out has also been described as “promotional reactance.” Ran Kivetz, *Promotion Reactance: The Role of Effort-Reward Congruity*, 31 J. CONSUMER RES. 725 (2005).

90. Frey & Jegen, *supra* note 2, at 593.

91. To the extent crowding raises concerns about autonomy, however, crowding-in should be equally distressing as crowding-out.

92. See Beretti et al., *supra* note 37, at 65.

93. Of course, any time an incentive is introduced, even a de minimis desire to obtain that incentive will alter the balance between intrinsic and extrinsic motivation. But the concerns of crowding theorists, particularly concerns about performance quality and future performance, are most acute when extrinsic motivations overtake intrinsic motivations as the primary motivators of behavior.

94. These impacts may include “choking,” changing behavior to emphasize short-term performance rather than long-term progress, and distraction. See Buccafusco et al., *supra* note 36, at 1972 (describing choking behavior, and finding no evidence of choking as defined by less creativity in the presence of incentives); Kristen Underhill, *The Harms and Benefits of Motivational Crowding-Out and a Justification for Regulating Incentive-Based Policies* 25 (2015) (unpublished manuscript) (on file with author).

95. Frey, *supra* note 44, at 431.

96. See, e.g., Kou Murayama et al., *Neural Basis of the Undermining Effect of Monetary Reward on Intrinsic Motivation*, 107 PROC. NAT’L ACAD. SCIS. 20911, 20912 (2010).

unpleasant tasks, such as donating blood or recycling. Second, the agent must know or expect the incentive before he undertakes the task. Third, because the mechanisms for crowding-out are cognitive, the task must demand a decision by the agent; activities beyond the agent's control or ability may respond to neither intrinsic nor extrinsic motivation.

Empirical evidence for motivational crowding-out comes from many settings,<sup>97</sup> including at least ten meta-analyses in psychology and economics.<sup>98</sup> The most recent systematic review of this literature identified 183 primary studies over four decades of research, documenting evidence of crowding-out in education, the workplace, and physical activity.<sup>99</sup> Econometric studies have modeled crowding-out effects in subject areas including volunteering, paid labor, tax evasion, and the siting of socially beneficial but locally undesirable projects;

---

97. See, e.g., Bowles, *supra* note 41, at 1606 (summarizing forty-one studies).

98. The most recent of these is Cerasoli et al., *supra* note 87. Past meta-analyses include the following: Judy Cameron et al., *Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues*, 24 BEHAV. ANALYST 1 (2001) (reviewing 145 studies to suggest that "[n]egative effects are found on high-interest tasks when the rewards are tangible, expected (offered beforehand), and loosely tied to level of performance," but casting doubt on a "pervasive" motivational crowding-effect); Judy Cameron & W. David Pierce, *Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis*, 64 REV. EDUC. RES. 363 (1994) (reviewing ninety-six experimental studies to suggest a negative effect of incentives "only . . . when expected tangible rewards are given to individuals simply for doing a task"); Edward L. Deci et al., *Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again*, 71 REV. EDUC. RES. 1 (2001) (identifying differences between meta-analyses by Cameron & Pierce (1994) and Deci et al. (1999) and arguing that the meta-analysis by Deci et al. is of higher methodological quality); Deci et al., *supra* note 62, at 627, 632 (reviewing 128 primary studies to conclude that "engagement-contingent, completion-contingent, and performance-contingent rewards significantly undermined free-choice intrinsic motivation . . . as did all rewards, all tangible rewards, and all expected rewards," and criticizing prior meta-analyses by Eisenberger, Cameron, and Pierce that had cast doubt on the crowding-out phenomenon); Robert Eisenberger & Judy Cameron, *Detrimental Effects of Reward: Reality or Myth?*, 51 AM. PSYCHOLOGIST 1153, 1153, 1162 (1996) (re-analyzing findings from Cameron & Pierce (1994) to conclude that "detrimental effects of reward occur under highly restricted, easily avoidable conditions," in which a reward is "presented on a single occasion without regard to the quality of performance or task completion," or when the reward is given "for a low degree of divergent [creative] thought;" this study also concluded, however, that "either a small reward or a large, nonsalient reward can be effectively used to increase generalized creativity"); Eisenberger et al., *supra* note 54, 677 (criticizing the findings of Deci et al. (1999)); Rummel & Feinberg, *supra* note 70, 159 (reviewing forty-five studies to conclude that "indeed, there is a detrimental effect of extrinsic rewards on intrinsic motivation"); Alexander D. Stajkovic & Fred Luthans, *A Meta-Analysis of the Effects of Organizational Behavior Modification on Task Performance, 1979-95*, 40 ACAD. MGMT. J. 1122, 1140 (1997) (reviewing nineteen studies to find that, when financial and nonfinancial incentives were combined, the financial incentive neither helped nor diminished the intervention effect); Shu-Hua Tang & Vernon C. Hall, *The Overjustification Effect: A Meta-Analysis*, 9 APPLIED COGNITIVE PSYCHOL. 365, 365, 368, 373 (1995) (reviewing fifty experimental studies to find evidence for crowding-out "across age, dependent measure, and design type in specific situations where it is predicted to occur," namely, where "intrinsic interest is initially high, the reward is task contingent, expected, and tangible with no additional feedback," or where the reward is given without any feedback affirming the agent's competence in performing the task); Uco J. Wiersma, *The Effects of Extrinsic Rewards in Intrinsic Motivation: A Meta-Analysis*, 5 J. OCCUPATIONAL & ORG. PSYCHOL. 101 (1992) (analyzing twenty studies of tangible incentives to find that "support for the overjustification effect occurs only when intrinsic motivation is operationalized as task behaviour during a free-time measure," not when it is assessed by task performance).

99. Cerasoli et al., *supra* note 87. Another recent overview of evidence for using incentives to motivate educational performance, prosocial behavior, and health behaviors appears in Gneezy et al., *supra* note 77.



economic laboratory and field studies have also documented crowding-out effects for motives as diverse as volunteerism, reciprocity, work effort, altruism, norm adherence, environmentalism, civic duty in the use of shared resources, trust, and civic virtue in the presence of adversary institutions.<sup>100</sup> A large body of research has also examined the impact of performance-related pay,<sup>101</sup> with varying findings that have included reduced satisfaction among highly motivated employees,<sup>102</sup> and reduced productivity among those who perceive bonuses to be “controlling.”<sup>103</sup> Empirical studies more frequently focus on crowding-out effects resulting from rewards instead of penalties, although evidence exists for crowding-out in both scenarios.<sup>104</sup> Several physiological studies have also documented biological evidence of crowding-out effects—changes in brain and muscle activity associated with the imposition and withdrawal of incentives for completing cognitive and physical tasks.<sup>105</sup>

Although legal scholarship has primarily engaged with motivational crowding-out through theory, several empirical papers have specifically considered crowding out in relation to legal rules.<sup>106</sup> In one analysis, Yuval Feldman and Oren Perez surveyed 1,800 residents of Israel about three hypothetical environmental laws—mandatory bottle deposits, fines for dumping bottles, and an ethical code without penalties—asking them about their intrinsic environmental motivation and predicted personal efficacy (compliance) in

100. See Frey & Jegen, *supra* note 2, at 599 (citing studies).

101. See, e.g., *id.* at 595-96 (describing research on pay-for-performance).

102. Kirsten Bregm, *Detrimental Effects of Performance-Related Pay in the Public Sector? On the Need for a Broader Theoretical Perspective*, 13 PUB. ORG. REV. 21, 28 (2013) (citing A.M. Bertelli, *Motivation Crowding and the Federal Civil Servant: Evidence from the U.S. Internal Revenue Service*, 9 INT'L PUB. MGMT. J. 3 (2006)).

103. Bregm, *supra* note 102, at 27 (citing L. Andersen & T. Pallesen, “Not Just for the Money?” *How Financial Incentives Affect the Number of Publications at Danish Research Institutions*, 11 INT'L PUB. MGMT. J. 28 (2008)).

104. Hollander-Blumoff, *supra* note 39, at 59-60.

105. See Censolo et al., *supra* note 49; Qingguo Ma et al., *The Dark Side of Monetary Incentive: How Does Extrinsic Reward Crowd out Intrinsic Motivation*, 25 NEUROREPORT 194 (2014); Murayama et al., *supra* note 96.

106. See, e.g., Bohnet et al., *supra* note 9, at 132 (finding that moderate contract enforcement can crowd out trustworthiness); Buccafusco et al., *supra* note 36, at 1971 (showing that incentives resembling patents do not impair creativity relative to other types of incentives); Alessandra Cassar et al., *Institutional Quality, Culture, and Norms of Cooperation: Evidence from Behavioral Field Experiments*, 57 J.L. & ECON. 821, 855 (2014) (finding that, in the presence of strong formal institutions, trustworthiness does not influence cheating behavior); Feldman & Lobel, *supra* note 1, at 1207 (finding that offering whistleblowers incentives may decrease the reporting of illegal activity); Feldman & Perez, *supra* note 27, at 433 (finding that among people with high environmental motivation, the behavioral effects of fines for not recycling decreased as fines increased); Steven Kelman, *Adversary and Cooperationist Institutions for Conflict Resolution in Public Policymaking*, 11 J. POL'Y ANALYSIS & MGMT. 178, 179 (1992) (suggesting that adversarial institutions crowd out public spirit); Scott E. Masten & Jens Prüfer, *On the Evolution of Collective Enforcement Institutions: Communities and Courts*, 43 J. LEGAL STUD. 359, 359 (2014) (finding that the existence of courts can weaken community enforcement of behavior); Brian Sheppard & Fiery Cushman, *Evaluating Norms: An Empirical Analysis of the Relationship Between Norm-Content, Operator, and Charitable Behavior*, 63 VAND. L. REV. 55, 88 (2010) (finding that mandates requiring students to give charitably reduced giving below aspirational statements encouraging giving).

response to a randomly-selected law.<sup>107</sup> Findings were consistent with crowding-out theory, showing that the already-motivated individuals predicted low personal responses to deposit programs and large fines.<sup>108</sup> Another empirical study by Feldman and Orly Lobel found that offering monetary incentives to whistleblowers may interfere with ethical and social image motivations to report others' wrongdoing.<sup>109</sup> As another example of empirical legal scholarship on crowding-out effects, Christopher Buccafusco and colleagues conducted a recent simulation study of incentives for innovation, but found that incentives in this instance did not impair creativity.<sup>110</sup>

The evidence for crowding-out is not uniform<sup>111</sup> or without controversy.<sup>112</sup> Many incentives do have a net beneficial effect on behavior, even those that cause some crowding-out, and some research has suggested that extrinsic rewards or penalties can sometimes *increase* perceived self-determination,<sup>113</sup> crowd *in* intrinsic motivation, or increase the internalization of norms.<sup>114</sup> Much evidence on crowding-out also supports the premise that it is context-specific and differs by culture,<sup>115</sup> setting,<sup>116</sup> age,<sup>117</sup> and behavior.<sup>118</sup>

### C. Crowding-Out Concerns and Incentive-Based Policies

As suggested by the examples in the Introduction,<sup>119</sup> crowding-out concerns have influenced commentary and at times rule development across a range of substantive legal areas. Crowding-out is more frequently associated with positive incentives compared to penalties or mandates, but all have given rise to

107. Feldman & Perez, *supra* note 27.

108. *Cf. id.* at 432-33 (also finding, however, that highly-motivated residents did not react negatively to low-level fines, which may give motivated people assurance that free riders would not take advantage of their good behavior).

109. Feldman & Lobel, *supra* note 1; *see also* Orly Lobel, *Linking Prevention, Detection, and Whistleblowing: Principles for Designing Effective Reporting Systems*, 54 S. TEX. L. REV. 37, 46-47 (2012).

110. *See* Buccafusco et al., *supra* note 36.

111. *See id.* at 19435-43 (examining inconsistent research findings on how incentives influence creativity).

112. Wendy Netter Epstein, *Public-Private Contracting and the Reciprocity Norm*, 64 AM. U. L. REV. 1, 42 (2014).

113. *See, e.g.*, Feldman & Lobel, *supra* note 1, at 1180.

114. *See infra* notes 184-186 and accompanying text for a discussion of this “expressive” effect of law.

115. *See* Costa-Font et al., *supra* note 58 (finding national differences in the acceptability of incentives for blood donation in fifteen European countries).

116. *Id.* at 6 (“Most of the empirical studies suggest that crowding out is specific of the particular settings individuals are in.”).

117. Deci et al., *supra* note 62, at 657 (finding more extensive crowding effects among children compared to adults).

118. *See, e.g.*, Buccafusco et al., *supra* note 36, at 1976 (“[W]hen it comes to incentives and creativity, context matters a lot”); Promberger & Marteau, *supra* note 2, at 954-55 (suggesting that crowding-out effects for health behaviors may differ from other contexts).

119. *See supra* notes 5-11 and accompanying text.

discussion in legal scholarship. (Mandates and negative incentives are particularly blurred, as most mandates are enforced by fines, tort liability, or criminal liability for noncompliance.) A select few illustrations will help to frame the following discussion of processes by which motivational crowding-out may occur.<sup>120</sup> I have organized these illustrations by the type of motivation that we seek to preserve, including altruism, professionalism, and civic duty in the following short, non-exhaustive list:

### 1. Altruism Crowding-Out

Perhaps the most well-known example of crowding-out concerns organ donation, where many have considered the argument that permitting payments for living donation (for example, of kidneys or bone marrow) may diminish altruism among potential donors.<sup>121</sup> The crowding-out argument provided one justification for the 1984 National Organ Transplant Act, which criminalized the acquisition, receipt, or transfer of human organs for “valuable consideration.”<sup>122</sup> The Ninth Circuit recently upheld the Act as applied to bone marrow donation through aspiration (an invasive procedure), finding that crowding-out concerns provided a rational basis for the ban on compensation.<sup>123</sup>

The preservation of altruistic motives has also been at the heart of law concerning Good Samaritan activity, on the basis that mandates or positive incentives for rescuing others may crowd out more noble motivations. For example, in *Stockberger v. U.S.*, the Seventh Circuit refused to impose a common-law duty to rescue (i.e., a mandate enforced by negative incentives through tort penalties), noting that “liability [for not rescuing] might actually reduce the number of altruistic rescues by depriving people of credit for altruism.”<sup>124</sup> The same argument has been applied to market-based rewards<sup>125</sup> and mandates. Moreover, although most states offer rescuers a reduced standard of care in tort for injuries inflicted during rescue, some only permit Good

120. See Underhill, *supra* note 94 (providing a more comprehensive overview of crowding-out across a range of incentive-based policies).

121. See DEBRA SATZ, WHY SOME THINGS SHOULD NOT BE FOR SALE: THE MORAL LIMITS OF MARKETS 192-95 (2010); Benjamin E. Hippen & Sally Satel, *Crowding Out, Crowding in, and Financial Incentives for Organ Procurement*, in WHEN ALTRUISM ISN'T ENOUGH: THE CASE FOR COMPENSATING KIDNEY DONORS 96 (Sally Satel, ed., 2008); Richard A. Epstein, *The Human and Economic Dimensions of Altruism: The Case of Organ Transplantation*, 37 J. LEGAL STUD. 459, 474-81 (2008) (describing crowding-out and noting uncertainty about the applicability of evidence from other fields); Horton, *supra* note 5.

122. 42 U.S.C. § 274e (2012).

123. *Flynn v. Holder*, 684 F.3d 852, 859-62 (9th Cir. 2012); see also John A. Robertson, *Paid Organ Donations and the Constitutionality of the National Organ Transplant Act*, 40 HASTINGS CONST. L.Q. 221 (2013).

124. *Stockberger v. U.S.*, 332 F.3d 479, 481 (7th Cir. 2003); see also Dan M. Kahan, *The Logic of Reciprocity: Trust, Collective Action, and Law*, 102 MICH. L. REV. 71, 79 (2003).

125. See Landes & Posner, *supra* note 6.

Samaritans to receive this benefit if their acts were “gratuitous and without expectation of compensation.”<sup>126</sup>

## 2. Professionalism Crowding-Out

Pay-for-performance (P4P) regimes, which reward employees or contractors based on job performance, have drawn increasing criticism on the grounds that they may crowd out professionalism, ethics, and inherent satisfaction in job tasks.<sup>127</sup> The federal tax code currently encourages pay-for-performance arrangements for executives of public corporations, which cannot deduct executive salaries above \$1 million unless the salary arrangement is linked to performance goals.<sup>128</sup> Pay-for-performance programs have been particularly controversial in health care as a means of paying individual physicians based on quality of care.<sup>129</sup> In 2013, Congressional testimony on Medicare payment structures included a description of potential crowding-out effects associated with P4P.<sup>130</sup> Empirical evidence for crowding-out associated with P4P in health care, however, has been mixed.<sup>131</sup>

## 3. Civic Duty Crowding-Out

Positive incentives, negative incentives, and mandates have all aroused concern on the basis that they crowd out citizens’ sense of civic duty. This commentary tends to focus on government-offered incentives or mandates for voting,<sup>132</sup> tax payment,<sup>133</sup> jury service,<sup>134</sup> and local acceptance of public works projects (“NIMBY” projects).<sup>135</sup> Frey has notably extended this principle to the design of public laws, on the theory that highly restrictive or controlling legal regimes communicate distrust of citizens.<sup>136</sup> In this view, laws that aim to

---

126. See Victoria Sutton, *Is There a Doctor (and a Lawyer) in the House? Why Our Good Samaritans Laws Are Doing More Harm Than Good for a National Public Health Security Strategy: A Fifty-State Survey*, 6 J. HEALTH & BIOMEDICAL L. 261, 277 (2010).

127. See, e.g., KOHN, *supra* note 36; Stout, *supra* note 7, at 533; David A. Weisbach, *Expenditures, Principal Agent Problems, and Redundancy*, 84 WASH. U. L. REV. 1823 (2006); Cynthia A. Williams & John M. Conley, *The Social Reform of Banking*, 39 J. CORP. L. 459, 474 (2014).

128. 26 U.S.C. § 162(m) (2012) (disallowing public corporations from deducting annual compensation above \$1 million for the CEO or top four highest compensated officers, if that compensation is not tied to performance goals); see also Stout, *supra* note 7.

129. See, e.g., Eijkenaar et al., *supra* note 25.

130. *Medicare Physician Payment System: Hearing Before the Subcomm. for Health of the H. Comm. on Energy and Commerce*, 113th Cong., 126-138 (2013) (statement of Robert A. Berenson, M.D., Institute Fellow, Urban Institute).

131. Jon B. Christianson et al., *Lessons from Evaluations of Purchaser Pay-for-Performance Programs*, 65 MED. CARE RES. & REV. 5S (2008).

132. Eric Fleisig-Greene, *Law’s War with Conscience: The Psychological Limits of Enforcement*, 2007 B.Y.U. L. REV. 1203 (2007).

133. Listokin & Shizer, *supra* note 11.

134. Seamone, *supra* note 10.

135. Frey & Oberholzer-Gee, *supra* note 85.

136. Frey, *supra* note 43.

“discipline” or control citizens through strict regulation or policing can erode public trust and reciprocity motivations to comply with the law, encouraging minor lawbreaking, tax evasion, and resistance to public works projects.<sup>137</sup>

Although these illustrations are far from exhaustive, they gesture toward the breadth of crowding-out concerns in the law. Despite the pervasiveness of crowding-out commentary, however, legal scholarship has an impoverished view of what happens when crowding-out occurs. Whether or not crowding-out occurs in any given domain is secondary to understanding precisely what crowding-out *is*, and what forms it may take in response to incentive-based legal policies. Although some legal scholars have noted that multiple types of mechanisms may drive crowding-out, none have undertaken a classification of these effects, and several have specifically declined to address this question.<sup>138</sup> This view impairs efforts to address crowding-out through improvements in incentive design. Crowding-out actually consists of several distinct processes, to which I will now turn.

