Decentralization Misplaced: On New Paternalism, Skepticism toward Experts and (Praise for) Conspiracy Theories

Nathan Berg

Abstract

Problems—both obvious and subtle—arise from the goal of making people better off “by their own standard” associated with Thaler-and-Sunstein-inspired ‘nudging’ and New Paternalism. Whose model of another person’s “own standard” can be trusted? This essay draws on neoclassical information economics, social welfare theory, and the methodological literature on normative behavioral economics. I enumerate a list of potential costs of nudges used in government policy and other recommendations under the heading New Paternalism—in units measuring informational loss, forgone individual payoffs, and social welfare losses. I analyze communication between a Non-Expert and an informationally advantaged Expert to characterize good and bad communication equilibria and demonstrate why decentralization is helpful for protecting fragile information flows from experts to non-experts. Governments that adopt nudging policies unwittingly introduce complexity to the strategic communication game by substantially expanding both players’ action sets. My analysis then proceeds to the question of socially optimal intensities of skepticism toward expert advice and government speech. Zero skepticism cannot be social-welfare maximizing insofar as skeptical individuals serve as a natural part of ‘checks and balances’ in the political economy. At the other extreme, maximal skepticism is unlikely to be social-welfare maximizing because it
wastes good information that both sides would voluntarily like to transmit and act upon. Therefore, the socially optimal intensity of skepticism is a strictly interior value somewhere between zero and maximal. One understanding of the “slippery slope” that some critics of New Paternalism have advanced is as follows. Because there is risk of other non-transparent objectives (e.g., lobbying) influencing paternalistic policies, the first-order effect of “choice architecture” on individual skepticism is to increase. As policy makers are emboldened to impose increasingly aggressive policy experiments in “choice architecture” under the cover of “objective” social science (behavioral economics, in this case), there is movement down the slippery slope along which individual response functions (e.g. updating of subjective beliefs) rationally select increasingly skeptical views of expert advice and government speech. The social costs from information loss and reduced coordination services that would otherwise have been achieved by decentralization without choice architecture provide two already well-understood economic arguments that can be used to better account for potential costs of proposals from New Paternalism, pointing to far more cautionary approaches to policy and regulation. New Paternalism risks rationalizing increased skepticism which, in its limit, may rationalize conspiracy theories about shrouded objectives influencing choice architects.
Decentralization Mislaid: On New Paternalism, Skepticism toward Experts and (Praise for) Conspiracy Theories

“When it becomes serious, you have to lie.” — Jean-Claude Juncker (2011)

1 Overview

I focus on a situation in which one agent, whom I refer to as the expert, has more information than a non-expert. The expert’s informational superiority is a maintained assumption throughout. Furthermore, I assume that the expert’s informational superiority is accepted as an objective fact by both expert and non-expert alike. The issue at hand is to investigate conditions under which information transmitted by the expert can be believed. The non-expert knows that the expert possesses a set of information which, if transmitted transparently (i.e., without selectively excluding subsets of the expert’s private information, shading or distorting), would (in expectation) enable the non-expert to improve his payoffs.

Based on Berg and Kim (2017), I present a model to characterize good and bad communication equilibria. The model suggests there are actions that an expert can take prior to transmitting her private information—or perhaps legal institutions—that could enable transparent communication as an equilibrium outcome characterized by full informational efficiency and a payoff profile that is Pareto-superior to bad communication equilibria in which experts paternalistically distort, exaggerate, strategically frame, or selectively exclude other contextualizing private information. Even when non-experts want to trust what experts say and act on the information they provide, the model shows that good communication equilibria are rather special and perhaps fragile with respect to nudging. Costly sacrifices by experts may be required in order for them to be believed. Avoiding nudges as an approach
to government policy—and leaving plenty of room for private information transmission, heterogeneous belief formation, heterogeneous behaviors, and decentralization of interpretation and filtering of information—would go a long way toward avoiding the waste of information and social welfare that results from getting stuck in a bad communication equilibrium with government experts. A good communication equilibrium may therefore prove elusive in the sense that there are many bad communication. Heterocultures of belief, action and information transmission can help avoid the waste of information and social welfare that is likely to transpire as a result of well-intentioned paternalism.\textsuperscript{1}

There is a large set of alternatives available to the non-expert when interpreting information transmitted by experts. In a good communication equilibrium (by the definitions I introduce below), the non-expert accepts expert information at face value (i.e., assuming it is expressed neutrally, without nudging). And if the expert has indeed transmitted her private information without distortion, then the non-expert can use that information to improve his payoff.

Interpretation of expert information by non-experts may, in the non-expert’s view, require filtering. If the non-expert believes the expert is nudging or distorting, then his filtering of the expert’s transmission of information could take the form of down-weighting (or otherwise transforming with a function that is not the identity function), on the presumption that the expert tends to paternalistically exaggerate or nudge based on the expert’s view of the non-

\textsuperscript{1}I investigate this situation using rational choice information economics and typical assumptions in this literature (which are stringent). Rhetorically, my goal is to demonstrate the severity of the problems of bad communication equilibrium (ignoring potentially valuable expert information as a rational reaction to experts’ paternalistic communication) even when unrealistic simplifying assumptions are in place (i.e., standard neoclassical rational choice methodology). Extrapolating from this finding, the implication would be that more complex objective functions and larger-dimensional spaces for signaling and choosing actions would likely increase the likelihood—and severity—of landing in a bad communication equilibrium. The modelling approach is borrowed from Crawford and Sobel (1982) and draws on an extension of their model, Berg and Kim’s (2017) “A Good Advisor.” In a subsequent section, I will expand the analysis to focus on skepticism toward government speech (i.e. transmissions of information), its costs, and its potential benefits.
expert’s payoff function which may diverge from the expert’s view of his own payoff function.

One “extreme” case of non-expert interpretation could mean ignoring what the expert says altogether. Ignoring experts could turn out to be irrational or rational depending on whether the expert is distorting her message. Another extreme case of interpretation would be for non-experts to believe the opposite—or do the opposite—of what experts recommend, raising interesting possibilities of pathological strategic communication where randomization could also be employed to reduce the power of experts to pursue paternalistic polices (including, but not limited to, nudging).

Sin taxes are an example of a paternalistic policy advocated by many behavioral economists (and presumably non-behavioral economists, too, e.g., public health experts) which Thaler says should not be regarded as nudges because they change the decision maker’s choice set (and similarly for policies that subsidize behavior that paternalistic experts seek to induce). While sin taxes may not be “nudges,” they do comfortably fit under the umbrella of “New Paternalism”, which takes behavioral economics as a new source of rationalizations of paternalism (Rizzo and Whitman, 2009). A problem with sin taxes is that they can paradoxically induce the opposite behavioral response as their designers intended while incurring a number of indirect social costs (e.g. low-income smokers in New Zealand responding to unusually high tobacco taxes by reducing food expenditures below the household’s nutritional requirements).

Berg, Berg and Ungor (2017) present a rational choice analysis of decisions about how much effort to allocate to changing one’s own preferences. A rational-choice decision maker deliberates how to rationally expend effort, time, money or other resources across two activities: (i) generate hedonic satisfaction given the preferences she currently possesses, and (ii) employ preference-change technology to change the hedonic satisfaction she will derive from the hedonic consumption she expects to have. That model’s main result is that the
demand function for preference-change technology (e.g., to eat less, exercise more, borrow less, become more financially literate, derive more wellbeing from social relationships and less from hedonic consumption) are quite often upward-sloping. Rational preference change generates outsized income effects that, as our simulations show, make rational agents interested in moderating their (possibly “excessive” by some standard) desire for hedonic consumption respond to price changes implemented by paternalistic polices in exactly the opposite direction as policy makers intended.