## II. Heterogeneity in Motivational Crowding-Out Processes

The distinct processes animating crowding-out have not yet been disaggregated in legal scholarship. Each of these behavioral models is grounded in field and laboratory experiments interpreting variables such as time spent on a task, willingness to spend resources, self-reported motivation, or physiological data as evidence of motivation. Multiple mechanisms may also operate simultaneously in a given context.<sup>139</sup> The remainder of this Part will consider four categories of crowding-out mechanisms that may be triggered when a principal offers an agent an incentive.

The first category includes signaling processes, by which incentives send or disrupt informational signals. Incentives themselves can convey information to agents, including information about the task itself (e.g., whether it is dangerous), information about the principal’s opinion of the agent (e.g., whether the principal finds the agent trustworthy), information about the principal (e.g., whether the principal is wasteful), information about social norms (e.g., whether the task is popular), and information about how the agent should make decisions about the task (e.g., whether the choice is economic or moral). Incentives can also interfere with the signals that an agent sends to *others* or *herself* about her behavior; an agent may undertake a task like voting to show she is public-

---

137. *Id.* at 1048.

138. See, e.g., Atiq, *supra* note 4, at 1081; Benkler, *supra* note 2, at 326 (“For purposes of my analysis here it is not necessary to pin down precisely the correct and most complete theory of motivation or the full extent and dimensions of crowding out.”); Richard M. Hynes, *Posted: Notice and the Right to Exclude*, 45 ARIZ. ST. L.J. 949, 981 (2013).

139. See Bowles, *supra* note 41, at 1606 (noting that separate crowding-out processes “often work jointly and sometimes with opposite effect”); Walton, *supra* note 28, at 440 (noting that while a single crowding mechanism may be “explanatory,” it is likely that multiple mechanisms are working simultaneously when crowding-out occurs).

spirited, and a fine for nonvoting can interfere with this message. The second category is the impairment of self-determination, by which incentives can reduce agents' perceived autonomy and cause them to disengage. A third category is the long-term adaptation of preferences, on the assumption that our values depend in part on the rules and entitlements that surround us, including the presence of incentives. Finally, I will note a less-accepted model of crowding-out known as "learned helplessness," by which certain incentives may discourage people who do not see a clear link between performance and reward.<sup>140</sup>

### A. Signaling Processes

All signaling processes assume that incentives affect the flow of information. Sometimes incentives transmit information directly, which occurs when incentivized agents interpret the incentive as information already known by the principal, who offered them the incentive. This Section will begin by noting these direct signaling processes, including how incentives may lead agents to draw conclusions about the value of a task, their own abilities or importance, the values of the principal who offered the incentive, social norms governing the task, or whether decisions about the task should be self-interested or not.

I will then consider how incentives disrupt the signals that agents try to send to *others* through their behavior. An agent may act in order to cultivate his reputation as altruistic, responsible, or civic-minded; if the agent receives a reward for his behavior, however, observers might attribute his behavior to greed rather than to his good character. This effect is image-spoiling when it refers to an agent's reputation. When incentives reduce the agent's own self-image, it is self-image spoiling. This Section will illustrate each of these processes below.

#### 1. Principal Signaling to Agents

When a principal offers an incentive, the agent may interpret it as a signal communicating the principal's private information,<sup>141</sup> regardless of the principal's intended meaning. This signal may cause the agent to update his expectations of how attractive or burdensome the task is, how competent he will

---

140. Samuel Bowles and Sandra Polanía-Reyes have also proposed a four-part categorization of crowding-out effects, including conveying information about the principal, framing choices as self-interested, reducing agents' sense of autonomy, and influencing learned preferences. Bowles & Polanía-Reyes, *supra* note 2; see also Bowles, *supra* note 41, at 1606-08. My classification differs from Bowles's four categories in several ways. I consider framing and "bad news" about principals both to be signaling processes, along with other signals sent by the principal to the agent (including task value, agent value, and social norms), signals sent by the agent to observers, and signals sent by the agent to him- or herself. Bowles does not consider this broad variety of signaling mechanisms. I also include the "learned helplessness" explanation as a fourth category, although this addition is secondary to the prior contributions.

141. See Roland Bénabou & Jean Tirole, *Intrinsic and Extrinsic Motivation*, 70 REV. ECON. STUD. 489, 491 (2003).

be in accomplishing it, what the principal's motives or values may be, how the task is perceived by others, and whether decisions regarding the task should be governed by market norms or moral reasoning.

#### a. Incentives Can Communicate Information about the Task

Incentives may reduce intrinsic motivation when they signal to the agent that the task is burdensome, unpleasant, or costly.<sup>142</sup> Larger incentives may increase the perception of the behavior as personally costly and thereby reduce willingness to engage in the activity for its own sake. For example, agents paid to donate blood may “view the payment . . . as compensation for the risks of donating.”<sup>143</sup> If the payments are higher than expected, indicating a higher-than-expected risk, a risk-averse individual may decide to stop donating.<sup>144</sup> In one potential test of this mechanism, a notable Swiss study asked citizens if they would allow the government to build a nuclear waste plant in their town, given the risks of accidents and groundwater contamination.<sup>145</sup> Although approximately 51% were initially willing to accept the repository, this percentage dropped to approximately 25% when an annual monetary payment was offered (ranging between \$2,175 and \$6,525 per year). One theoretical explanation was that the payment signaled to the residents that the facility was dangerous; although only 6% of respondents believed that the compensation affected their risk perceptions in this case,<sup>146</sup> the explanation merits investigation in other contexts.

Small incentives may also reduce intrinsic motivation when they signal that a task is unimportant.<sup>147</sup> Rewards provide a market valuation of agents' performance (a price): when the price is lower than the agent expected, the agent revises her belief in the task's value, and thereby becomes less motivated to undertake the task.<sup>148</sup> This effect may also be explained through “equity theory,”

---

142. See *id.* at 494; Bénabou & Tirole, *supra* note 47, at 1654; Bolle & Otto, *supra* note 39; Censolo et al., *supra* note 49, at 1-2; Gneezy et al., *supra* note 77, at 192; Emir Kamenica, *Behavioral Economics and Psychology of Incentives*, 4 ANN. REV. ECON. 13.1, 13.18 (2012).

143. Bolle & Otto, *supra* note 39, at 18.

144. *Id.* Bénabou and Tirole describe this as the converse of the “forbidden fruit” effect, in which penalties or prohibitions on an activity can increase its allure. Bénabou & Tirole, *supra* note 141, at 490, 498.

145. Frey & Oberholzer-Gee, *supra* note 85.

146. *Id.* at 749-50 (considering but rejecting the idea that incentives signaled danger in this instance); see also Bolle & Otto, *supra* note 39, at 18 (suggesting that paying blood donors may cause them to revise upward their estimate of the physical risks associated with the donation process).

147. “If paid, [the agent] receives a signal, namely a price, which she may perceive as proportionate to the market value of her activity or as an estimate thereof by others. The consequence of the adoption of this valuation is crowding-out if and only if her own estimation had been considerably higher than . . . the price offered.” Bolle & Otto, *supra* note 39, at 10; see also Gneezy, *supra* note 77, at 20 (noting that incentives act to “complete” incomplete contractual arrangements, in effect setting task price or value).

148. Bolle & Otto, *supra* note 39, at 11-13.

which posits that agents try to balance their contributions with their rewards, decreasing their effort when they believe the reward is too small.<sup>149</sup>

#### b. Incentives Can Signal the Value of the Agent

Incentives may also provide information to agents about how the principal views their motivation, competence, and trustworthiness. The effect of this signal will vary depending on agents' interpretation. To the extent that large rewards reinforce agents' belief in their own competence, these incentives may "impart a competence-boosting message," particularly when incentives are contingent on high-quality performance.<sup>150</sup> The signaling function of incentives may backfire, however, if the agent interprets an incentive to mean that the principal is pessimistic about the agent's competence, work ethic, or motivation, and that the principal has therefore concluded that incentives are necessary.<sup>151</sup> This signal may lower agents' confidence in their own skills, undermining motivation to attempt the activity. When the agent perceives but disagrees with the principal's view, he may feel that the principal undervalues his competence or motivation, leading to a similar loss of self-esteem.<sup>152</sup> Moreover, if the agent perceives an incentive as indicating that the principal's expectations are low, the agent will anticipate feeling less guilt for underperformance, which excuses a lower level of engagement.<sup>153</sup>

Several studies have found that agents view incentives as information about whether the principal finds them trustworthy, each demonstrating that agents perform better on principals' behalf when the principal refrains from imposing fines, lower boundaries on production, or penalties for shirking.<sup>154</sup> Indeed, where

149. Bregm, *supra* note 102, at 29. This effect may occur regardless of general compensation; for example, Uri Gneezy and Aldo Rustichini found that when subjects were paid a participation fee for attending an experiment, and then offered an additional per-question amount to answer questions, the per-question fee became their reference point for deciding how much effort to exert. Subjects paid less per question exerted less effort, compared to those paid more. Participants who were not paid an additional amount per question, however, often believed it was "their side of the bargain" to answer the questions in return for the participation fee alone. Uri Gneezy & Aldo Rustichini, *Pay Enough or Don't Pay At All*, 115 Q.J. ECON. 791, 803 (2000).

150. Cerasoli et al., *supra* note 87, at 4; *see also* Deci et al., *supra* note 62, at 628-29 (contending that rewards are most likely to affirm competence where rewards are linked to performance and the person receives an award signaling excellent performance). Frey suggests this may be one reason why incentive pay is used more frequently for managers than for lower-level employees, as managers may be more likely to understand incentives as performance feedback. Frey, *supra* note 44, at 437.

151. *See* Bénabou & Tirole, *supra* note 141, at 494; Bruna Bruno, *Rewarding My Self: The Role of Self Esteem and Self Determination in Motivation Crowding Theory* 5 (MPRA Paper No. 23117, 2010), [http://mpra.ub.uni-muenchen.de/23117/1/MPRA\\_paper\\_23117.pdf](http://mpra.ub.uni-muenchen.de/23117/1/MPRA_paper_23117.pdf) ("When individuals perceive an external intervention as reducing their self determination, intrinsic motivation is substituted by external control.").

152. Frey & Jegen, *supra* note 2, at 594; *see also* Benkler, *supra* note 2 (describing both Frey's and Bénabou and Tirole's conceptions of the self-esteem theory).

153. *See* Armin Falk & Michael Kosfeld, *The Hidden Costs of Control*, 96 AM. ECON. REV. 1611, 1623 (2006).

154. *See id.*; Bénabou & Tirole, *supra* note 47, at 1655; Ernst Fehr & Armin Falk, *Psychological Foundations of Incentives*, 46 EUR. ECON. REV. 687, 698 (2002); Fehr & Fischbacher,



agents can interpret explicit incentives—prizes, rules, or penalties—as a signal of distrust, they may exert more effort under incomplete contract than when provided with clear incentives.<sup>155</sup> This result aligns with Bruno Frey’s theory of why corporations rarely seek bondage commitments from their employees; a principal that requires agents to make such commitments reveals that he distrusts the employee, which crowds out intrinsic motivations and skews the selection pool of job applicants.<sup>156</sup> Frey has also commented on this dynamic in relation to public laws, in that laws that “impl[y] a fundamental distrust of its citizens and seek[] to discipline them tend[] to crowd out civic virtue and undermine[] the support which citizens are prepared to exert towards the basic law.”<sup>157</sup>

### c. Incentives Can Reveal the Principal’s Values

Incentives can signal information about the principal’s character, including that the principal is selfish, hostile,<sup>158</sup> immoral, or even in violation of an implicit contract to behave reciprocally.<sup>159</sup> This is particularly true of fines as compared to rewards;<sup>160</sup> and one economic game study has shown that agents interpret

*supra* note 18, at C22 (noting that incentives may create “a hostile atmosphere of threat and distrust”); Ernst Fehr & Simon Gächter, *Fairness and Retaliation: The Economics of Reciprocity*, 14 J. ECON. PERSP. 159, 177 (2000); Ernst Fehr & John A. List, *The Hidden Costs and Returns of Incentives—Trust and Trustworthiness Among CEOs*, 2 J. EUR. ECON. ASS’N 743 (2004); Ernst Fehr & Bettina Rockenbach, *Detrimental Effects of Sanctions on Human Altruism*, 422 NATURE 137 (2003); Frey, *supra* note 68 (referring to this dynamic as the “misattribution effect”); Gneezy et al., *supra* note 77, at 192; Florian Herold, *Contractual Incompleteness as a Signal of Trust*, 68 GAMES & ECON BEHAV. 180 (2010). *But see* Mary Rigdon, *Trust and Reciprocity in Incentive Contracting*, 70 J. ECON. BEHAV. & ORG. 93 (2009) (finding that reward and punishment at low levels are no different from the absence of incentives, that the availability of high punishment can induce investors to invest more, and that returns are highest when the investor can provide a large reward).

155. See Epstein, *supra* note 112, at 40–41 (noting that less specific contracts may improve agent performance because they avoid “giv[ing] the agent the impression of a lack of trust”); Falk & Kosfeld, *supra* note 153, at 1613 (noting that many contracts are deliberately left incomplete as a means of signaling trust in the agent); Herold, *supra* note 154. An incomplete contract also leaves open the potential for additional rewards or fines, which may exert its own motivational effect. *See, e.g.*, Uri Gneezy & Aldo Rustichini, *A Fine Is a Price*, 29 J. LEGAL STUD. 1, 10 (2000).

156. Frey, *supra* note 44, at 437; Frey, *supra* note 68, at 1530–31.

157. Frey, *supra* note 43, at 1048.

158. *See, e.g.*, Ellingsen & Johannesson, *supra* note 30, at 991 (noting that principals who do not exert control over agents “signal[] a prosocial attitude,” and that although this makes no difference to a selfish agent, “a prosocial agent has a stronger desire to make a good impression on a prosocial principal than on a selfish one”); Herold, *supra* note 154, at 188 (presenting an economic model in which complete contracts providing for incentives signal distrust).

159. Frey, *supra* note 68, at 1526. A more recent modeling study makes an analogous claim that motivational crowding-out is more likely among agents who are already specifically motivated to advance the principal’s interests. Harvey S. James, Jr. *Why Did You Do That? An Economic Examination of the Effect of Extrinsic Compensation on Intrinsic Motivation and Performance*, 26 J. ECON. PSYCHOL. 549, 562 (2005) (presenting an economic model suggesting that agents may perceive incentives to be controlling when they are offered by employers or when they are too large).

160. Fehr & Falk, *supra* note 154, at 695 (using results of gaming studies to show that incentives can interact with intrinsic enjoyment, reciprocity, and image motivations). Agents may also perceive small incentives to convey that the principal is insulting them. Crowding-out may depend on the size of the incentive, with the potential interpretation that “people simply find it demeaning to work for a small amount of money.” Kamenica, *supra* note 142, at 13.10.

manifestations of control as evidence of the principal's hostility, diminishing reciprocity motivations to perform well.<sup>161</sup> Lars Feld and Bruno Frey's model of tax compliance suggests that when tax administrators treat taxpayers in an authoritarian or arbitrary way, they violate an implicit "psychological tax contract" committing to mutual respectful treatment, which undermines reciprocity and increases evasion by taxpayers.<sup>162</sup> Other studies have had similar findings,<sup>163</sup> and some have suggested that stiff punishments may even activate contrarian behavior due to spite.<sup>164</sup> Indeed, individuals may go out of their way to punish people who are perceived as violating reciprocity norms, even when the act of punishing is costly.<sup>165</sup>

Researchers have named several specific crowding-out effects attributable to signals about principals' values, including moral repugnance (which rewards the signal that the principal has bad moral values) and the eviction effect (which rewards the signal that the principal manages resources poorly). Antoine Beretti and colleagues have suggested that positive incentives, particularly financial rewards for altruistic behavior, may crowd out motivation by activating "moral repugnance."<sup>166</sup> This effect arises when a principal introduces incentives in "a territory previously immune to [] market forces," causing agents to disengage out of distaste.<sup>167</sup> Uri Gneezy and colleagues illustrate with the example of asking an attractive person for sex, and then offering to pay \$20 for the privilege: "only a certain kind of economist would expect [his or her] partner to be happier" under these conditions.<sup>168</sup> Govind Persad has suggested the same dynamic for offering

161. Falk & Kosfeld, *supra* note 153, at 1625-26 (using results of a laboratory gaming experiment to show that agents viewed the principal to be controlling when the principal restricted their choices, and that a majority of agents reacted negatively to this behavior by penalizing the principal, resulting in lower returns).

162. Lars P. Feld & Bruno S. Frey, *Tax Compliance as the Result of a Psychological Tax Contract: The Role of Incentives and Responsive Regulation*, 29 LAW & POL'Y 102, 104 (2007).

163. See, e.g., Fehr & Gächter, *supra* note 154, at 170-72 (finding lower effort among reciprocity-motivated agents when incentives were used); Ernst Fehr & Simon Gächter, *Do Incentive Contracts Undermine Voluntary Cooperation?* (Zurich IEER, Working Paper No. 34, 2002), <http://ideas.repec.org/p/zur/iewwpx/034.html> (finding that explicit performance incentives and a fine for shirking reduced agents' average effort).

164. Gneezy, *supra* note 77, at 24.

165. See, e.g., Bowles, *supra* note 41, at 1608 (citing studies); Ernst Fehr & Simon Gächter, *Cooperation and Punishment in Public Goods Experiments*, 9 AM. ECON. REV. 980 (2000) (finding in a public good contributions game that there is heavy punishment of free riders even when punishment is economically costly); Fehr & Falk, *supra* note 154, at 706; Fehr & Gächter, *supra* note 154, at 159; Fehr & Rokenbach, *supra* note 154, at 13.

166. Beretti et al., *supra* note 37, at 67. This explanation recalls Yuval Feldman and Oren Perez's discussion of motivational crowding-out when there is "a mismatch between . . . subjects' moral views and a certain regulatory instrument . . . lead[ing] to behavior reversal." Feldman & Perez, *supra* note 27, at 411.

167. Beretti et al., *supra* note 37, at 68 (using a field survey to show that monetary payments diminish participation in an environmental survey).

168. Gneezy et al., *supra* note 77, at 201; Gneezy, *supra* note 77, at 6; see also BENKLER, *supra* note 50, at 97 (noting the social impropriety of offering payment after sex, or of paying friends for hosting dinner in their home).

teenagers money to take the morning-after pill.<sup>169</sup> One empirical study offered traditional healers a cash incentive to refer patients with certain ulcers to the public health service; although healers in one country found it acceptable, those in a nearby country revolted against the incentive as an effort to “pay for diseased bodies.”<sup>170</sup> Moral repugnance has also been raised as an objection to paying research subjects<sup>171</sup> and paying environmentalists to participate in an environment-related survey.<sup>172</sup>

The “eviction effect” occurs when the principal is a charity, or is otherwise using resources to achieve goals that are meaningful to the agent.<sup>173</sup> If the incentive is costly, then the agent may infer that the incentive depletes the principal’s resources to do good work. The agent will then be less motivated to act in a way that triggers the payout.<sup>174</sup> Some support for this hypothesis arises from a set of interviews with Australian blood donors, who worried that “the introduction of a cash incentive would be costly to the Blood Service and that these resources would be better spent on operational needs.”<sup>175</sup> These donors held similar views about other tangible rewards and in some cases would not accept gifts unless they were donated or industry-sponsored, instead of purchased using charity funds.<sup>176</sup>

When different incentives are offered across a group of agents, motivational crowding-out may also be a response to perceived unfairness of the principal across different agents. Simulation studies and field research have found that employees reduce their effort if they know they are being paid less than a co-worker.<sup>177</sup> Notably, one study found that this reduction in effort was less severe when participants were told that wages were randomly generated (not decided by the principal);<sup>178</sup> this suggests that the driving force behind the motivational effect is how agents interpret the incentive as information about the principal.