Before describing the model of expert communication in greater detail, I want to take advantage of the opportunity afforded by Richard Epstein and Mario Rizzo’s conference on “Behavioral Economics and New Paternalism”—and the esteemed group of scholars assembled—to offer a broader enumeration of problems raised by new paternalism advocated more and more throughout the behavioral economics literature. Section 2 presents an eclectic list of problems with new paternalism, many of which have yet to receive attention (on the cost side of the ledger—or any acknowledgement whatsoever as a potential “unintended consequence”) when thinking through benefits and costs of government policies that claim “findings from behavioral economics” as theoretical or empirical rationales for reducing irrationality, backed by the monopolistic force of the government. Section 3 presents a slight re-interpretation of Berg and Kim’s (2017) model of good and bad communication equilibria between experts and non-experts. Section 4 investigates the question skepticism toward experts and governments, allowing for the possibility of both social costs and benefits. Section 5 discusses robust heterocultures of belief and behavior and the question of what normative behavioral economics might seek to measure following the approach of ecological rationality. Section 6 is a note on the recent history of economic thought from Thaler (1991) to Thaler and Sunstein (2008) and the emergence of New Paternalism. Section 7 contains a concluding discussion drawing on Epstein (1995, 1998, 2003, 2011, 2014) to argue that the basic insights
about the social benefits of decentralization, in conflict with New Paternalism, risks being mislaid.

2 Risks of Nudging and New Paternalism

2.1 Risk (a): Correcting behavior inconsistent with consistency axioms can lower payoffs

Proponents of the biases program in behavioral economics define rationality as an absence of inconsistencies with certain sets of axioms of internal logical consistency (e.g., transitive preferences, risk preferences that satisfy the Savage axioms, Bayesian belief updating, time-consistent discounting rates). These consistency axioms have a continuum of free parameter values that permit infinitely many ways of being perfectly rational (i.e., perfectly consistent). Thaler (1991) encourages us to think of violations of axiomatic consistency as analogous to optical illusions.

In the analogy he proposes, he associates axiomatic rationality (i.e., the continuum of behavior patterns or individual choice data that satisfy axiomatic consistency) with objective units of physical measure, namely, measures of physical distance (e.g. metres). This unhelpful analogy fails for several reasons, the most obvious being that—unlike an optical illusion where perceived distance can be corrected by a single, well-defined standard of accuracy—there is no single choice behavior that is maximally rational in Thaler’s analogy (i.e., no single way for ‘behavioral deviations from consistency axioms’ to be corrected).

Yes, the axiomatic standard of consistency is well-defined. But it is both promiscuous in admitting many patterns of behavior to satisfy axiomatic rationality, and it is overly restrictive, ruling out many inconsistent behavior patterns that perform well in the world. Thus, behavioral anomalies reported in the behavioral economics literature do indeed violate
a well-defined standard of internal axiomatic consistency.

But that standard does not tell us what we should do. There are infinitely many ways of being internally consistent (e.g., consistently patient, consistently impatient; consistently risk-averse, consistently risk-accepting; etc.). Because there are infinitely many ways of being rational (by the axiomatic definition) and an even larger infinity of ways of being irrational, optical illusions are a bad analogy for understanding the alleged biases that behavioral economics is commonly associated with. In particular, nudges do not “correct” mistaken behavior by returning them to a high level of performance based on well-defined unit of measure.

When nudgers say that people should eat less, exercise more or save more for retirement, nudgers are arguing that the nudge provides a crutch that “improves rationality” based on inducing (a greater degree of) internal consistency with what people actually want. Thus, nudgers implicitly define rationality as internal logical consistency—following the technical requirements on ordinal preference orders that were developed originally for utility representation theorems and, subsequently, interpreted by behavioral economists as domain-general requirements for human rationality. But unlike the measurement of physical distance where there is a clear objective standard for measuring how accurate perceived distances are, there are infinitely many ways to conform perfectly consistency axioms and satisfy this definition (borrowed from neoclassical economics) of rationality.

Nudges are promoted as restoring rationality merely by reducing internal inconsistencies with a person’s multiple selves or among their multiple preferences. Nudges are promoted as a means of shifting behavior in the direction of internal logical consistency, without necessarily achieving substantive improvements measured in any compelling metrics of human performance.

Counter to nudgers and New Paternalists, there is a substantial literature showing do-
mains in which systematic deviations from axiomatic rationality can generate improvements in own payoffs (e.g., Berg and Hoffrage, 2008, 2010; Bookstaber and Langsam, 1985; Gigerenzer, Todd and The ABC Group, 1999, Todd, Gigerenzer and the ABC Research Group, 2012; Hertwig, Hoffrage and The ABC Research Group, 2013). Perhaps as a rhetorical work-around, the behavioral economics literature seems to have focused on documenting biases and irrationality and reporting those deviations as the headline finding rather than investigating their economic costs (or benefits). That deviations from axiomatic rationality “must be costly” is an auxiliary and largely unexamined assumption in many studies in this literature.