---

169. Govind Persad, *Libertarian Patriarchalism: Nudges, Procedural Roadblocks, and Reproductive Choice*, 35 WOMEN’S RTS. L. REP. 273, 292-93 (2014) (suggesting that “it may seem repugnant to take [the pill] in order to receive \$50—and the offer of money may convert” the decision from acceptance to refusal).

170. Priscilla Magrath & Mark Nichter, *Paying for Performance and the Social Relations of Health Care Provision: An Anthropological Perspective*, 75 SOC. SCI. MED. 1778, 1783 (2012).

171. Martin Wilkinson & Andrew Moore, *Inducements Revisited*, 13 BIOETHICS 114, 122-27 (1999) (summarizing and rejecting this argument).

172. Beretti et al., *supra* note 37, at 64.

173. Bénabou & Tirole, *supra* note 47, at 1658.

174. *Id.*

175. Danielle Chmielewski-Ralmondo et al., *A New Perspective on the Incentive-Blood Donation Relationship: Partnership, Congruency, and Affirmation of Competence*, 52 TRANSFUSION 1889, 1894-95 (2012).

176. *Id.*

177. Bregm, *supra* note 102, at 30.

178. *Id.* at 30 (citing Simon Gächter & Christian Thöni, *Social Comparison and Performance: Experimental Evidence on the Fair Wage-Effort Hypothesis*, 76 J. ECON. BEHAV. & ORG. 531 (2010)).

## d. Incentives Can Reveal the Supportiveness of Social Norms

An incentive may also signal information about the way that other agents behave.<sup>179</sup> This effect may work in two opposing ways. When the incentive suggests that social norms run counter to the principal's priorities, this may induce crowding-out; but when the incentive exerts an "expressive" effect that announces social norms that align with the principal's priorities, this may conversely induce crowding-in.

Agents may interpret an incentive (particularly negative) as evidence that intrinsic motivation is insufficient—that is, incentives are necessary because social norms do not support the task.<sup>180</sup> Agents who seek to conform to or reciprocate social norms will therefore update their own preferences with reduced intrinsic or reputational motivation to engage. Dan M. Kahan has noted, for example, that when the government conspicuously raises penalties for tax evasions, this signal "causes individuals to infer that more taxpayers than they thought are choosing to evade. This inference, in turn, triggers a reciprocal motive to evade. . . ."<sup>181</sup> When signs beseech National Park visitors to refrain from stealing park artifacts, they may encourage theft by communicating that many tourists steal.<sup>182</sup> The escalation in bad behavior could be due to conformism (people want to be like others), or reciprocity motivations vis-à-vis social norms—people are "afraid of being chumps and are thus unwilling to comply unilaterally."<sup>183</sup>

Conversely, an incentive might also have an "expressive" effect that operates to align social norms with the principal's priorities, thereby crowding *in* or augmenting intrinsic and reputational motivation. A number of legal

179. Bowles & Hwang, *supra* note 42, at 6 (citing additional references); Dan M. Kahan, *supra* note 124, at 79 (2003) ("Incentives do more than affect individuals' calculations of the costs and benefits of particular forms of conduct; they also shape their impressions of the attitudes and intentions of those around them."); Dan M. Kahan, *Social Influence, Social Meaning, and Deterrence*, 83 VA. L. REV. 349 (1997).

180. See, e.g., Wendy J. Gordon, *Discipline and Nourish: On Constructing Commons*, 95 CORNELL L. REV. 733, 749 n.76 (2010); Wendy J. Gordon, *Of Harms and Benefits: Torts, Restitution, and Intellectual Property*, 21 J. LEGAL STUD. 449, 457 (1992) ("[P]laying people to refrain from doing harm is likely to encourage precisely the wrong sorts of behavior . . . [M]oral people might (inaccurately) infer that one has no moral obligation to do the right thing unless one is paid."); Adrian Vermeule, *Precautionary Principles in Constitutional Law*, 4 J. LEGAL ANALYSIS 181, 207 (2012).

181. Kahan, *supra* note 19, at 342; see also Michael Doran, *Tax Penalties and Tax Compliance*, 46 HARV. J. ON LEGIS. 111, 112, 133-37 (2009); Fleisig-Greene, *supra* note 132, at 1212-13; Galle, *supra* note 63, at 834-35 (citing research by Cass Sunstein to suggest that "one possible symbolic implication of paying actors to do good is that the 'expected' or default state is that they have no such obligations"); Stephen W. Mazza, *Taxpayer Privacy and Tax Compliance*, 51 U. KAN. L. REV. 1065, 1079-80 (2003) (rejecting this theory on the basis that penalties may be viewed as "just deserts" for "dishonest outliers" who evade taxes). Similarly, Frank Cross has examined (and ultimately rejected) the analogous claim that law itself can crowd out trust in society because "the protective nature of law signals that trust is absent from the relationship." Frank B. Cross, *Law and Trust*, 93 GEO. L.J. 1457, 1498-1500 (2005); see also Bowles, *supra* note 73, at 73 (attributing this view to economist Robert Lucas).

182. See IAN AYRES, *CARROTS AND STICKS* 79 (2010) (discussing the work of Robert Cialdini).

183. Vermeule, *supra* note 180, at 207; see also Vermeule, *supra* note 43, at 424-25.

scholars including Robert Cooter, Dan Kahan, and Cass Sunstein have written extensively on the expressive effects of law, suggesting that some laws “tip[] the balance in favor of informal enforcement” even when formal enforcement is rare (depending on the extent to which the public internalizes the norm).<sup>184</sup> This may be illustrated by Patricia Funk’s study of voter turnout in Switzerland. When several Swiss jurisdictions *reversed* their policy of fining nonvoters, these jurisdictions actually experienced a decline in voting, and the impact was more severe in jurisdictions that previously had a high voter turnout.<sup>185</sup> Funk interpreted this as evidence that the fine had previously had an expressive effect, inducing citizens to vote due to “civic duty or fear of social sanctions.”<sup>186</sup> Removing the fine, therefore, may have expressed that voting was no longer as important.

Crowding-out may also occur when agents interpret the size of an incentive to signal the magnitude of social approval or disapproval associated with a behavior. When a penalty is small, it is “a signal that the social damage is not as high as she initially believed,” causing agents to escalate bad behavior.<sup>187</sup> This explanation may be supported by Uri Gneezy and Aldo Rustichini’s study from an Israeli daycare center, which imposed a small fine for parents who picked up their children late.<sup>188</sup> After establishing the fine, the daycare center experienced an *increase* in the frequency of late pickups. This effect may be explained by several different signaling mechanisms. One may be that the fine signaled to parents that late pickups were common and that social norms about late pickups were permissive. The small size of the fine might also have conveyed that even the daycare center considered late pickups to be inconsequential. The incentive thus influenced behavior not only via the desire to conform to others’ actions, but also via reduced reputational motivations to pick up children on time.<sup>189</sup>

184. Robert D. Cooter, *Three Effects of Social Norms on Law: Expression, Deterrence, and Internalization*, 79 OR. L. REV. 1, 10-11 (2000) [hereinafter Cooter, *Three Effects*]; see also Robert Cooter, *Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms*, 86 VA. L. REV. 1577 (2000); Robert Cooter, *Expressive Law and Economics*, 27 J. LEGAL STUD. 585 (1998); Alex Geisinger, *A Belief Change Theory of Expressive Law*, 88 IOWA L. REV. 35 (2002); Hollander-Blumoff, *supra* note 39, at 62-63 (describing how law’s expressive effect and “the moral judgment that law carries[] may provide protection against crowding out”); Dan M. Kahan, *What Do Alternative Sanctions Mean?*, 63 U. CHI. L. REV. 591 (1996); Eric A. Posner, *Law and Social Norms: The Case of Tax Compliance*, 86 VA. L. REV. 1781 (2000); Cass R. Sunstein, *On the Expressive Function of Law*, 144 U. PA. L. REV. 2021 (1996); Feldman, *supra* note 2, at 16-22 (describing literature on competing models for legal compliance, including the expressive effect of law).

185. Patricia Funk, *Is There an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines*, 9 AM. LAW & ECON. REV. 135, 138 (2007). Funk cites similar evidence of effective small incentives, such as fining company directors for missing meetings or fining players in public goods games. *Id.* at 139.

186. *Id.* at 138.

187. Bolle & Otto, *supra* note 39, at 18 (suggesting that this mechanism may explain the daycare study).

188. Gneezy & Rustichini, *supra* note 155, at 14.

189. Fehr & Falk, *supra* note 154, at 709.

## e. Incentives Can Reframe Choices as Self-Interested

Apart from conveying information about tasks, agents, principals, and social norms, incentives may also shift agents' point of view by signaling that self-interested behavior is appropriate.<sup>190</sup> This phenomenon has variously been called "framing" or shifts in "decisional mode." Incentives reframe tasks as transactions with the possibility for selfish benefit, which may undermine valuable other-regarding motivations such as altruism or a sense of fiduciary duty. In effect, offering an incentive shifts the "decisional mode" from moral reasoning to exchange.<sup>191</sup> Another way to describe this phenomenon is that incentives "change[] . . . the relationship between the principal and the agent,"<sup>192</sup> yielding a more transactional interaction. This phenomenon is particularly associated with financial incentives, as even small amounts transform tasks into market transactions.<sup>193</sup> These "subliminal effects"<sup>194</sup> resonate with psychological research showing that even *thinking* about money can make individuals less likely to make and respond to requests for help.<sup>195</sup> Framing effects are particularly problematic for actions that rely on relationships to provide intrinsic motivation (for example, altruism and reciprocity), and it aligns with the observation that crowding-out may be exacerbated when incentives are introduced despite a close relationship between the principal and agent.<sup>196</sup>

A large body of empirical data provides support for framing effects. For example, James Heymann and Dan Ariely have demonstrated that in social-market relationships (wherein effort is shaped by altruism), the "mere mention of monetary payment" can transform perceived relationships into money-market

---

190. Bowles, *supra* note 41, at 1606; *see also* Bowles & Hwang, *supra* note 42 (citing additional studies supporting this view); Gneezy, *supra* note 77, at 25 (describing a shift from a "communal" to an "exchange" interaction when incentives are present).

191. *See, e.g.*, STEPHAN MEIER, *THE ECONOMICS OF NON-SELFISH BEHAVIOUR: DECISIONS TO CONTRIBUTE MONEY TO PUBLIC GOODS* 35 (2006). Schnedler and Vanberg have also noted that offering a reward makes a behavior "tradeable" in social exchange. Wendelin Schnedler & Christoph Vanberg, *A Rationale for Motivational Crowding Out*, 1-2 (Discussion Paper Series in Economics and Management, Paper No. 10-09, 2011), <http://geaba.de/DP/DP-10-09.pdf>.

192. Kamenica, *supra* note 142, at 13.9.

193. *See, e.g.*, Gneezy, *supra* note 77, at 27 (describing a W-shaped curve of incentives, where behavior is higher with no incentive than it is with either a small fine or a small reward—"[E]ven the smallest amount of extrinsic motivation can destroy the intrinsic motivation completely."); Gneezy & Rustichini, *supra* note 149 (describing their finding that a group of subjects offered a tiny amount to answer questions performed worse than those offered nothing); Gneezy & Rustichini, *supra* note 155, at 15 (reiterating this finding); Feldman & Lobel, *supra* note 109, at 47 (noting in the context of whistleblowing activity that "the introduction of money can reclassify action as an economic, rather than pro-social, behavior and crowds out ethical intentions").

194. Censolo et al., *supra* note 49, at 8.

195. Kathleen D. Vohs et al., *The Psychological Consequences of Money*, 314 *SCIENCE* 1154, 1154 (2006); *see also* Feldman, *supra* note 2, at 23-24 (describing research on the priming effect of money).

196. *See* Frey, *supra* note 43, at 1046; Stout, *supra* note 7, at 552-53 (noting that incentives "change[] the social context," signaling "that the employer . . . views the employment relationship as an arm's length exchange in which self-interested behavior is appropriate, expected, and even encouraged").

exchanges (wherein effort is shaped by compensation).<sup>197</sup> Some have suggested that corporate ethics programs may replace moral motivations with extrinsic incentives for ethical behavior; one study has found that individuals facing decisions in conditions with compliance programs are more likely to view those decisions as cost-benefit “business” choices, rather than ethical choices, because incentives “put[] a price tag on behaving ethically.”<sup>198</sup> Framing effects may also explain the Israeli daycare study described above: “the imposition of a price [for lateness] conveyed the message that the commodity of ‘being late’ could now be bought,”<sup>199</sup> transforming parents’ perception of late pickups from an imposition into a service that could be fully paid for (and absolved) by the fine.<sup>200</sup> Another example may be the American Association of Retired Persons’ transactions with lawyers, who refused to serve retirees at the low cost of \$30 per hour, but agreed to provide services for free.<sup>201</sup> “Once it was clear they were being asked to engage in a charitable activity rather than a market transaction, the lawyers responded charitably.”<sup>202</sup>

In another example, Rebecca Hollander-Blumoff suggests that tort law damages may reframe the act of harming others as a commodity that tortfeasors can purchase, thereby lessening the motivation to avoid harming others for moral reasons.<sup>203</sup> Funk suggests a similar explanation for reduced voting rates in a Swiss jurisdiction that imposed a new fine on nonvoting; the fine allowed citizens “to ‘buy’ not voting and removed . . . feelings of guilt.”<sup>204</sup> Another Swiss example arises from the study offering residents incentives for accepting a nuclear waste facility in their town.<sup>205</sup> This incentive may have signaled to the residents they should reason based on self-interest rather than on the public good; when they considered their own self-interest, the payment was insufficient to compensate for the facility’s risks.<sup>206</sup> Similarly, in studies of contribution to

197. James Heymann & Dan Ariely, *Effort for Payment: A Tale of Two Markets*, 15 PSYCHOL. SCI. 787, 780, 792 (2004).

198. Maurice E. Stucke, *In Search of Effective Ethics & Compliance Programs*, 39 J. CORP. L. 769, 823 (2014) (citing MAX H. BAZERMAN & ANN E. TENBRUNSEL, BLIND SPOTS: WHY WE FAIL TO DO WHAT’S RIGHT AND WHAT TO DO ABOUT IT (2011)).

199. Fehr & Falk, *supra* note 154, at 709.

200. Barry Schwartz and Kenneth Sharpe analogize this to parking illegally on the street where a parking ticket is \$25, but a garage spot costs \$30; “you’re not doing the ‘wrong’ thing; you’re doing the economical thing.” BARRY SCHWARTZ & KENNETH SHARPE, PRACTICAL WISDOM: THE RIGHT WAY TO DO THE RIGHT THING 190-91 (2010).

201. SANDEL, *supra* note 72, at 121 (citing research by Dan Ariely).

202. *Id.*

203. Hollander-Blumoff, *supra* note 39, at 61.

204. Funk, *supra* note 185, at 152-53.

205. Frey & Oberholzer-Gee, *supra* note 85.

206. For an explanation of this effect, see Barry Schwartz, *Money for Nothing*, N.Y. TIMES, July 2, 2007, at A21. Ultimately, the study authors determined that the crowding-out effect was strongest among individuals who had indicated general support for nuclear technology as socially beneficial, suggesting that the offer of a financial incentive specifically reduced “civic spirit” among those who previously supported the project. Frey & Oberholzer-Gee, *supra* note 85, at 753-54; see also SCHWARTZ & SHARPE, *supra* note 200, at 192-93. This explanation may fall under several of the mechanisms described here, including priming, image-spoiling, or moral repugnance. Similar dynamics

public goods (or non-exploitation of a common resource), incentives have been ineffective in encouraging socially-optimal behavior, perhaps because they transformed contributing from a social act into “an individual choice motivated . . . by private benefits.”<sup>207</sup>

## 2. Agent Signaling to Observers: Image-Spoiling

Incentives may also cause crowding-out when they interfere with a second set of signals: the signals agents send to observers by their behavior. Social capital is a critical resource, more valuable in some contexts than financial rewards,<sup>208</sup> and incentives can interfere with agents’ opportunities to increase their social standing through activities that are prosocial, altruistic, or moral (e.g., contributing to charity, performing professional jobs,<sup>209</sup> or voting<sup>210</sup>). That is, incentives can crowd out reputational motivation.<sup>211</sup> When observers can attribute an agent’s behavior to greed, the incentive “spoils”<sup>212</sup> or weakens the good-character signal sent by his acts<sup>213</sup>—or it may even send a bad-character signal by marking him as greedy. If the agent worries that others will see him as simply “doing well” [for himself] instead of “doing good,” his reputational motivation will diminish.<sup>214</sup> For example, someone recycling bottles may ordinarily be seen as environmentally responsible; if others know that she

---

may also explain the results of a U.S.-based study, which found that tax rebates did not increase Nevada residents’ willingness to accept a nuclear waste facility in their town. Frey, *supra* note 43, at 1047 (citing work by Howard Kunreuther and Douglas Easterling).

207. Andreas Fuster & Stephan Meier, *Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement* 24 (Fed. Res. Bank of Boston, Working Paper No. 09-2, 2009), <http://www.econstor.eu/bitstream/10419/55597/1/593927435.pdf>; see also Bowles, *supra* note 41, at 1606-07 (referencing a similar study).

208. See BENKLER, *supra* note 50, at 95-96. This social approval is valuable for its own sake, or it can be used to generate material benefits later. See Fehr & Falk, *supra* note 154, at 705; see also Ariely et al., *supra* note 48, at 544 (noting that social capital is tradeable for “future extrinsic rewards, like a political career or admission to a college”); Fleisig-Greene, *supra* note 132, at 1211-12 (interpreting scholarship by Eric A. Posner to conclude that an individual who visibly abides by social norms “demonstrates [to others] that he is a long-term type, desirable as a partner” for commerce).

209. See, e.g., Eijkenaar et al., *supra* note 25, at 116 (noting “reputational incentives” for physician performance compared to their peers).

210. See, e.g., Patricia Funk, *Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System*, 8 J. EUR. ECON. ASS’N 1077 (2010) (identifying the role of social incentives in voting behavior).

211. See Ariely et al., *supra* note 48 (presenting two experiments consistent with this theory); Bénabou & Tirole, *supra* note 47 (explaining an economic model of individual behavior incorporating intrinsic, extrinsic, and reputational motivations); Gneezy et al., *supra* note 77, at 192-93 (suggesting that incentives can cause prosocial behaviors to signal “greediness” rather than “prosocial preferences,” thereby depleting agents’ image motivations).

212. Bénabou and Tirole have named this the “image-spoiling effect of rewards,” which I follow here. Bénabou & Tirole, *supra* note 47, at 1655. Frey has called this “impaired expression possibility.” FREY, *supra* note 2, at 17.