Thousands of behavioral economics papers report damning rates of “irrational” behavior (i.e., internally inconsistent) without reporting mean payoffs among rational versus irrational types (Berg and Gigerenzer, 2010; Berg 2014). For all the unfair criticism by non-economists of economists’ narrow focus on money, it is astonishing how little interest this literature documenting behavioral anomalies appears to have regarding its costs (e.g., reporting mean contrasts breaking out mean earnings, wealth, health, happiness, etc., in the axiomatically consistent versus inconsistent subsamples). It seems that documenting instances of logical inconsistency in the behavioral economics literature has crowded out careful analysis of the costs (and benefits) of logically inconsistent decision rules—whether these are compared by the metrics of mean payoffs, income, wealth, lifespan, self-reported happiness, numbers of friends, etc.! More specificity about the costs and benefits of inconsistent behavior and mores straightforward justification of any performance metrics invoked would be helpful. Doing so would at least help focus New Paternalists’ nudge program on reducing inconsistency in settings where inconsistency has been demonstrated to be damaging in substantive

2See Binder (2010; 2013) for further critique of paternalism from a happiness perspective; Bruni and Porta (2007) regarding subtleties and multi-dimensionality of happiness; and Khalil (2017) for a rational choice interpretation of Adam Smith’s ‘sympathy’ and pro-social preferences.
ways—and avoid unwittingly harming the performance generated by useful inconsistent behavior rules in the domains where they perform well (c.f. Caplan’s [2000] analysis of a ‘rationally irrational’ individual who optimally allocates irrationality in domains where it is least costly).

Behavioral biases and “foibles” are not analogous to optical illusions. From Berg’s (2003) “Normative Behavioral Economics”:

Thaler (1991, p. 138) argues that the major contribution of behavioral economics has been the discovery of a collection of “illusions, completely analogous to optical illusions. Thaler interprets these illusions as unambiguously incorrect departures from the rational or correct way of making decisions. Thaler is explicit in accepting neoclassical axioms of individual preferences (e.g. transitivity, completeness, non-satiation, monotonicity, and the Savage axioms which guarantee that preferences over risky payoffs can be represented by an expected utility function) as his normative ideal when he writes: ‘It goes without saying that the existence of an optical illusion that causes us to see one of two equal lines as longer than the other should not reduce the value we place on accurate measurement. On the contrary, illusions demonstrate the need for rulers!’

Yet, in showing that human decisions contradict the predictions of expected utility theory, there is no analogue to the straight lines of objectively equal length. Unlike the simple geometric verification of equal lengths against which incorrect perceptions may be verified, the fact that human decisions do not satisfy the axioms underlying expected utility theory in no way implies an illusion or a mistake. Expected utility theory is, after all, but one model of how to rank risky alternatives. . .
2.2 Risk (b): Individually costly “biases” can improve aggregate payoffs

Systematic deviations from axiomatic rationality can, in surprisingly wide classes of decision domains (or environments), generate Pareto improvements (i.e., increases in aggregate payoffs or wellbeing). Although there are individual costs to having biased beliefs, they can provide a market-wide liquidity improvement that more than offsets their costs. This finding holds even in relatively orthodox models of financial market micro-foundations with no exotic preferences or pro-social motives—where subjective beliefs deviate from objective frequencies and are indeed individually costly. A heterogeneous ecology of subjective beliefs (perhaps comprised of mostly false subjective beliefs) can, however, confer important social benefits (Arthur, 1984; Berg and Lien, 2005; Berg and Gigerenzer, 2007; Gintis, 2007). To intervene, to de-bias, or to nudge in the direction of fewer (or less severe) deviations from axiomatic consistency would risk perpetrating an aggregate-payoff-decreasing shift from Pareto-superior to Pareto-inferior allocations or action profiles.