213. Bénabou & Tirole, *supra* note 47, at 1654; see also Dan M. Kahan, *supra* note 179, at 351 (“Individuals also draw such inferences [perceptions about social valuations] from the behavior of other individuals; when the law regulates such behavior, it can either accentuate or mute these signals.”).

214. Ariely et al., *supra* note 48; see also Brian Galle, *Tax, Command . . . or Nudge?: Evaluating the New Regulation*, 92 TEX. L. REV. 837, 881 (2014); Kamenica, *supra* note 142, at 13.



receives rebates for each can, however, she may be seen as simply “cheap.”<sup>215</sup> The same is true of legal incentives, such as increased penalties for tax nonpayment; when legal penalties require compliance, rule followers can no longer signal their intrinsic commitment to social norms through their behavior.<sup>216</sup> (That is, when there is “order without law,” there is some concern that imposing a law may degrade the order.) As Yochai Benkler has noted, “for any given culture, there will be some acts that a person would prefer to perform not for money, but for social standing, recognition, and probably, ultimately, instrumental value, obtainable only if that person has performed the action through a social, rather than a market, transaction.”<sup>217</sup>

Several empirical tests have demonstrated how incentives may undermine reputational motivations. In one study, students who collected money for charity collected less when they could keep a percentage of their collections, which lessened the basis for social approval of their actions.<sup>218</sup> In another test, undergraduates were assigned to press keyboard keys that controlled donations to two charities, one “good” (the American Red Cross) and one “bad” (the National Rifle Association) according to social norms on the campus.<sup>219</sup> Individuals were randomly assigned to charities, and then further randomized to perform in private or with a peer audience, with or without compensation based on their performance. In private, individuals assigned to both causes pressed significantly more keys when they were incentivized, compared to their effort without the incentive. In front of peers, however, individuals pressing keys for a “good” cause actually pressed fewer keys when the task was incentivized, while those assigned to the “bad” charity did not perform differently when the incentive was offered. A replication study asking undergraduates to bike for “good” or “bad” charities had similar findings.<sup>220</sup> These results suggest that the public visibility of a task matters to the design of an incentive scheme, as does the visibility of the incentive. When reputational motivations loom large, “crowding out can . . . mak[e] all but very large rewards inferior to [no rewards].”<sup>221</sup>

Conversely, negative incentives such as fines may have the effect of reducing anticipated social disapproval for undesirable activity. This may also be an explanation for the crowding-out effects of the Israeli daycare study;<sup>222</sup> when parents knew that others would be aware they had paid for their tardiness, this payment reduced or eliminated their anticipated social disapproval consequences. A similar dynamic may be at play in the case of the Boston Fire

---

215. Gneezy, *supra* note 77, at 5.

216. See Posner, *supra* note 184, at 1791.

217. BENKLER, *supra* note 76, at 96.

218. Gneezy & Rustichini, *supra* note 149. For a description of this study, see Fehr & Falk, *supra* note 154, at 709-10.

219. Ariely et al., *supra* note 48.

220. *Id.* at 552-53.

221. Bénabou & Tirole, *supra* note 47, at 1662.

222. Gneezy & Rustichini, *supra* note 155.

Commissioner's decision to dock firemen's pay when they called in sick too often; sick calls doubled during the year after this change.<sup>223</sup>

### 3. Agent Self-Signaling: Attribution and Self-Image-Spoiling

A third set of signals are those that the agent sends to himself, and accepting a reward can change the agent's perception of his own motivation. People draw conclusions about their own motives based on their actions; when an agent can look back at his behavior and ascribe it to an incentive, he may subsequently view the task as less intrinsically valuable,<sup>224</sup> and he is less likely to internalize his behavior as a norm.<sup>225</sup> This is also known as "attribution;" attributional explanations for motivational crowding-out focus on individuals' perceptions of their *own* motives, rather than perceptions of control or of a principal's motives.<sup>226</sup> Larger, more salient incentives may be more likely to influence self-signaling.<sup>227</sup> For example, studies have found that people who are paid *less* to engage in experimental tasks, including boring tasks, are subsequently *more* likely to report finding the task intrinsically interesting, compared to participants paid larger amounts.<sup>228</sup>

---

223. See Sung-Ha Hwang & Samuel Bowles, Are Incentives Overused in Cases Where They Crowd out Prosocial Motivations? (Nov. 24, 2010) (unpublished manuscript), <http://www.hec.unil.ch/documents/seminars/deep/358.pdf> (citing Douglas Belkin, *Boston Firefighters Sick—or Tired of Working*, BOS. GLOBE, Jan. 18, 2002, at B1).

224. See Deci et al., *supra* note 62, at 630; see also Katharine T. Bartlett, *Showcasing: The Positive Spin*, 89 N.C. L. REV. 1055, 1066 (2011) (describing how people infer their own motives from the actions they take, and noting that people who complete tasks for no incentive are more likely to rate those tasks as pleasant, compared to those who complete the same tasks for rewards). This concept has also been extended to whether individuals perceive their prior behaviors to be trustworthy when those actions were constrained by a contract. See Brent Simpson & Kimmo Eriksson, *The Dynamics of Contracts and Generalized Trustworthiness*, 21 RATIONALITY & SOC'Y 59 (2009). This has also been described as part of the "overjustification" effect—when we can attribute our behavior to external causes that "overjustify" our actions, we tend to overlook our intrinsic motives. See, e.g., Sherry F. Colb, *Oil and Water: Why Retribution and Repentance Do Not Mix*, 22 QUINNIPIAC L.R. 59, 81, 83 (2003) (suggesting that punishment leads criminal offenders to attribute their criminal acts to an intrinsic enjoyment of crime, thereby strengthening their criminal motivations); Reed Elizabeth Loder, *Tending the Generous Heart: Mandatory Pro Bono and Moral Development*, 14 GEO. J. LEGAL ETHICS 459, 472-74 (2001) (suggesting that viewing oneself as externally motivated to engage in volunteer work may undermine "psychic satisfaction" and reduce volunteering behavior). This has also been described as a form of overjustification. See, e.g., James, *supra* note 159, at 553; Winick, *supra* note 44, at 768-69, 810 (describing overjustification effects through self-signaling in the presence of either incentives or coercion).

225. See Fleisig-Greene, *supra* note 132, at 1213.

226. See Deci et al., *supra* note 62, at 628, 654. One legal scholar also ascribed this effect to "the discounting principle," which suggests that "when there is more than one apparent cause of our behavior, we will discount the causal priority of any particular cause." Mark S. Sobus, *Mandating Community Service: Psychological Implications of Requiring Prosocial Behavior*, 19 L. & PSYCHOL. REV. 153, 160 (1995).

227. James, *supra* note 159, at 553 (citing research suggesting that "reward size is negatively related to intrinsic motivation").

228. See Hanson & Yosifon, *supra* note 84, at 108-09 (describing the results of a 1959 study that asked participants to complete a boring task for money, tell others that the task was interesting, and then state how they actually felt about the task; those paid \$1 subsequently reported much higher

When an activity has a social valence, such as charitable giving, blood donation, or recycling, this attributional effect may produce *self*-image-spoiling, an analogue to the image-spoiling mechanism described above. Our desire for a positive self-image may indeed drive some prosocial acts, particularly when behavior is anonymous,<sup>229</sup> and we often look at our actions as “diagnostic of [our] preferences,” such as preferences for altruism.<sup>230</sup> For example, when research subjects take part in an incentivized study, they subsequently view themselves as less altruistic compared to subjects who were not paid.<sup>231</sup> Other scholars have referred to the positive self-image feedback loop as “impure altruism,”<sup>232</sup> the “warm glow,”<sup>233</sup> or the “joy of giving” associated with prosocial behavior.<sup>234</sup> But if our good deeds are incentivized, the incentive may affect our ability to signal good character to ourselves, even when no one else is watching.<sup>235</sup> This may dampen the self-image benefits we expect from our good deeds, reducing self-image motivation.<sup>236</sup> This theory resonates with qualitative research with blood donors, which suggests that financial compensation “pollutes” altruism.<sup>237</sup>

We also send signals to ourselves through acts and omissions that cause guilt, and guilt aversion can often be a powerful motivation. Negative incentives can disrupt signals that previously induced guilt, reducing self-image motivations for good behavior. This provides yet another explanation for the daycare study. When parents could pay for their transgression of picking up their children late, internalizing the costs they had imposed on the providers, they “may no longer have felt bad.”<sup>238</sup> Another study incentivized members of a group to contribute to a public good and permitted them to punish “free-riding” (noncontributing) group members; free riders were less sensitive to punishment

---

levels of interest, suggesting that they “subconscious[ly] alter[ed] their attitudes” to make their perceptions consonant with their statements).

229. Bénabou & Tirole, *supra* note 47, at 1653 (citing several studies that “confirm the importance of such self-image concerns in explaining prosocial behavior in anonymous settings”).

230. *Id.* at 1657 n.11.

231. *See, e.g.,* Sobus, *supra* note 226, at 171-72 (summarizing this research).

232. Frey, *supra* note 98, at 410 (2008) (citing James Andreoni, *Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving*, 100 *ECON. J.* 464 (1990)).

233. Alex Imas, *Working for the “Warm Glow”: On the Benefits and Limits of Prosocial Incentives*, 114 *J. PUB. ECON.* 14 (2014).

234. Bénabou & Tirole, *supra* note 47, at 1654.

235. *See id.* at 1657 (“Later on . . . information [about intrinsic incentives] may no longer be perfectly ‘accessible’ in memory—in fact, there will often be strong incentives to recall it in a self-serving way. Actions, by contrast, are much easier to remember than their underlying motives.”).

236. *See* Vermeule, *supra* note 43, at 424-25 (“[T]he provision of the reward reduces the utility that public-spirited actors derive from performing it, if the reward suggests that the behavior is motivated by venality rather than altruism.”).

237. Chmielewski-Ralmondo et al., *supra* note 175, at 1894.

238. Fehr & Falk, *supra* note 154, at 709; *see also* AYRES, *supra* note 182, at 73-75 (suggesting that a more effective design would be to emphasize the injustice of late pickups by “forc[ing] one of the (poorer) caregivers to pay money to the offending parent”).

under these conditions, perhaps because they were able to feel less guilt about small contributions.<sup>239</sup>

### *B. Overjustification and Impaired Self-Determination*

Besides signaling effects, a second broad explanation for crowding-out is “overjustification” that culminates in a loss of self-determination. This theory initially stipulates that individuals benefit from behaving in accordance with their intrinsic motivations, and one benefit is the perception of control over their behavior.<sup>240</sup> When a principal offers an incentive, agents perceive their intrinsic motivations to be unnecessary—that is, the task is now “overjustified” by the combination of intrinsic and extrinsic motivations.<sup>241</sup> The addition of an incentive also displaces a measure of control from agent to principal, because the incentive necessarily exerts some influence on extrinsic motivation.<sup>242</sup> Although the agent could reject the incentive by declining the task, this choice would be “reactive,”<sup>243</sup> displaying his response to the incentive rather than his intrinsic preferences regarding the task. Without the benefits of perceived self-determination, the agent is less intrinsically motivated to engage in the task at all.<sup>244</sup> This theory underlies the description of incentives as “controlling,” leading some to suggest that crowding-out is most likely when agents perceive incentives to impair autonomy in some way.<sup>245</sup>

Some scholars have criticized the self-determination explanation of motivational crowding-out, citing research to suggest that “people believe they have more control over their behavior when offered reward.”<sup>246</sup> Indeed, under this theory, the use of an incentive actually telegraphs the principal’s “lack of control over the [agent]; the person, group, or institution offering the reward believes that favorable consequences are needed to obtain the cooperation of the person asked to perform the task.”<sup>247</sup> Thus, people actually “feel freer to decline

239. See Fuster & Meier, *supra* note 207, at 23-25.

240. Bowles, *supra* note 41, at 1607 (citing E.L. DECI, *INTRINSIC MOTIVATION* (1975)).

241. *Id.* at 1607; see also Benkler, *supra* note 2, at 324; Frey, *supra* note 68, at 1526 (citing research suggesting that overjustification may diminish both intrinsic motivation and work quality); Frey & Jegen, *supra* note 2, at 594.

242. Frey & Jegen, *supra* note 2, at 594; see also Bregm, *supra* note 102, at 24 (“When individuals perceive an intervention as reducing their self-determination, intrinsic motivation is substituted by extrinsic control.”); Bruno, *supra* note 151, at 5; Deci et al., *supra* note 62, at 628.

243. Bénabou & Tirole, *supra* note 141, at 504.

244. See Galle, *supra* note 214, at 881 (describing several theories of crowding-out, including “that monetary incentives are particularly apt to generate resistance because they reduce our sense of autonomy”).

245. See, e.g., Frey & Jegen, *supra* note 2, at 594; see also Beretti et al., *supra* note 37, at 65 (“The crowding-out effect is more likely to occur when (i) the external intervention is perceived as controlling rather than supportive, (ii) there is a high level of self-determination of individuals and (iii) there is a high level of trust between people.”).

246. Eisenberger et al., *supra* note 54, at 687 (criticizing the self-determination explanation found in cognitive evaluation theory).

247. *Id.*

a task when they are offered reward as opposed to simply being asked to perform the task.”<sup>248</sup>

### C. Endogenous Preference Adaptation

A more recent contribution to the motivational crowding-out literature is the idea that incentives lead to lasting changes in actual preferences, which are learned and reinforced by experience.<sup>249</sup> This is referred to as “the endogenous formation of preferences:” although standard economic models assume that individual preferences are independent of (or exogenous to) public policy, research has increasingly focused on the adaptability of preferences in response to the policy environment.<sup>250</sup> Within a population, preferences may change over time due to “cultural transmission . . . across generations” or across society as “individuals imitate other more ‘successful’” people.<sup>251</sup> As preferences change, so too might the impact of policy choices, including future incentive programs.<sup>252</sup> This mechanism differs from other crowding-out explanations because it proposes enduring changes in individual preferences,<sup>253</sup> rather than task-specific changes in the balance of motivations.<sup>254</sup>

248. *Id.*

249. See, e.g., BOWLES, *supra* note 41 (suggesting that market institutions may crowd out virtues through both framing effects and changes in preferences); *id.* at 1607 (explaining the “endogenous preferences” model of crowding-out and noting that preference changes are “a slow process more akin to acquiring an accent than to choosing an action in a game,” given “population-level effects such as conformism, schooling, religious instruction, and other forms of socialization that are not readily captured in experiments”); see also Oren Bar-Gill & Chaim Fershtman, *Public Policy with Endogenous Preferences*, 7 J. PUB. ECON. THEORY 841 (2005) (modeling how a policy that incentivizes contributions to a public good can actually diminish the proportion of agents who value social rewards).

250. See Bar-Gill & Fershtman, *supra* note 249, at 842; Vermeule, *supra* note 43, at 425 (noting that crowding-out effects propose “that the motivations of officials are endogenous”). For additional legal scholarship on endogenous preferences, see Kenneth G. Dau-Schmidt, *Legal Prohibitions as More than Prices: The Economic Analysis of Preference Shaping Policies in the Law*, in LAW AND ECONOMICS: NEW AND CRITICAL PERSPECTIVES 153 (Robin Paul Malloy & Christopher K. Braun eds., 1995); Samuel Bowles, *Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions*, 36 J. ECON. LIT. 75 (1998); Gregory Scott Crespi, *The Endogeneity Problem in Cost-Benefit Analysis*, 8 GEO. J.L. & PUB. POL’Y 91 (2010); 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, *Endogenous Preferences*, *Environmental Law*, 22 J. LEGAL STUD. 217 (1993); and Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 U. CHI. L. REV. 1129 (1986).

251. Bar-Gill & Fershtman, *supra* note 249, at 842-43; see also Bowles & Polania-Reyes, *supra* note 2, at 383 (describing a model of cultural evolution in which the distribution of preferences depends on the structure of incentives).

252. Bar-Gill & Fershtman, *supra* note 249, at 843.

253. See *id.* at 844.

254. The endogenous preference adaptation model may be more successful for explaining how behaviors evolve across an entire population than for explaining how a given individual’s intrinsic preferences change over time. There are, however, intriguing possibilities for studying these changes. For example, a recent paper reports that corporate CEOs who receive large incentive payouts “may become overconfident,” leading them to take more risks over time and engage in “value-destroying” activities that ultimately harm the firm. Michael J. Cooper et al., *Performance for Pay? The Relation between CEO Incentive Compensation and Future Stock Price Performance* 2 (Oct. 1 2014) (unpublished manuscript), [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1572085](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1572085). But as Bar-Gill and

Frey and colleagues provide an interesting illustration for this dynamic in their follow-up study on the siting of the Swiss waste treatment plant. Initially, local residents were less willing to accept the plant when offered compensation.<sup>255</sup> But one year later, a developer re-extended the offer of compensation, and 60% voted to accept. Frey and colleagues argue that some residents believed that the plant would be built eventually; the initial compensation offer inspired these residents to anticipate increased income and to perceive opportunity costs associated with turning down the offer. But to avoid visibly hypocritical behavior, the residents began to develop “new moral arguments” in favor of the facility.<sup>256</sup> When the developer repeated its offer, individuals with changed preferences carried the vote. Frey reports an analogous process in a New York town preceding construction of a solid-waste landfill.<sup>257</sup>

Barry Schwartz and Kenneth Sharpe provide an analogy from the U.S. medical profession. In the early twentieth century, professional ethics guidance and societies like the American Medical Association (AMA) imposed strict rules separating clinical practice from financial incentives; these included, for example, a ban on physician advertising and financial arrangements with drug and device manufacturers.<sup>258</sup> Professional norms shifted, however, with the development and widespread use of cost-control efforts over the past half-century, such as HMO capitation systems, prepaid treatment plans, using physicians as gatekeepers for more expensive tests and services, and risk-sharing agreements.<sup>259</sup> Schwartz and Sharpe argue that these arrangements exerted a powerful framing effect, “encouraging money-making in doctors’ choices about patient care.”<sup>260</sup> Over time, the AMA guidelines shifted to permit advertising and entrepreneurial activities, and the organization now views medicine as “both a profession and a business.”<sup>261</sup> The organization’s preferences have changed, responding in part to the influence of incentives over several decades.<sup>262</sup> Schwartz and Sharpe identify similar pressures operating in the legal profession,<sup>263</sup> and Nina Walton has suggested similar processes at work in the financial sector.<sup>264</sup>

---

Fershtman have recognized, “little has been said regarding the precise mechanism through which public policy may affect norms and preferences.” Bar-Gill & Fershtman, *supra* note 249, at 853.

255. Bruno S. Frey et al., *The Old Lady Visits Your Backyard: A Tale of Morals and Markets*, 104 J. POL. ECON. 1297, 1306 (1996).

256. *Id.* at 1308-10.

257. *Id.* at 1310; *see also* SANDEL, *supra* note 72, at 116-17 (noting that communities often find public goods, such as libraries or parks, to be more palatable compensation for accepting undesirable public projects, and that they do not crowd out norms because they “repay civic sacrifice in the same coin”).

258. *See* SCHWARTZ & SHARPE, *supra* note 200, at 201-03.

259. *Id.* at 204-07.

260. *Id.* at 205.

261. *Id.* at 206 (emphasis omitted).

262. *Id.* at 206-08.

263. *Id.* at 216.

264. Walton, *supra* note 28, at 455.

Several modeling studies, buttressed by game experiments, provide additional support for endogenous preference adaptation. Oren Bar-Gill and Chaim Fershtman have developed a model of contributions to a public good, identifying the impact of a subsidy on increased contributions.<sup>265</sup> Before the subsidy is introduced, individuals who contribute to public goods for reputational motivations only contribute when there is a social image payoff. When they are subsidized, these individuals will contribute more frequently, including in some scenarios without a reputational reward, because the subsidy reduces the costs of contribution.<sup>266</sup> In the long run, selfish individuals will “take advantage of [contributors’] generosity and proliferate [at their] expense.”<sup>267</sup> This effect will eventually lead to a “decline in the share of socially minded individuals” in the population, undermining the incentive policy and resulting in a stable, lower level of contributions.<sup>268</sup>

Iris Bohnet and colleagues have modeled the impact of incentives in the form of penalties for breach of contract, suggesting that preferences evolve within a population and are disciplined by economic success.<sup>269</sup> This model assessed the effect of weak versus strong contract enforcement on “trust” and “trustworthiness,” measured respectively by first movers’ likelihood of entering a contract and second movers’ likelihood of performance. When the probability of punishment for breach is high, second movers can maximize their gain through performance; this environment may crowd out first movers’ preferences for contracting only with honest parties because “interpersonal trust is replaced by institutional trust in the legal system,” making it inefficient to exclude potentially selfish second movers.<sup>270</sup> When the probability of punishment is “medium,” second movers can maximize their gain by breaching, causing selfishness to spread as a “successful ‘trait’” and displacing moral preferences for performance.<sup>271</sup> But when the probability of punishment for breach is *low*, these conditions crowd *in* both second-movers’ preferences for honesty and first movers’ preferences for dealing only with honest parties.<sup>272</sup> A recent modeling study by Scott Masten and Jens Prüfer echoes part of this finding, suggesting that the existence of courts tends to crowd out “informal, reputational enforcement” of cooperation among members of a community, because—in the presence of

---

265. Bar-Gill & Fershtman, *supra* note 249.

266. This proposition is the reverse of most crowding-out theories, which would suggest that the socially-minded individuals would contribute less due to image-spoiling concerns. *See supra* Section II.A.2.

267. Bar-Gill & Fershtman, *supra* note 249, at 852.

268. Bar-Gill and Fershtman further suggest that, given this dynamic, using anti-incentives rather than rewards may produce the opposite effect by amplifying reputational motivations to contribute: “a tax policy may be effective in promoting contributions to a public good . . . independent of any direct spending of the tax revenues on the public good.” *Id.* at 853.

269. Bohnet et al., *supra* note 9.

270. *Id.* at 141.

271. *Id.* at 135.

272. *Id.* at 140-41.

courts—contractors are willing to continue dealing with parties who have defected in prior transactions.<sup>273</sup> Another recent study by Alessandra Cassar and colleagues found that, in the presence of contract enforcement institutions, both trust and trustworthiness cease to predict market participation and cheating behaviors.<sup>274</sup>

In a similar project, Bruno Deffains and Claude Fluet have modeled the extent to which different regimes of legal liability for tortfeasors may crowd out guilt and social image concerns as intrinsic motivations to avoid causing uncompensated harm to others.<sup>275</sup> According to Deffains and Fluet's model, the desires to avoid reputational damage and guilt govern behavior in regimes without legal liability for torts. With perfect enforcement, a *strict liability* regime may crowd out both guilt and reputational motivations—reputational motivations disappear because adverse court decisions do not indicate negligence, while guilt disappears because tortfeasors always pay for damages.<sup>276</sup> A perfectly enforced *negligence* regime may crowd out guilt but not reputational motivation, given that adverse court rulings communicate that a tortfeasor was negligent. Under strict liability with *imperfect* enforcement, Deffains and Fluet argue that there will be “some crowding-out” of both motivations; guilt avoidance still matters because not all tortfeasors are held liable, and reputational motivations still matter because only some tortfeasors are sued. But an imperfectly enforced negligence regime may actively crowd *in* social image concerns: not all tortfeasors are punished (preserving some guilt motivation), and an adverse court decision has the maximum social cost. In this way, legal rules may “have normative power . . . not only by shaping [] material payoffs but also by directly influencing motives.”<sup>277</sup>

#### D. Learned Helplessness

The final explanation of crowding-out may have less modern support, but I have included it here for the sake of completeness. One research team has suggested that crowding-out occurs only when incentives encourage mere participation or low-effort activity, rather than high-quality performance. Based on a systematic review of psychological research on incentives, these authors suggest that crowding-out occurs when agents believe that their level of effort is

---

273. See Masten & Prüfer, *supra* note 106, at 362.

274. See Cassar et al., *supra* note 106, at 857.

275. See Bruno Deffains & Claude Fluet, *Legal Liability When Individuals Have Moral Concerns*, 29 J.L. ECON. & ORG. 930 (2013) [hereinafter Deffains & Fluet, *Legal Liability*]. For further analysis of social image concerns in the design of tort liability regimes, see Bruno Deffains & Claude Fluet, *The Role of Social Image Concerns in the Design of Legal Regimes* (CIRPÉE Working Paper, No. 13-21, Aug. 2013) [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2317319](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2317319).

276. Deffains & Fluet, *Legal Liability*, *supra* note 275.

277. *Id.* at 20.



divorced from rewards.<sup>278</sup> To support this view, the authors note studies showing that when a principal has rewarded an agent regardless of performance quality, the agent subsequently spends less free time on the task.<sup>279</sup> The disassociation between reward and performance may cause “generalized motivational deficits” and a loss of intrinsic interest,<sup>280</sup> leading to less creativity and effort. These scholars moreover suggest that rewards offered *only* for high-effort performance can lead to “learned industriousness,” whereby exertion takes on “secondary reward properties” that increase agents’ effort.<sup>281</sup> This body of research, however, is mixed, as some studies indicate that extrinsic rewards may enhance creativity, while others suggest that they may undermine creative performance.<sup>282</sup>

As an explanation for motivational crowding-out, learned helplessness theory has been criticized on the ground that little evidence supports the link between “uncontrollable positive outcomes” and the feeling of helplessness, particularly when rewards are somehow linked to activity such as engagement with a task.<sup>283</sup> Further evidence suggests that incentives can work to encourage creative activity, especially when agents know they will be rewarded based on their creativity.<sup>284</sup> I will rely on the learned helplessness explanation less in the remainder of this Article, but I note it here due to the several studies with supportive findings.<sup>285</sup>

### III. Designing Incentive Architecture to Accommodate Motivational Crowding-Out

Motivational crowding-out provokes concern for several reasons. Most commentators raise instrumental questions about crowding-out, which focus on how crowding-out affects behavior. If crowding-out overwhelms an incentive’s disciplining effect, incentives may lead to temporary or permanent disengagement—that is, incentives will backfire entirely, resulting in worse behavior than before. Even if incentives are effective, a loss of intrinsic motivation could mean that behavior becomes dependent on incentives—if incentives are later withdrawn, behavior could drop below pre-incentive levels. Others have worried that incentives may reduce performance quality if they distract agents, cause agents to “choke,” encourage gaming, or induce agents to

---

278. Eisenberger & Cameron, *supra* note 98, at 1156; *see also* Deci et al., *supra* note 62, at 630.

279. Eisenberger & Cameron, *supra* note 98, at 1162.

280. *Id.* at 1156; *see also* Deci et al., *supra* note 62, at 630.

281. Eisenberger & Cameron, *supra* note 98, at 1161; *see also* Deci et al., *supra* note 62, at 630.

282. For a helpful overview, *see* Buccafusco et al., *supra* note 36, at 1935-43 (finding a “murky picture of the relationship between incentives and creativity”).

283. *See* Deci et al., *supra* note 62, at 654.

284. *See* Buccafusco et al., *supra* note 36, at 1938-39.

285. *See* Eisenberger & Cameron, *supra* note 98, at 1160-62 (citing studies).

focus on immediate performance instead of long-term task mastery. Further concerns include whether incentives will attract different populations of agents, or whether the displacement of motivation in one area may also reduce it in other areas (for example, when fines for nonvoting displace civic duty, people will disengage from jury duty as well).<sup>286</sup>

Crowding-out has also inspired anxiety on the basis of autonomy and morality. Some worry that incentives (particularly large rewards) limit autonomy by interfering with motivations, sometimes expressed as the view that incentives provide “undue inducement” to perform. This is particularly problematic when incentives encourage potentially risky tasks, such as participation in human subjects research. Others have expressed concern about the inherent moral or attitudinal harms caused by the erosion of intrinsic motivations, such as civic virtue or altruism.<sup>287</sup>

This Article is focused on ways to alleviate the instrumental impacts of crowding-out. When considering the design of incentive policies, it is important to remember that incentives will *not* produce crowding-out when the prerequisites described in Part I are not present: agents must have intrinsic motivation for the task,<sup>288</sup> the incentive must be observable or expected when the agent decides whether to perform the task, and the task must be within the agent’s control and ability. Crowding-out may be less worrisome, but not absent, for tasks that are uninteresting, tasks that are costly (which agents may be unwilling to undertake without compensation),<sup>289</sup> tasks with a tradition of monetary compensation,<sup>290</sup> and tasks for which there is “little or no normative stance on the behavior” at the outset.<sup>291</sup> In contrast, crowding-out may be *more* worrisome when intrinsic motivations matter greatly, such as when behavior is difficult to monitor,<sup>292</sup> pricing is difficult, the funds for incentives are low,<sup>293</sup> or the behavior

286. For a detailed discussion of these impacts, see Underhill, *supra* note 94.

287. See, e.g., SANDEL, *supra* note 72, 93-130; SCHWARTZ & SHARPE, *supra* note 200, at 129-34; Atiq, *supra* note 4, at 1073; Bowles, *supra* note 41. A common argument in this literature is that efforts to “fine-tune” reward systems or make incentives “smarter” are counterproductive. See e.g., Kohn, *supra* note 73, at 4-5. I do not seek to answer these concerns at present; as Sandel has suggested, the moral effects of market incentives require a more participatory and reflective conversation, with the goal of identifying permissible and impermissible uses of incentive mechanisms. SANDEL, *supra* note 72, at 14-15. This is not the task of this Article, which takes a more instrumental view of the costs, benefits, and tailoring of incentives. But to the extent that incentives do produce long-term erosion of valuable motivations, including moral preferences, I agree that this is an important area for future work, and policy tweaks may not always be an appropriate solution.

288. See, e.g., MEIER, *supra* note 191, at 37; Bowles & Hwang, *supra* note 42, at 17 (investigating the extent of optimal incentives that can contribute to a public good when intrinsic motivation is crowded out); Stern, *supra* note 46, at 564-65 (noting when monetary rewards can crowd out intrinsic motivation).

289. Stern, *supra* note 46, at 565.

290. Fehr & Falk, *supra* note 154, at 718.

291. Fleisig-Greene, *supra* note 132, at 1242 (referring to these as “low-commitment laws”).

292. Feldman, *supra* note 2, at 50-51.

293. BENKLER, *supra* note 50, at 95.

has a social valence. Unexpected rewards generally do not affect motivation,<sup>294</sup> but principals must be wary lest “unexpected” rewards come to be expected.<sup>295</sup>

With the overarching goal of addressing motivational crowding-out, my project in this last Part is to introduce incentive architecture: the deliberate design and implementation of incentive programs to maximize their effectiveness. Because crowding-out can undermine the goals of incentive plans, one objective of incentive architecture is to accommodate—and, where possible, minimize—crowding-out effects. This Part discusses a series of decisions that governmental and private principals must make when designing incentive programs, focusing specifically on decisions that may mediate crowding-out effects. These decisions fall into seven categories: program development and implementation, framing choices, incentive definition, contingency decisions, information provision, publicity choices, and options for managing heterogeneity.

For each category of decisions, I identify strategies for intervening in the different crowding-out mechanisms identified in Part I. These strategies are drawn from the crowding-out literature. Because the causes of crowding-out are heterogeneous, principals who are concerned about crowding-out effects must choose solutions that match the specific processes at work.<sup>296</sup> I conclude by suggesting that where program modification is impossible or unsuccessful, the type or extent of harms linked to crowding-out may weigh in favor of discontinuing incentive programs entirely.

### *A. Program Development and Implementation*

Before setting an initial incentive scheme, principals should develop empirical evidence to identify the potential for crowding-out effects that considers knowledge regarding existing motivations, behavioral contexts,<sup>297</sup> and the distribution of agents with different motivations.<sup>298</sup> It is important to acknowledge, however, that some evidence may only be gathered after an incentive scheme is in place, given that agents may have difficulty accurately predicting their own responses to an incentive.<sup>299</sup>

---

294. See Deci et al., *supra* note 62, at 639-40 (finding no effect of unexpected and task-noncontingent rewards on intrinsic motivation, and casting doubt on the “learned helplessness” explanation); Tang & Hall, *supra* note 98, at 379.

295. Deci et al., *supra* note 62, at 656 (“[O]ne must be mindful that people may begin to expect the ‘unexpected’ rewards if they are given very often.”).

296. Although other legal scholars have suggested policy tweaks to reduce motivational crowding-out in specific areas, this is the first Article to argue across subject areas that these solutions must match the behavioral dynamics that drive crowding-out effects.

297. See Feldman & Lobel, *supra* note 1, at 1206-07; see also Stern, *supra* note 46, at 580 (arguing that agents should be involved “early in the process” to enable the principal to gauge intrinsic motivation and optimize program design).

298. Different types of agents may be affected differently by incentives. Feldman & Perez, *supra* note 27, at 436-37.

299. Gary Charness & Uri Gneezy, *Incentives to Exercise* (UC Santa Barbara Economics Dept., Working Paper, Aug. 22, 2008), <http://escholarship.org/uc/item/3tc3j5x7> (finding that

Preparatory research for the design of the incentive scheme should also assess agent competence. To avoid “demoralizing” agents, incentive schemes should aim for a match between agent competence and the incentivized goal: goals should be “attainable by the majority of participants,” with informational feedback on good performance as it is achieved and technical support for agents at risk of failing.<sup>300</sup> Empirical studies have found that “the failure to get the maximum reward . . . could be highly detrimental because the reward structure is not only controlling but also conveys negative-competence information.”<sup>301</sup> Accurately assessing agents’ abilities in advance can reduce the potential for crowding-out due to signals regarding agent competence.

A powerful tool for addressing crowding-out due to impaired self-determination, as well as for addressing the problem of autonomy, is ensuring that the incentive structure is designed using a participatory process. Mechanisms for participatory design might include surveying agents during early stages of incentive development, conducting a well-publicized “notice-and-comment”-type process to solicit agent feedback, allowing agents to vote or otherwise participate in the decision to implement the incentive scheme, and choosing administrators from among the agent population to implement the incentive program. When agents perceive that they have meaningful input on the design of an incentive plan, it may increase the likelihood that they will view the incentive to be supportive, rather than controlling.<sup>302</sup> For example, P4P programs in healthcare have been more successful in increasing provider engagement when providers were themselves involved in selecting and defining performance targets, and when communication with providers was “direct and extensive” regarding measurements and rewards.<sup>303</sup>

The use of a participatory process for incentive development alleviates autonomy concerns by giving agents a voice in incentive development, placing agents on both sides of the principal-agent line. Intrinsic motivation may be demonstrably enhanced when agents participate in the decision-making process,<sup>304</sup> and agents who perceive incentives as supportive are likely to experience crowding-in of intrinsic motivations rather than crowding-out.<sup>305</sup> Frey has also noted that when agents have the ability to participate in institutional processes, they are more likely to view incentives as supportive of their self-esteem, alleviating the potential for crowding-out due to the principal’s signaling

---

individuals’ predictions about whether a financial incentive will motivate their behavior did not predict actual behavioral impacts).

300. Stern, *supra* note 46, at 579.

301. Deci et al., *supra* note 62, at 657.

302. Frey, *supra* note 43.

303. Eijkenaar et al., *supra* note 25, at 126; Magrath & Nichter, *supra* note 170, at 1783.

304. Bruno S. Frey & Alois Stutzer, *Environmental Morale and Motivation*, in THE CAMBRIDGE HANDBOOK OF PSYCHOLOGY AND ECONOMIC BEHAVIOUR 412-14 (Alan Lewis, ed., 2008).

305. *Id.* Crowding-in, as I have noted, does not solve the autonomy problem because it is nonetheless an interference with individuals’ motivations, but the behavioral impacts of crowding-in may be less damaging than those of crowding-out.

of agent value.<sup>306</sup> Agents who meaningfully participate in program design may also be less likely to view incentives as a signal that the principal is hostile, selfish, or in violation of an implied reciprocity agreement. Obtaining agents' views may even help foster intrinsic motivation, which is enhanced when principals and agents have close or communicative relationships that generate reciprocity.<sup>307</sup> Interestingly, using a participatory approach for the development of an incentive program may be insufficient to prevent the extreme autonomy problem of undue inducement. If the program relies on positive incentives that are indeed "too good" to refuse, agents may very well advocate implementing the scheme in order to obtain the incentives. Undue inducement may therefore need to be policed by third parties and remedied by removing incentives or reducing their size.

The design of monitoring and surveillance options for the administration of an incentive scheme should accommodate concerns about intrusiveness. Soliciting agent feedback in advance on plans for monitoring and enforcement may help to mitigate this problem. Surveillance by self- or peer-reported behavior may be preferable to direct observation or evaluation by the principal to avoid crowding-out due to impaired self-determination or signaling the principal's hostility. Where an incentive is positive, some surveillance mechanism is needed for administration of the rewards. But where surveillance is impossible or perceived as intrusive, negative incentives may be preferable, and some evidence suggests that mandates with weak or absent penalties may be effective. Successful examples of such mandates include no-smoking signs, tax compliance, pooper-scooper laws,<sup>308</sup> seat belt laws that are rarely enforced,<sup>309</sup> and parking laws for diplomats exempt from fines.<sup>310</sup>

One of the major harms arising from motivational crowding-out is sustainability and reduced long-term performance if incentives are withdrawn. At the planning stage, principals should account for this by ensuring that sufficient resources are devoted to support long-term and perhaps even perpetual implementation of the incentive program, including the costs of incentives, administration, and surveillance.<sup>311</sup> Allocating resources to identify and counteract "gaming" behavior may also be necessary if the principal is concerned about crowding-out harms due to reduced quality of performance; if gaming

---

306. See Frey, *supra* note 43 (suggesting that opportunities for institutional participation increase the likelihood that an incentive will support an individual's self-esteem).

307. Frey, *supra* note 304, at 412-14.

308. Cooter, *Three Effects*, *supra* note 184, at 4.

309. See, e.g., Maggie Wittlin, *Buckling Under Pressure: An Empirical Test of the Expressive Effects of Law*, 28 YALE J. ON REG. 419, 429-31 (2011); Fleisig-Greene, *supra* note 132, at 1241.

310. See Funk, *supra* note 185, at 137 (describing these examples as indicative of an "expressive" function of law).

311. Extending an incentive for long-term use may also deserve future study; as Emir Kamenica has noted, there is as yet "no systematic evidence . . . that indicates long-term monetary incentives can backfire." Kamenica, *supra* note 142, at 13.5.

occurs, efforts to address it may be in tension with concerns about program intrusiveness. Once an incentive plan is in place, continued program evaluation and adjustment can identify crowding-out effects and opportunities for program modification as they arise. Trial periods of incentives may sometimes help to facilitate their acceptance; as endogenous preference adaptation takes place, incentives may grow to be more acceptable over time.<sup>312</sup>

The implementation of incentive programs can also embed respect for agents' autonomy, reducing the likelihood of crowding-out due to impaired self-determination. As Stephanie Stern has noted in her discussion of conservation programs, staff with a "personalized tenor" may improve the effectiveness of incentive plans by emphasizing the ways in which agents are treated as individuals.<sup>313</sup> Empirical evidence from several experimental studies suggests that rewards are more likely to crowd out intrinsic motivation when the interpersonal climate (the context of the reward and the style of the principal) is controlling and pressuring; when the context is relatively non-controlling, rewards are more likely to be experienced as informational, once again alleviating crowding-out due to impaired self-determination.<sup>314</sup> The same is true of verbal rewards or praise; research indicates that when praise is expressed using controlling or pressuring language, it leads to a decrease in intrinsic motivation, while praise in purely informational language does not.<sup>315</sup>

### *B. Framing Decisions*

In the broadest terms, "framing decisions" refer to an incentive architect's option to select a positive incentive, negative incentive, or mandate or prohibition without an explicit enforcement mechanism. Several legal scholars have already noted pressures in favor of using positive incentives rather than negative incentives or mandates. For example, Galle has noted that political processes are set up to favor the creation of carrots rather than sticks, particularly in tax policy; sticks operate by transferring resources from a relatively small (offending) group to the larger society, which mobilizes a concentrated interest group to oppose the policy.<sup>316</sup> Carrots, on the other hand, represent transfers from the larger society to a smaller group, which may draw less organized opposition.<sup>317</sup> Rewards may also be more desirable than penalties for disciplining the behavior of individuals or households; "sticks may be undesirable when they fall on households that are

---

312. See Frey et al., *supra* note 255.

313. Stern, *supra* note 46, at 580.

314. Deci et al., *supra* note 62, at 629, 653. See also Stern, *supra* note 46, at 566 (noting research suggesting that "threats, high-pressure evaluation, intensive surveillance, and deadlines can reduce intrinsic motivation"). But see Tang & Hall, *supra* note 98, at 379 (summarizing several studies finding no effect of feedback "designed to provide negative or controlling information" on intrinsic interest).

315. Deci et al., *supra* note 62, at 629-30.

316. Galle, *supra* note 63, at 802-03, 840-41.

317. See *id.*

poorer than average,” or they may fail to deter judgment-proof households.<sup>318</sup> Government-imposed rewards may also receive more favorable treatment than fines during judicial review, and federal standing rules that give individual beneficiaries a right to sue may encourage coalitions to seek incentive programs more frequently than other types of regulation.<sup>319</sup> Moreover, incentives may sometimes be more desirable, more effective, and fairer than other forms of regulation when society is highly-specialized or tasks are complex.<sup>320</sup> When a task demands far more effort from some agents compared to others, using incentives instead of penalties can ensure a more equitable balance of burdens and benefits.<sup>321</sup> And because penalties can impoverish agents, carrots are “less prone to abuse than sticks.”<sup>322</sup> Sticks, on the other hand, are typically better than incentives for managing externalities.<sup>323</sup>

Beyond these concerns, the decision of how to frame an incentive program has ramifications for crowding-out effects. Where there is a concern about motivational crowding-out due to impaired self-determination, it may be advisable to use positive rather than negative incentives, as penalties may give a stronger impression of coercion.<sup>324</sup> Incentives framed as bonuses may also be less likely to trigger crowding-out by signaling a hostile or non-reciprocal principal; fines may convey a stronger message of distrust or hostility, while bonuses may be more likely to convey a message of kindness.<sup>325</sup>

Other mechanisms of crowding-out, however, surprisingly counsel in favor of using negative incentives. Where crowding-out occurs because a positive incentive signals unsupportive social norms, the use of penalties may actually increase confidence and engagement by agents who fear being taken advantage of by free riders.<sup>326</sup> When crowding-out occurs because a reward deprives individuals of the opportunity to signal their good character to others or themselves, using negative incentives may induce less crowding-out than positive rewards. Feldman notes, for example, that penalties may be more effective than rewards for encouraging socially controversial behaviors such as whistleblowing, where acting to obtain incentives may be perceived as opportunistic.<sup>327</sup> Negative incentives may also be preferable when positive incentives lead to framing or priming effects, or where offering positive incentives induces moral repugnance. Where there are concerns about incentive sustainability and long-term behavioral impacts, using negative incentives may

---

318. *Id.* at 818.

319. *See id.* at 842-44.

320. *See* De Geest & Dari-Mattiacci, *supra* note 17.

321. *See id.*

322. *Id.* at 393.

323. Galle, *supra* note 63.

324. Kohn, *supra* note 73, at 5.

325. *See* Fehr & Gächter, *supra* note 163, at 31.

326. *See* Kahan, *supra* note 19, at 346.

327. *See* Feldman, *supra* note 2, at 39; *see also* Feldman & Lobel, *supra* note 1.

also be advantageous; fines and penalties may cost less than financial rewards (and may even represent a net gain for the principal, depending on the transaction costs involved in administering the incentive program).

The use of mandates, prohibitions, and negative incentives may also be useful to avoid crowding-out due to image-spoiling. Where prohibitions are aligned with social norms, either from the outset or due to the expressive effect of law, they harness rather than counteract reputational motivations. “By explicitly proclaiming a behavior as antisocial, a prohibition may be complementary with individual values, affirming a citizen’s moral predisposition to not behave antisocially rather than crowding out moral sentiments.”<sup>328</sup> The use of mandates or fines may be particularly effective when accompanied by public education and outreach framing the mandate within the context of social norms.<sup>329</sup> One example may be Ireland’s highly-visible campaign to stigmatize the use of plastic shopping bags as wasteful and harmful to the country’s natural beauty, which increased the effectiveness of its newly-imposed tax on plastic bags.<sup>330</sup> Even unenforced mandates or “aspirational norms” may be considered a form of “soft law,” encouraging but not requiring agents to act as desired.<sup>331</sup> Experimental research by Brian Sheppard and Fiery Cushman found that using aspirational norms encouraged more charitable giving than mandatory norms, suggesting that such norms were less likely to crowd out charitable motivations.<sup>332</sup>

### C. Incentive Definitions

Incentive architects must make a range of decisions about the type, size, and choice set of incentives. This Section will refer primarily to positive incentives, but penalties can also differ along these lines. The choice of an incentive can drive crowding-out effects in a number of ways. Here I will consider five attributes of incentives: their type, size, specificity, fungibility, and recipients.

#### 1. Incentive Type

First, the type of incentive—cash, an in-kind incentive, verbal praise, an award, or another prize—can influence whether the program activates several crowding-out mechanisms. If crowding-out occurs because agents view the

---

328. Bowles & Hwang, *supra* note 42, at 18.

329. See, e.g., Bowles & Polania-Reyes, *supra* note 2, at 417-18 (suggesting that combining negative incentives with information supporting social norms can be an effective deterrent to antisocial behavior, and noting that fines can “work more as messages than as incentives”).

330. See Hwang & Bowles, *supra* note 223, at 21.

331. See Brian Sheppard, *Norm Supercompliance and the Status of Soft Law*, 62 BUFF. L. REV. 787, 802 (2014).

332. Sheppard & Cushman, *supra* note 106.



incentive as a signal of their own competence, it may be desirable to use in-kind incentives (for example, T-shirts, pens, bags); these may limit agents' impulse to gauge performance by the size of the reward, attenuating the informational signal of the reward.<sup>333</sup> The use of an in-kind incentive may also reduce crowding-out due to framing and priming effects (avoiding "a detrimental 'market instinct'"),<sup>334</sup> the eviction effect, moral repugnance triggered by using cash, image-spoiling, and self-image-spoiling. When incentives are provided in-kind, however, it is important to note that communicating the cash value of the gift to the agent may produce the same effect as cash alone.<sup>335</sup> Using verbal praise (or shame) instead of tangible incentives may also avert image-spoiling for tasks associated with reputational motivation.<sup>336</sup> Another option may be to give agents a choice of incentives. Bénabou and Tirole suggest that offering agents a "menu" of incentives may diminish image-spoiling effects,<sup>337</sup> suggesting that agents who wish to preserve their reputational benefits may choose small rewards or forego them entirely. This opportunity for agent choice may also reduce crowding-out due to impaired self-determination and diminish autonomy concerns. The use of in-kind incentives may be particularly useful for counteracting image-spoiling and self-image-spoiling when the goods can be easily obtained elsewhere. That is, an agent is unlikely to have volunteered solely in order to obtain a mug, for example, if he can get a mug somewhere else.<sup>338</sup>

Even within the category of in-kind incentives, there is some evidence that crowding-out due to framing and priming may be lessened further when the incentive is "congruent" with the incentivized behavior. For example, research with blood donors suggests that incentives are perceived to be either congruent or incongruent with the blood donation process;<sup>339</sup> congruent incentives were those that made the blood donation process itself easier, more comfortable, or more frequent (e.g., blood screening for health conditions, paid time off from work, recognition letters, refreshments, and branded "reminders" to donate

---

333. See Frey, *supra* note 44, at 433.

334. Beretti et al., *supra* note 37, at 75; see also Stout, *supra* note 7, at 557-58 (suggesting that nonfinancial rewards may be superior ways of rewarding employees compared to pay-for-performance).

335. See Heymann & Ariely, *supra* note 197, at 792.

336. This may be more effective "where gaining distinction is the dominant reputational concern (self-sacrifice, heroism, great inventions)," rather than where agents are concerned with simply avoiding stigma. Bénabou & Tirole, *supra* note 47, at 1672.

337. *Id.* at 1669 (noting that menus "may be a good strategy for increasing contributions," and that menus may also benefit principals by yielding information about the agents' own preferences). A danger in offering menus or give-backs, however, is that the good deed of declining an incentive may be "too obvious," revealing to observers that the agents who give back rewards are in fact motivated by image concerns. In this scenario, agents may never turn down the incentive, and the menu or give-back option would not remedy the problem. *Id.*

338. *Id.* at 1663.

339. See Chmielewski-Ralmondo et al., *supra* note 175, at 1896.

again).<sup>340</sup> These congruent incentives were less likely to prompt negative reactions than incentives that were unrelated to the activity (including cash). This outcome echoes the results of a study finding that loyalty programs for coffee shop customers were more acceptable if they rewarded consumers with coffee shop offerings (e.g., baked goods or mugs) than if they provided unrelated rewards such as movie tickets.<sup>341</sup> The congruency perception may also underlie the greater willingness of citizens to receive public goods, such as parks and bike trails, instead of cash as compensation for accepting NIMBY projects; the exchange is perceived as an opportunity to strengthen the community in exchange for a “civic sacrifice.”<sup>342</sup>

## 2. Incentive Size

Second, the size of the incentive matters for crowding-out. Here, different mechanisms for crowding-out counsel different policy tweaks. Mechanisms such as impaired self-determination, image-spoiling, self-image-spoiling, framing and priming effects, moral repugnance, and the eviction effect all operate in favor of *reducing* award size (or eliminating incentives altogether). The same is true of some of the downstream effects of crowding-out, including autonomy concerns, temporary reductions in effort, choking, distraction, performance orientation, gaming, and agent selection effects. Minimizing some of these effects, such as distraction and performance orientation, may be especially useful for encouraging creative behavior.<sup>343</sup>

In contrast, *increasing* incentive size may alleviate other crowding-out effects, particularly if the desired behavior relies on effort or engagement alone (rather than on quality of performance). Across the board, once the incentive is large enough, the relative price effect will dominate the crowding-out effect, remedying concern about reduced effort.<sup>344</sup> For example, a fining system could provide escalating fines that eventually reach a level that ceases to be a “price” and becomes a sufficient prize or penalty to induce performance.<sup>345</sup> If agents temporarily reduce their engagement due to playing hard-to-get (an example of relative reductions in intrinsic motivation), increasing award size will also remove the problem.<sup>346</sup> Where crowding-out is due to signaling the value of the

---

340. See *id.* Other studies have also found that effective incentives for donors may include screening for health conditions such as high cholesterol, giving “blood credits,” and public prizes for reaching donation quotas. See Costa-Font et al., *supra* note 58, at 5-6.

341. Chmielewski-Ralmondo et al., *supra* note 175, at 1896 (citing Kivetz, *supra* note 89).

342. SANDEL, *supra* note 72, at 116-17.

343. See Eisenberger & Cameron, *supra* note 98, at 1162.

344. This is one reason why some scholars have noted that crowding-out appears most likely when incentives are “intermediate.” See, e.g., Feldman, *supra* note 2, at 26.

345. See, e.g., Katrina Fischer Kuh, *When Government Intrudes: Regulating Individual Behaviors That Harm the Environment*, 61 DUKE L.J. 1111, 1179 n.270 (2012).

346. Where strategic crowding-out effects are observed, principals may act most wisely by initially offering a small incentive, then increasing it. But this is in tension with crowding-out due to

task, increasing incentive size may have two opposing effects: larger incentives may communicate greater task importance, but also greater task burden or danger, so effects are unclear. Similar contradictions vex the design of incentives that may signal the agent's competence (larger incentives may communicate more positive information about performance, but also may signal that the principal is pessimistic about the agent's intrinsic motivation), social norms (larger incentives may signal either high social expectations or high social costs of performing), and the principal's value (larger incentives may signal either generosity or a manifestation of control).

Some have argued that, instead of considering absolute incentive size, a more relevant metric is the extent to which the incentive is proportional to the effort or task burden.<sup>347</sup> To minimize crowding-out due to image-spoiling, for example, “the items should . . . be cheap compared to the donation,”<sup>348</sup> taking advantage of (and preserving) social norms that already favor the behavior. This also applies where crowding-out occurs due to self-signaling; to reduce self-image-spoiling effects, incentives should be “proportional to behavioral costs and performance,” and avoid being “disproportionately high.”<sup>349</sup> Writing in the area of environmental conservation, Stern has suggested that one way to achieve an appropriate balance is to use competitive bidding, in which agents “submit bids stating the amount of compensation required to elicit their participation.”<sup>350</sup> But where image-spoiling and self-image-spoiling are concerned, even reducing the award size may sometimes be unhelpful; “even a minimal concern about appearing greedy is sufficient to cause a sharply negative response to small incentives.”<sup>351</sup>

### 3. Incentive Uncertainty and Specificity

Third, the specificity with which the incentive is defined may matter. Some have noted that where rewards are interpreted as signals of agent competence, trustworthiness, or principal reciprocity, it may be preferable to leave incentive agreements incomplete,<sup>352</sup> which can convey faith in the agent. Wendy Epstein has noted the conventional view that detailed contracts—those that specify tasks and compensation in advance—make it easier for principals to control and

---

social norms; raising an incentive may communicate that the initial offer was insufficient to motivate good behavior, and that social norms are therefore highly unsupportive of the task.

347. See, e.g., Stern, *supra* note 46, at 575-77; see also Winick, *supra* note 44, at 792 (warning that crowding-out can result when the principal selects a reward or penalty that is too severe).

348. Bénabou & Tirole, *supra* note 47, at 1663. In effect, this may be viewed as a “tax” on reputation-seeking motives to act, because reputation-seeking is socially wasteful; similarly, when the incentive is a penalty for bad behavior, using smaller penalties can take advantage of social norms such as “opprobrium” already working against the behavior. *Id.* at 1672.

349. Stern, *supra* note 46, at 575-77, 565.

350. *Id.* at 578.

351. Bénabou & Tirole, *supra* note 47, at 1663.

352. See Herold, *supra* note 154; Epstein, *supra* note 112.

monitor agents.<sup>353</sup> But more recent research on reciprocity, including several papers cited above, suggests that less specific contracts can improve agents' perception of autonomy,<sup>354</sup> avoid signaling distrust of the agent, and encourage cooperation by "fram[ing] the relationship from the outset as a cooperative one."<sup>355</sup> Although reducing specificity may pose other problems—such as complicating litigation in the event of an action for breach—principals concerned about preserving long-term reciprocity norms and intrinsic motivation may seek to experiment with less complete arrangements.

Where incentives are negative, such as fines, vagueness about potential penalties may be offensive, but nonetheless effective. Feldman and Teichman have found that agents may perceive uncertain or probabilistic negative incentives—such as a fine or tort judgment with imperfect enforcement—as less acceptable compared to penalties that are certain and known in advance, but they may be a more effective deterrent to antisocial behavior.<sup>356</sup> To isolate this effect, Feldman and Teichman asked students to play the role of a factory owner who could choose to dump chemicals in a nearby lake. The penalty for pollution was randomly specified as a fine paid in advance, a fine paid after the fact, or a fine that *may* be enforced after the fact. Students were most averse to the probabilistic penalty, which most effectively deterred pollution and signaled most strongly that polluting is immoral.<sup>357</sup> As the authors note, one explanation for these results is that wholly certain penalties inspire "cost-benefit analysis" that encourages agents to breach and pay,<sup>358</sup> that is, reliable fines may become prices, but agents are less likely to view a probabilistic fine as a price that licenses breach.<sup>359</sup> (Even if probabilistic fines work, however, fines that arouse such aversion may nonetheless cause *relative* crowding-out—compliance for the sake of the fine rather than intrinsic motivation—and future research might consider this question.)

#### 4. Incentive Fungibility

The fungibility, or tradability, of an incentive refers to the ease with which it might be converted to a substitute or another good. Crowding-out effects due to framing and priming are associated with cash payments, suggesting that incentives that more strongly approximate cash may provoke stronger crowding-

---

353. See Epstein, *supra* note 112, at 38-39.

354. *Id.*, at 40-41 (citing studies).

355. *Id.*, at 41.

356. See Yuval Feldman & Doron Teichman, *Are All "Legal Dollars" Created Equal?*, 102 NW. U. L. REV. 223 (2008); Feldman, *supra* note 2, at 27-28 (describing Feldman & Teichman's paper).

357. Feldman & Teichman, *supra* note 356, at 248.

358. *Id.*, at 253.

359. See *id.*, at 251 (suggesting that a state wishing to deter harmful activity, even when such activity is efficient, should make the penalty "as different from a price as possible: making the payment probabilistic, assessing it after the harmful activity, and directing it to the state").

out effects. Fungibility is an essential difference between cash and in-kind incentives, of course, but in-kind incentives may themselves fall along a spectrum of difficulty in exchange. When incentives are more difficult to replace or trade, agents may be less able to interpret them as signals of competence or task value, and perhaps less likely to conclude that market-based reasoning is the most appropriate mode of decision-making. In one test of this idea, Nicola Lacerta and Mario Macis surveyed blood donors in Italy about their likely acceptance of an incentive, which was randomly specified as cash or as a voucher of equivalent value to be used for books or food.<sup>360</sup> Approximately 21% of women and 11% of men predicted that they would reduce or stop donating if offered cash, but these rates were far lower (both below 5%) when the incentive was a voucher.<sup>361</sup> Another test of this hypothesis would be to compare equivalent incentives with a broader range of fungibility, such as \$20 in cash, a \$20 restaurant gift card (perhaps with an offer to buy it back in cash), a personalized voucher for a meal at the restaurant, or a buffet provided directly after the donation. This type of test would help to clarify how fungibility or easy tradability can exacerbate crowding-out due to framing, priming, or other types of signaling. To cite another example, some have suggested that creating tradable emission rights (which create opportunities for individual gain) may produce a stronger crowding-out effect than imposing emission taxes.<sup>362</sup> The tradability of those rights—the opportunity to convert them into cash or an option to pollute—is integral to their crowding-out effects. To minimize these types of crowding-out effects, it may be advisable to use personalized incentives that offer less opportunity for trade.

### 5. Incentive Recipient

Finally, the incentive architect may also vary the recipient of the incentive; the agent can receive the incentive herself, the incentive may be remitted to a charity or *anti-charity* (for penalties) of the principal or agent's choosing, or the agent may have the freedom to designate the incentive recipient.<sup>363</sup> Several articles have now noted the possibility of remedying crowding-out effects due to image-spoiling, self-image-spoiling, and moral repugnance by allowing the agent to designate her incentive as a charitable contribution. For example, allowing female blood donors to donate their incentives to charity eliminated

---

360. Nicola Lacerta & Mario Macis, *Do All Material Incentives for Pro-Social Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations*, 31 J. ECON. PSYCHOL. 738 (2010).

361. *Id.* at 742.

362. See Feldman, *supra* note 2, at 38 (citing work by Bruno Frey). This example may also reflect the difference between positive and negative incentives, but the opportunity to trade emission rights for other items of value makes them similar to cash payments rather than in-kind incentives.

363. One could argue that an agent who receives an incentive always has the freedom to donate it to charity, but the empirical studies cited here found an effect of embedding the charity option directly into the incentive plan.

crowding-out;<sup>364</sup> in the case of financial incentives for landowners to allow public access to their property, Hynes has interpreted this research to suggest that programs may decrease crowding-out by allowing landowners to designate incentives for charity instead of accepting them.<sup>365</sup> If agents can direct their incentives back toward the principal (for example, if the principal is a charity), this may also diminish motivational crowding-out due to the eviction effect. Beretti and colleagues, however, note that a principal's decision to automatically direct agents' incentives to charity may not always have the desired impact.<sup>366</sup> Rather, the principal should give the agent the opportunity to direct incentives to himself or a charitable cause; in Beretti's research, rates of behavior were highest when agents had this option.<sup>367</sup> Similar to the incentives "menu" discussed above, giving agents the option to identify the recipient of their contributions may also foster autonomy. Additionally, when a principal makes or allows contributions to charity, this may promote agents' happiness and cause them to feel positive about the principal's value and reciprocity;<sup>368</sup> these dynamics may reduce motivational crowding-out due to perceived hostility by the principal. One caveat here may be the need to ensure that the agents support the charity's mission, or at least have the option of transferring the incentive to multiple charities that support their views.<sup>369</sup>

Allowing agents to donate their incentives to a charity may be particularly important when the incentives are small, which may assist in the implementation of a low-budget program. An experimental study has suggested that when incentives are low, individuals may in fact exert more effort when they perceive the beneficiary to be a charity, and that many even choose to work for the charity rather than self-gain under these conditions.<sup>370</sup> (The effect was reversed, however, when incentive stakes were high, in which case individuals exerted more effort on their own behalf.)<sup>371</sup>

When the incentive is a fine, the recipient of the fine may also matter. The study by Feldman and Teichman described above—in which students acted as factory owners—also tested whether the participants would be more likely to pollute a lake if they paid a neighbor harmed by the pollution, compared to if they paid a fine to the state.<sup>372</sup> Students who were asked to compensate the neighbor directly were more likely to pollute, in part because the payments

---

364. See Mellström & Johannesson, *supra* note 58.

365. See Hynes, *supra* note 138, at 982.

366. See Beretti et al., *supra* note 37, at 75.

367. See *id.* at 75.

368. See Imas, *supra* note 233.

369. The results of the study by Arieli et al., see *supra* text accompanying notes 219-220, in part show agents' concern about the recipients of their efforts, at least while performing in public, although this was not the central focus of their work.

370. Imas, *supra* note 233, at 17.

371. *Id.*

372. Feldman & Teichman, *supra* note 356.

displaced their “intrinsic motivation not to harm others.”<sup>373</sup> This study suggests that negative incentives may be less likely to induce crowding-out of prosocial obligations if they are remitted to a third party, such as the state (which could then compensate individuals injured by noncompliance), without explicitly changing the duties owed among citizens themselves.

#### *D. Contingency Decisions*

In the context of incentive architecture, “contingency” refers to the strength of the link between the incentive and performance. Here again, the different crowding-out mechanisms demand different remedial responses.

First, some mechanisms of motivational crowding-out suggest that incentive architects may avoid these effects by making incentives noncontingent on behavior, such as by giving agents “burn money” in advance to encourage them to attempt a task.<sup>374</sup> This may reduce crowding-out due to an unfavorable signal regarding agent competence or perceived principal hostility: providing the incentive in advance and on a noncontingent basis is likely to signal trust in the agent and a generous principal, activating reciprocity motivations for performance. Research has documented that unexpected favors or gifts may “create feelings of indebtedness obliging many people to repay the psychological debt.”<sup>375</sup> Moreover, some research suggests that agents who receive noncontingent incentives exhibit more interest in a task compared to those receiving contingent rewards,<sup>376</sup> and noncontingency may promote both agent creativity and autonomy.<sup>377</sup> Noncontingent incentives can be provided on an individual or group basis; tax reform, for example, may increase favorable views of the law among the general population and therefore motivate higher tax compliance.<sup>378</sup> These noncontingent rewards, however, “do not provide a means of expressing appreciation for doing a task or doing it well because they are not linked to the task . . . [Therefore,] their primary utility may be in situations where rewards are necessary but are not used to convey information or to motivate performance (e.g., paying people a salary for occupying a job).”<sup>379</sup> A noncontingent reward may also have undesirable agent selection effects (if it is not limited to certain agents); the promise of a noncontingent incentive may attract unqualified agents to attempt performance.<sup>380</sup> A similar way of insulating

---

373. *Id.*, at 250-51.

374. See Bénabou & Tirole, *supra* note 141, at 501.

375. Fehr & Gächter, *supra* note 154, at 161.

376. See, e.g., Tang & Hall, *supra* note 98, at 373.

377. See e.g., Bruno S. Frey, *State Support and Creativity in the Arts: Some New Considerations*, 23 J. CULTURAL ECON. 71, 79 (1999).

378. Kahan, *supra* note 19, at 343.

379. Deci et al., *supra* note 62, at 656.

380. Bénabou & Tirole, *supra* note 141, at 501-02 (noting that noncontingent rewards are akin to “burning money,” which can attract lazy or undesirable types if the principal cannot restrict such rewards to high performers).

agents from crowding-out effects at the time of performance is to provide incentives *ex post*,<sup>381</sup> as a subsequent reward for good performance that takes advantage of the extent to which incentives signal agent competence. Unexpected rewards, as described above, do not give rise to crowding-out, and some research suggests that *ex post* or discretionary bonuses may “enhance motivation by functioning as a credible feedback mechanism.”<sup>382</sup>

Second, incentive architects who choose a contingent reward must decide whether to link the incentive to engagement (expending any effort), process indicators (completing interim steps or following a pre-specified process), or performance (achieving a measurable outcome, whether absolute or relative to other agents). Incentive plans seeking to avoid crowding-out due to impaired self-determination, self-image-spoiling, and perceived agent competence should be designed to reward outcomes, rather than process indicators. This approach allows agents to retain independence in deciding how to achieve the goal,<sup>383</sup> allows more latitude for creative tasks,<sup>384</sup> ensures a more informational rather than controlling delivery of the incentive,<sup>385</sup> and may limit autonomy concerns. To avoid the impression that incentives in this context are controlling, it may also be useful to design incentive plans to reward *high-quality* performance rather than mere participation or task completion: although performance-contingent rewards embed more intrusive surveillance (for performance evaluation), they are also more likely to convey positive information about competence for those who meet the standard.<sup>386</sup> (Importantly, performance-contingent rewards may backfire among agents who do not meet the specified standards for the reward,<sup>387</sup> which is why prior assessment of agent competence is important.) Although the learned helplessness theory of motivational crowding-out is not commonly accepted, the remedy for this effect is also to tailor the incentive to high-quality performance (rather than, for example, to engagement).<sup>388</sup> Incentive architects concerned about learned helplessness

---

381. *Id.* at 504; *see also* Stout, *supra* note 7, at 559 (advocating *ex post* rewards rather than incentive-based pay for employees).

382. Bénabou & Tirole, *supra* note 47, at 1655 (citing research by Anton Suvorov and Jeroen van de Ven).

383. *See* Stern, *supra* note 46, at 579.

384. *See, e.g.,* Frey, *supra* note 377, at 79-80 (1999) (stressing the importance of allowing artists control over their product when incentivizing artistic works).

385. *See* Deci et al., *supra* note 62, at 655-57 (citing research by Richard M. Ryan and Miron Zuckerman).

386. *See id.* at 628, 641.

387. *See* Stern, *supra* note 46, at 579 (“Research has found that participants who fail at projects for which a performance-contingent reward was possible suffer large reductions in intrinsic motivation.”).

388. *See* Eisenberger & Cameron, *supra* note 98, at 1164 (discussing strategies for averting learned helplessness effects).



should link rewards to an explicit performance criterion, preferably one that benchmarks performance relative to other agents.<sup>389</sup>

Third, principals must decide precisely what behavior to reward, including options such as a temporary or one-off behavior, a corpus of good work, the maintenance of a behavior or goal (e.g., maintaining weight loss),<sup>390</sup> or an undesirable *anti*-behavior. Most incentive programs reward desirable behaviors or the maintenance thereof, and maintenance incentives can be one way to avoid the sustainability problems caused by motivational crowding-out. But several intriguing programs have taken advantage of the image-spoiling and self-image-spoiling mechanisms for crowding-out effects to impose anti-incentives, which reward *undesirable* behavior. For some years now, the online retailer Zappos has offered new employees substantial sums of money (\$2,000) to quit their jobs.<sup>391</sup> Employees' refusal to accept the offer communicates to themselves and others that they value their job highly, and incurs for them a "sunk cost" that increases their motivation "to make good on that initial choice by working hard to succeed."<sup>392</sup> Amazon has now instituted a similar program for full-time employees, offering employees \$2,000 in their first year, then escalating the offer by \$1,000 each year until it reaches \$5,000 (under the heading "Please Don't Take This Offer").<sup>393</sup> Ian Ayres has catalogued other anti-incentives, such as requiring payment to use a carpool lane,<sup>394</sup> or punishing someone innocent if an agent engages in bad behavior.<sup>395</sup> Bar-Gill and Fershtman have also suggested that, where reputational motivations are at stake, *taxing* good behaviors (such as contributions to public goods) may amplify their reputational payoff and solidify reputational motivations, staving off the endogenous preference adaptations—a shift away from social motivations and toward incentives—that may ensue from an opposite policy of subsidizing good behavior.<sup>396</sup> It would perhaps be unwise, however, to use anti-incentives when reputational and self-image motivations are absent (that is, where there is no social cost for the anti-behavior); Amazon, for

---

389. See Eisenberger et al., *supra* note 54, at 687 (using meta-analysis methods to conclude that "rewards for meeting more restrictive performance standards produce more positive effects on intrinsic motivation than do rewards for meeting vague standards"). Research by this group found that incentives that rewarded people for meeting "vague performance standards" reduced engagement in the activity during "free time," but did not affect self-reported interest in the activity; incentives rewarding people for meeting "absolute performance standards" did not affect free-time activity but increased self-reported interest; and incentives rewarding people for performing well compared to *others* ("normative performance standards") increased both free-time activity and self-reported interest. *Id.*

390. See, e.g., AYRES, *supra* note 182, at 100-21 (discussing maintenance contracts); AYRES, *supra* note 41.

391. See AYRES, *supra* note 182, at 33-36.

392. *Id.* at 34.

393. Emily Cohn, *Why Amazon Pays Some Workers up to \$5000 to Quit*, HUFFINGTON POST, Apr. 10, 2014, [http://www.huffingtonpost.com/2014/04/10/amazon-pay-to-quit\\_n\\_5128035.html?utm\\_hp\\_ref=politics&ir=Politics](http://www.huffingtonpost.com/2014/04/10/amazon-pay-to-quit_n_5128035.html?utm_hp_ref=politics&ir=Politics).

394. See AYRES, *supra* note 182, at 35-36.

395. See *id.* at 75.

396. See Bar-Gill & Fershtman, *supra* note 249, at 853.

example, does not offer its anti-incentive to temporary workers, who may have less personal or social investment in the position.<sup>397</sup>

Fourth, principals must consider the extent to which a contingent incentive is salient to the agent during the time of performance. This is particularly important for alleviating the effects of crowding-out on performance quality through mechanisms such as choking, distraction, or performance orientation. Recent work on motivational crowding-out has drawn a distinction between the operation of directly performance-salient and indirectly performance-salient incentives.<sup>398</sup> When an incentive is directly salient, agents can perceive “a clear, proximal, unambiguous link between the incentive and performance,” such as a sales commission.<sup>399</sup> Incentives of this kind are both visible and immediate to agents, and they are likely to lessen the strength of the link between intrinsic motivation and behavior.<sup>400</sup> An *indirect* performance-salient incentive has a more ambiguous link with performance, such as a salary, and may influence behavior to a lesser degree due to its reduced immediacy and visibility.<sup>401</sup> A recent meta-analysis did indeed find that the link between intrinsic motivation and performance was stronger when incentives were indirectly performance-salient than when incentives were directly performance-salient.<sup>402</sup> Other ways of reducing an incentive’s salience may include reducing the size of the incentive, linking it to a large body of tasks,<sup>403</sup> or making an incentive more like a prize (something that cannot consciously be worked toward).<sup>404</sup> Furthermore, introducing a delay can diminish salience. For example, if the time horizon of incentives is such that they are only realized after a delay of several years, the incentives may exert less influence on day-to-day activity. Concerns about long-term performance may be alleviated by structuring incentives so that they may only be realized in the long-term, such as Professor Romano and colleagues’ proposal on deferred-incentive compensation for bank executives.<sup>405</sup>

An incentive architect’s final contingency decision is to design the metrics by which agents’ performance and qualification for incentives are measured. Here, the concern about reduced quality of performance (through mechanisms

397. See Cohn, *supra* note 393.

398. See e.g., Cerasoli et al., *supra* note 87, at 4.

399. *Id.*

400. See *id.* at 4-5.

401. See *id.* at 4.

402. See *id.* at 15.

403. See Frey, *supra* note 44, at 433 (noting that “prizes given for a particular performance tend to negatively affect intrinsic motivation”).

404. *Id.* at 432; see also Bruno S. Frey & Susanne Neckermann, *Abundant but Neglected: Awards as Incentives*, 2009 ECONOMISTS’ VOICE 1 (noting that “indirect incentives” such as awards that individuals “cannot or do not consciously work towards” can have beneficial effects, such as identifying role models and reinforcing social norms). Using an indirect incentive may also move the reward closer to a credible signal of agent competence, rather than a signal of principal distrust. Frey, *supra* note 44, at 433.

405. Sanjai Bhagat & Roberta Romano, *Reforming Executive Compensation: Focusing and Committing to the Long-Term*, 26 YALE J. ON REG. 359, 359-63 (2009).

such as choking, distraction, or performance orientation) could be addressed by incorporating quality metrics as well as effort metrics in the calculations. For example, if a school board is concerned that paying children to read books will result in a “performance orientation” that induces children to forgo difficult books in favor of quick reads, imposing length or difficulty requirements (or difficulty bonuses) may help to mitigate this concern during the period when the incentive is in force. In the healthcare context, payment for performance provides some instructive lessons on the importance of building flexibility into metrics, enabling adjustments for difficulty. In order to avoid some “gaming” behavior, such as cherry-picking healthy patients who will show improvement, some scholars have advocated metrics that allow exceptions for difficult or noncompliant patients;<sup>406</sup> many have also remarked on the importance of consulting physicians during metric development. It is impossible to avoid gaming behavior entirely, and increasing the complexity of metrics may also increase transactions costs due to monitoring and incentive program administration. But in some contexts, considering quality in metric design may help to avoid some of the downstream consequences of crowding-out effects.

### *E. Bundling Incentives with Information*

Several types of motivational crowding-out may be alleviated through the provision of information, particularly crowding-out due to signaling mechanisms. When the agent interprets the reward as a sign of the task’s value or difficulty—for example, as a signal that the task is dangerous, boring, or unattractive—pairing the reward with additional factual information about the task may alleviate these concerns. Another option is to incentivize only repeated tasks, so that agents are aware of what the task entails before they are incentivized. Some have also advocated that principals should only reward tasks that require “specific high task-performance,” which may increase intrinsic motivation by conveying to agents that the task holds “personal or social significance.”<sup>407</sup>

When the agent interprets an incentive as a signal of his or her competence or trustworthiness, information provision is likely to be particularly important for avoiding motivational crowding-out. Participants should receive informational feedback on good performance;<sup>408</sup> several meta-analyses have indicated that verbal rewards can increase self-reported interest and engagement

---

406. See, e.g., Eijkenaar et al., *supra* note 25, at 116 (suggesting that programs may “allow[] providers to exclude noncompliant patients from performance calculations”); Adam Oliver & Lawrence D. Brown, *Incentivizing Professionals and Patients: A Consideration in the Context of the United Kingdom and the United States*, 36 J. HEALTH POL. POL’Y & L. 59, 71-72 (2011) (citing empirical support for the idea that allowing providers in P4P programs to exclude patients from performance calculations does not lead to widespread gaming of the system).

407. Eisenberger et al., *supra* note 54, at 687.

408. See Stern, *supra* note 46, at 579.

when they include positive feedback on the agent's competency at the task.<sup>409</sup> For example, research on P4P programs for physicians has found that giving physicians direct performance feedback may be as effective as incentives; several studies found that adding P4P mechanisms to feedback programs added little to program outcomes.<sup>410</sup> Principals can also avoid delivering rewards in a controlling way by emphasizing "the interesting or challenging aspects" of the task during feedback.<sup>411</sup>

Similar effects obtain when an agent interprets an incentive as a signal of prevailing social norms. In order to avoid signaling unsupportive social norms, incentives should be paired with information about prevailing rates of compliance with desirable behavior; it may even be more effective to provide agents with information on their performance relative to their peers.<sup>412</sup> Some have suggested that communicating about social norms may be an effective replacement for incentive programs; for example, Kahan has described a strategy for promoting tax compliance that relies on communicating about the high proportion of people who already comply, rather than providing incentives or penalties to discourage evasion.<sup>413</sup> Of course, these strategies depend on the extent to which pre-existing social norms *are* supportive of the incentivized behaviors; if incentives are used to encourage a socially stigmatized task, or a new task that is rarely performed, providing information about prevailing social norms may do more harm than good.

Communicating about the source of incentives may also help reduce crowding-out due to the eviction effect. The eviction effect occurs when an agent believes that the cost of an incentive will harm the principal's mission. It may therefore ameliorate the effect if the principal clarifies the source of the rewards (i.e., funds or a donor source that does not impact the principal's activity).<sup>414</sup> Communicating about the value of the incentive may also help avoid the eviction effect when the incentive is small, but as noted above, identifying the cash value of an incentive may simultaneously exacerbate crowding-out due to framing effects or moral repugnance.

---

409. See Tang & Hall, *supra* note 98, at 379; Deci et al., *supra* note 62, at 638-39, 653.

410. Eijkenaar et al., *supra* note 25, at 125-26.

411. Deci et al., *supra* note 62, at 655-57 (citing research by Richard M. Ryan and Miron Zuckerman).

412. AYRES, *supra* note 182, at 79-86 (2011); Ian Ayres et al., *Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage*, 29 J. L. ECON. & ORG. 992, 993.

413. See Kahan, *supra* note 19, at 342-43.

414. Chmielewski-Ralmondo et al., *supra* note 175, at 1894-95 (suggesting that rewards for blood donations would be acceptable if donated or industry-sponsored, but not if they were purchased using charity funds).

### F. Publicity Choices

Principals designing incentive schemes have the additional task of managing the flow of information (to agents and other observers) about incentives and the identity of their recipients. These publicity choices are primarily important for crowding-out mechanisms that depend on reputational motivations.

The privacy or publicity of incentives and their recipients may be most important for crowding-out due to image-spoiling. If an agent is concerned that others will believe he engaged in good behavior simply to obtain an incentive, one potential remedy may be to make the incentive private, by keeping secret the agent's identity, the incentive amount, or the incentive scheme as a whole. There may also be notice requirements for some incentives, such as negative incentives (penalties), and incentives may be particularly difficult to obscure when they are offered by a governmental principal. Bénabou and Tirole have discussed the option of publicly disclosing rewards that are small, but then allowing the principal and agent to renegotiate reward size privately.<sup>415</sup> Astutely, they have also noted that although avoiding publicity may diminish crowding-out due to image-spoiling, it may do little to avoid the problem of crowding-out due to *self*-image-spoiling. "[T]aking secret rewards does not help with self-image, and may even damage it."<sup>416</sup> An alternative means of averting crowding-out due to image-spoiling or self-image-spoiling may be to give agents the opportunity to *publicly* refuse a reward, or to divert the value of their incentive to charity. Charity options can reduce motivational crowding-out in a number of ways, as described above, but they can *only* work to reduce crowd-out due to image-spoiling if the "give-backs [are] observable by the audience to whom agents are trying to signal."<sup>417</sup>

When there is a concern that the availability of incentives may communicate unsupportive social norms, it may help to remedy these problems by publicizing the number or names of agents who receive the incentive for good behavior. Publicizing a large number of agents may help to demonstrate that social norms favor the behavior, reducing this source of crowding-out effects. When social norms are already supportive of compliance, public appeals reinforcing these social norms may sometimes be more effective than incentives. Regrettably, however, public appeals may not always be good alternatives to incentives; specifically, appeals to conserve resources may actually increase demand.<sup>418</sup> One study, for example, found that a public appeal asking healthy individuals to refrain from getting flu shots (to conserve them for the elderly)

---

415. This may, however, cause several difficulties in practice, including commitment problems that would increase privately negotiated incentives. Bénabou & Tirole, *supra* note 47, at 1669 n.35.

416. *Id.* at 1669.

417. *Id.*

418. See Steven E. Sexton, *Paying for Pollution? How General Equilibrium Effects Undermine the "Spare the Air" Program*, 53 ENVTL. RESOURCE ECON. 553, 555 (2012).

actually *increased* demand because it elevated the salience of shots and reduced healthy people's procrastination.<sup>419</sup>

Instead of relying on public appeals or social norms education, a more powerful strategy to capitalize on reputational motivations may be public reporting on agent performance. Public reporting programs harness social dynamics, reputational motivations, and the desire to punish free riders by releasing information to make social shaming (or reputational payoff) easier. Peer punishment is a powerful mechanism that can "induce the self-interested to act as if they were civic-minded,"<sup>420</sup> and incentives are more likely to trigger crowding-in when they harness "moral approval or disapprobation from a subject's peers."<sup>421</sup> Eric Fleisig-Greene has noted this strategy at work in efforts to publicize the names of individuals who bounce checks, purchase services from sex workers, evade taxes, or fail to pay child support.<sup>422</sup> The extent to which an act is publicly observable may also influence voting behavior. Funk notes that when Swiss jurisdictions offered citizens the opportunity to vote by mail, the option did not increase voter turnout; she suggests that the "nonverifiability of the [postal] voting" behavior may have undermined motivation.<sup>423</sup> Other examples of effective publicity schemes include a program for Thai police officers, who must wear pink "Hello Kitty" armbands when they fail to follow rules,<sup>424</sup> and a program in Mumbai to deter spitting among governmental employees, whose photos and identities are publicly posted if they are caught in the act.<sup>425</sup> Reviews of P4P for physicians have found that publicly reporting on a physician's performance may be just as effective, if not more so, than linking payments to performance indicators.<sup>426</sup> We might also view such naming-and-shaming programs as their own form of incentive scheme, which pays incentives in the currency of reputational impact. Publicity strategies, however, are in tension with concerns about intrusiveness and surveillance in program administration; agents may also experience such programs as controlling, thereby exacerbating motivational crowding-out due to impaired self-determination.

### G. Managing Heterogeneity

Both the measurement and alleviation of crowding-out effects are afflicted by the problem of heterogeneity across agents in their intrinsic motives, the

---

419. *Id.* at 556 (citing Alain de Janvry et al., *Can Appeals to Cooperation Be Effective in Managing the Scarcity of a Vital Good? Responses to the 2004 Flu Vaccine Shortage* (Dep't Agric. And Resource Econ., Working Paper Series 1013, 2008)).

420. Bowles & Hwang, *supra* note 42, at 19.

421. Walton, *supra* note 28, at 438.

422. Fleisig-Greene, *supra* note 132, at 1239.

423. Funk, *supra* note 185, at 156.

424. See Rigdon, *supra* note 154, at 96; Fleisig-Greene, *supra* note 132, at 1240.

425. Fleisig-Greene, *supra* note 132, at 1240.

426. Eijkenaar et al., *supra* note 25, at 125-26.

extent to which they value a given incentive, and their susceptibility to crowding-out effects.<sup>427</sup> This problem can lead to agent selection effects, or it can pose the problem of multiple mechanisms of crowding-out that demand different solutions for different subgroups of agents. For example, Armin Falk and Michael Kosfeld's study found groups of agents that responded negatively, positively, or neutrally to their principal's imposition of extrinsic control in an economic game; although approximately 55% of participants responded negatively to the principal's use of controls, another 27% reacted favorably.<sup>428</sup> It may be difficult, the authors note, to achieve the goal of "disciplin[ing] the opportunistic agents without reducing the motivation of the intrinsically motivated ones."<sup>429</sup> As described above, crowding-out effects may depend on variables like gender,<sup>430</sup> culture,<sup>431</sup> or prosocial orientation.<sup>432</sup> The choice to refrain from incentivizing or controlling selfish agents may be unwise, but the choice to incentivize or control *uniformly* may also undermine the activity of unselfish or prosocial agents due to crowding-out.<sup>433</sup>

When setting incentives for a heterogeneous population, an incentive architect might consider several options for designing a uniform incentive scheme. One is to simply set incentives to motivate the modal agent—the type of agent representing the largest share of the population. Another may be to set incentives for only the agents who are *already unmotivated* or *noncompliant*, because intrinsic motivation is a prerequisite for crowding-out to occur. This strategy has sometimes been used in incentives for health behavior, which are offered to those with unhealthy behavior to stimulate change.<sup>434</sup> Notably, although these designs may avoid crowding-out among the incentivized segment of the population, both may exacerbate agent selection effects (particularly if agents are permitted to move among categories). Even though an unselfish agent may not have wanted a reward for good behavior, she may resent being denied an incentive that is offered to others; this may communicate that her contribution

---

427. See, e.g., Feldman & Perez, *supra* note 27, at 412 (noting the challenge of regulating different "types" of individuals in a community); Shogren, *supra* note 33, at 356 (noting "the regulator's dilemma" when trying to mobilize both motivated and unmotivated populations); Bowles & Hwang, *supra* note 42, at 18-19 (noting differences in "the crowding parameter" across individuals).

428. Falk & Kosfeld, *supra* note 153, at 1624 tbl.2 (pooling the percentage of participants who reacted negatively, positively, and neutrally to the principal's use of control in all three experimental conditions; across 209 participants, 114 reacted negatively by reducing their performance, 57 reacted positively, and 38 reacted neutrally).

429. *Id.* at 1628. Indeed, as Fehr and Falk write, "the negative side effects of . . . explicit incentives . . . do not apply to selfish subjects because these subjects do not exhibit voluntary cooperation." Fehr & Falk, *supra* note 154, at 698.

430. Mellström & Johannesson, *supra* note 58.

431. Magrath & Nichter, *supra* note 170, at 1783.

432. Censolo et al., *supra* note 49, at 1.

433. See Ellingsen & Johannesson, *supra* note 30, at 991.

434. See, e.g., Promberger & Marteau, *supra* note 2, at 954 (noting that incentives for protective health behaviors are often only offered to individuals with low initial behavior levels and that existing evidence on motivational crowding-out, which stipulates high initial motivation, may be less applicable in the health behavior context).

is valued less, exacerbating crowding-out due to signals of low task value, low perceived agent competence, or principal hostility. If an incentive is high enough, agents may desist from their good behavior temporarily, so that they may become part of the incentivized group. Moreover, when the incentive is negative, such as a pro bono requirement for non-serving lawyers,<sup>435</sup> any agent who *disengages* will then become part of the population to which the requirement is applied—this type of incentive is indistinguishable from a blanket mandate.

A third option for designing a uniform incentive scheme to avoid crowding-out, as recommended by Rebecca Hollander-Blumoff and others, is to simply set incentives so high that the relative price effect always dominates over motivational crowding-out. As she writes, “law ought to be careful to pitch penalties at the right—high—level” to overwhelm the crowding-out effect: “those with some intrinsic motivation may find such motivation crowded out by the penalties—but if the penalty is high enough, they will still refrain from the behavior in order to avoid the penalty, even if they ignore the expressive function” of the law.<sup>436</sup> “[B]ecause the law cannot tailor itself to different populations . . . a better focus remains calibrating the penalty high enough to deter both groups.”<sup>437</sup> This may work better for negative than for positive incentives, however, because it may be very costly to principals to provide positive incentives at a high enough level to overwhelm crowding-out effects. This solution does little to avoid motivational crowding-out, but it is a useful instrumental fix when increasing the level of behavior is the primary or only concern. Increasing incentive size may diminish program sustainability (increasing vulnerability to long-term behavior reductions when incentives run out). It may also exacerbate crowding-out due to choking, distraction, performance orientation, moral repugnance, framing effects, the eviction effect, and long-term endogenous preference formation.

In some scenarios, it may be worth the administrative and transaction costs to design a program that differentiates among agents. This can help to reduce crowding-out due to signaling of agent competence or principal hostility, and it can limit inefficient competition among agents. When rewards make no differentiation between agents, agents who believe they have above-average intrinsic motivation may be offended by being treated the same as agents who shirk; over time, this will lead to a reduction in intrinsic motivation among these agents.<sup>438</sup> Frey suggests this is a particular problem in public institutions, where highly motivated employees “over time become increasingly disillusioned,” leading to a lower morale than their counterparts in private industry.<sup>439</sup>

---

435. Loder, *supra* note 224, at 474-75.

436. Hollander-Blumoff, *supra* note 39, at 63; *see also* Sheppard & Cushman, *supra* note 106, at 64 (noting that “monetary incentives . . . must at least be powerful enough to cover the spread created by the crowding out of antecedent intrinsic motivation”).

437. Hollander-Blumoff, *supra* note 39, at 65.

438. Frey, *supra* note 44, at 433; *see also* Frey, *supra* note 68, at 1527-29.

439. Frey, *supra* note 44, at 433; Frey, *supra* note 68, at 1528-29.



One way to achieve better tailoring of incentives to agents is to create a program with decentralized administration; decentralizing programs can take advantage of local knowledge, promoting a better fit between incentive plans and the distribution of types in the local population, as well as more sensitive “fine-tuning” over time.<sup>440</sup> Frey advocates the use of decentralized means for providing state funds to artists, suggesting that support should be channeled through private parties who are able to delineate among individual artists.<sup>441</sup> A second, potentially less effective way of achieving differentiation is to impose incentives or penalties on a sliding scale to account for heterogeneity in agent resources. For example, Sandel describes a \$217,000 speeding ticket for one of the richest motorists in Finland, a jurisdiction that imposes fines based on income.<sup>442</sup> A third strategy, mentioned briefly above, may be the use of a “menu” of incentives from which agents can choose. The use of individualized incentives, however, has drawbacks. These programs are likely to be more difficult to administer than those with uniform incentives, and they may be more likely to “violate liberal legal and ethical norms . . . and prove politically infeasible.”<sup>443</sup>

How should the principal decide whether to use a uniform or non-uniform incentive plan? One strategy is to let the decision of a uniform or non-uniform incentive scheme be guided by the task itself. In this area, a sophisticated taxonomy has been offered by Feldman, who distinguishes between three types of scenarios, with different implications for targeted or “differentiated” regulation: (1) actions that only require participation by a segment of the population (e.g., whistleblowing), for which incentivizing everyone may not be worth the cost of crowding out motivations for the few; (2) actions that require everyone to act (e.g., trade secret protection), for which it is worthwhile to incentivize “the lowest common denominator”; and (3) actions where maximal participation is desirable (e.g., recycling), for which it is helpful to motivate “both populations” of motivated and unmotivated agents.<sup>444</sup> It is in the second category where a non-uniform incentive scheme may be of most use, while tasks in the third category may be best addressed by a uniform incentive targeting unmotivated agents; in the first category, incentives may do more harm than good. Incentives may also be most useful for activities with broad participation goals, such as “a target level of compliance” for vaccinations.<sup>445</sup>

---

440. See, e.g., Feldman & Perez, *supra* note 27, at 437 (“Providing municipal authorities with broader regulatory autonomy can assist them in employing regulatory strategies that better fit the cultures and norms of their communities.”); Stern, *supra* note 46, at 580.

441. Frey, *supra* note 377, at 79-80.

442. See SANDEL, *supra* note 72, at 66.

443. Bowles & Hwang, *supra* note 42, at 18-19.

444. Feldman, *supra* note 2, at 42.

445. Bowles, *supra* note 41, at 1609; see also Bowles & Hwang, *supra* note 42, at 17. As Bowles and Hwang note, the question is whether the principal seeks to “maximize[] the material net benefits of the project,” or “to maximize citizens’ utility including their values.” *Id.*

### *H. Discontinuing Incentive Programs*

The prior Section discussed strategies by which incentive architects may intervene in the mechanisms by which motivational crowding-out occurs, or diminish the downstream behavioral impacts or autonomy harms caused by crowding-out effects. But in some scenarios, crowding-out may be impossible to remedy by the use of “smarter” incentives. One particularly intractable crowding-out mechanism may be endogenous preference adaptation. Besides the use of anti-incentives in the presence of reputational concerns, my review of the literature on crowding-out has identified no empirically-supported strategies to minimize this dynamic, which may mean there is little scope for preventing spillovers and horizontal creep of motivational crowding-out effects. Other scenarios in which crowding-out effects may be especially resistant to remedy include framing and priming (if not alleviated by the use of in-kind incentives) and moral repugnance. Several downstream consequences of motivational crowding-out may also be intractable, including sustainability in the presence of limited incentive resources, distraction, performance orientation, gaming behavior, motivation atrophy (if it occurs), and the erosion of individual or societal values over time.

In these cases, the principal must weigh the benefits of the incentive scheme (including any benefits of crowding-out) against its administrative costs, autonomy impacts, behavioral harms (especially harms externalized to third parties), and attitudinal or moral damage. The decision to discontinue an incentive is unlikely to cure any harms immediately if the lesson from Gneezy and Rustichini’s study (“Once a commodity, always a commodity”)<sup>446</sup> holds true. Additional steps may be needed to reverse the effects of crowding-out, such as public education to rebuild social norms, public disclosure of behaviors to kick-start social mechanisms of punishment or reward, dismantling of surveillance mechanisms originally used to administer incentives, or demonstrations of principal generosity to counteract the perception of principal hostility or intrusiveness. But the recovery of motivations and behavioral levels may indeed be possible, as demonstrated by at least one dismantled incentive program.<sup>447</sup>

As this Article has stressed, there are limits to what incentives may accomplish—but there may also be limits to how much damage incentives can do. In the absence of incentives, there are nonetheless many ways to steer behavioral choices in desired directions, which are the subject of scholarship in fields such as behavioral economics, public health, psychology, and more; some may include simplifying choices to avoid providing too many options, setting

---

446. Gneezy & Rustichini, *supra* note 155, at 14.

447. See Kohn, *supra* note 73, at 3-4 (describing a welding company that removed an incentive program, which resulted in temporary losses but did not affect long-term production rates).

default options, asking individuals to verbally announce whether and how they will implement desired choices,<sup>448</sup> and public education.<sup>449</sup>

## Conclusion

Incentives, both financial and non-financial, are common contemporary tools for steering individual behavior, with applications ranging widely from smoking cessation, to tax payment, to support for local development. Governmental and private institutions alike participate as principals in the deliberate structuring of incentives to accomplish public and private goals. But the discourse regarding deliberate incentives reflects long-standing anxiety over the extent to which incentives—particularly financial incentives—may be ineffective, inefficient, disempowering, or downright corruptive to incentivized parties and society as a whole. The concern about motivational crowding out has been foremost among these worries, and legal scholarship increasingly invokes this literature without a precise understanding of the processes animating crowding-out. To date, the legal field has lacked a classification of different types of crowding effects, and little attention has focused on the specific mechanisms by which crowding-out may impact the outcomes of incentive-based policies. But, as I have argued here, a closer look at crowding may help identify a more precise set of tailored policy design options for limiting the damage that crowding-out may cause.

This Article is the first legal scholarship to demonstrate how motivational crowding-out actually encompasses a heterogeneous set of causal mechanisms and effects that demand tailored responses by incentive architects. In order to intervene in these dynamics productively, we need a practical set of tailored policy options that suggest *how* best to alleviate crowding-out concerns. To that end, I have introduced many dimensions of incentive architecture, which refers to the deliberate design and implementation of incentive-based policies.

Regulators and principals seeking to design incentive-based policies that accommodate crowding-out effects will face challenges. This Article has identified how crowding-out arises from diverse processes, and given that each policy context may be unique, a uniform approach to addressing crowding-out in policy development may be out of reach. But these concluding remarks may note four common challenges. First, it may be difficult to predict crowding-out in advance of program implementation. Initial formative research should consider the types and strength of existing motivations before introducing incentives, with particular attention to image motivations, self-image motivations, and perceived social norms. Crowding-out effects may be most

---

448. Kamenica, *supra* note 142, at 13.11-13.17.

449. Frey & Stutzer, *supra* note 304, at 412-22 [hereinafter “proposition 5”] (also noting that, when incentives are less controlling, agents without an intrinsic motivation to protect the environment may free ride, causing motivated agents to disengage to avoid being “suckers”).

damaging where social norms are strong, and where tasks are viewed as prosocial or done for the benefit of society as a whole. Qualitative research is particularly important when existing motivations are not well known. Pilot testing of incentive programs should also help identify crowding-out effects and potential remedies before incentive-based policies are brought to scale. Second, principals and regulators must make a judgment call on the extent to which these effects demand correction. Some crowding-out may be tolerable, if the disciplining effect of an incentive is large enough—that is, the allure of incentives can outweigh crowding-out, and the resulting levels of behavior may be enough to fulfill the program goals. Third, as noted above, heterogeneity across a population is a particularly thorny problem. Formative research on crowding-out should canvass different types of agents, and although the distribution of agents may be unknown, incentive architecture decisions about tailoring programs should consider whether the overall goal is broad participation (for which incentives may be more useful) or whether performance by comparatively few, intrinsically-motivated agents is sufficient. Finally, little research exists exploring remedies for crowding-out effects; remedies for crowding-out are in their infancy, and regulators have little concrete evidence to guide policy design as of yet.

Despite these challenges, this Article has sought to define crowding-out and its potential incentive-architecture remedies, and to begin a more systematic conversation about incentives, motivation, and the design of incentive-based policies.