2.3 Risk (c): Nudging toward axiomatic rationality may induce worse payoffs even when perfect rationality is first-best

Following Lancaster’s theory of the second best, another normative problem with the nudge program arises. If a nudge does not succeed fully at shifting logically inconsistent behavior to achieve perfect axiomatic consistency, why are we confident that payoffs should improve? If a nudge does induce more behavioral consistency, then (following Lancaster’s analysis) we should worry that ‘somewhat more consistent’ behavior induced by the nudge may leave the individual worse off. Without guaranteeing that the nudge succeeds at achieving fully optimal behavior in a given domain, there is the important theoretical possibility that moving
an individual who deviates from full optimization in the direction of optimization may make him or her worse off, as the following trivial example demonstrates.

Consider an agent choosing a triple \( x \), whose veridical (not small-world) payoff function is \( u(x) = x_1^{0.89}x_2^{0.10}x_3^{0.01} \). Boundary solutions achieve minimum utility (zero). Therefore they can be ruled out. The first-order conditions for an interior optimum require equating marginal rates of substitution with price ratios, which (given Cobb-Douglas utility) implies allocating expenditures with the following expenditure ratios: 0.89: 0.10: 0.01.

In the biases program promulgated by nudgers, the agent is “behavioral” because she fails to maximize. Suppose that she uses a ‘1/N’ heuristic by allocating equal expenditure to each of the three goods regardless of their prices, violating all first-order conditions (\( MRS_{ij} \neq p_i/p_j \) for all \( i \neq j, i, j \in 1, 2, 3 \)). Failure to optimize is costly and easy to verify.

Now suppose a nudger enacts a policy that induces partial conformity with the first-order conditions. Suppose the nudge succeeds at inducing the agent to satisfy one of the first-order conditions (which I interpret as nudging the agent toward rationality\(^3\): an optimal ratio of expenditures, \( p_2x_2/p_2x_3 = 10 \). Of course there are many ways to respect one first-order condition that give lower utility than the heuristic does. Suppose, for example, she spends her entire budget on goods 2 and 3 in the correct ratio, leaving zero units of \( x_1 \) and, therefore, minimal (zero) utility—despite having responded in the desired way to the nudge. The agent would have been better off continuing to use the heuristic than following this nudge which induces partial conformity with optimization conditions.

It is trivial demonstration. But the relevance of the point to the nudge program is worth

\(^3\)A nudger might object that nudges, by definition, seek to improve payoffs and therefore “satisfying one more first-order condition” is not a nudge. This objection raises the spectre of tautological reasoning, however. If a policy intervention must raise payoffs in order to be a “nudge,” then nudges can never fail to raise payoffs (by definition). Unsuccessful attempts are not nudges. Successful ones are. My point is that, among the ways we might try to help this agent raise her payoffs, one (possibly) reasonable way would be to recommend satisfying the first-order conditions. If a nudge only gets her part of the way toward perfect optimization, then the nudge may have hurt her payoffs.
emphasizing. The choice problems that regulated citizens of the state face are so radically heterogeneous that there is little hope for a nudger, in general, to be sure if “nudging people to more closely conform with axiomatic rationality” winds up achieving a net gain or loss in payoffs. In general, we do not know the answer to this question. Therefore, I argue that a prudent state would, in general, avoid engaging in nudges, given that they can (and are sometimes likely to) induce more harm than benefit.

2.4 Risk (d): Lost benefits of heterogeneous ‘ecologies’ of belief and action

When considering benefits and costs of pursuing New Paternalism policy making, the benefits of heterogeneity itself are worth considering. If the goal of the policy is to induce as many people to follow the recommended behavior as possible, then it stands to reason that the policy’s end goal would reduce some forms of heterogeneity. Insofar as government nudges are successful at influencing a population to conform more strictly to a profile of beliefs and behavioral targets that experts deem to be “best,” “optimal on average,” or “optimal for the average person,” then heterogeneity of beliefs and behavior can be expected to decline (all else equal). It is not inconceivable that fewer discoveries, lower rates of innovation, and more psychological hardship among those with unusual or outlier points of view might transpire (even though there may be a smaller number of such individuals if the nudging programs realize their intentions).

For example, as improbable as it sounds, having a population that includes both smokers and non-smokers might enable us to discover as-yet-undiscovered benefits that some people derive from smoking. Or not! An analogous observation is that there have been scientific and epidemiological findings suggesting some benefits from moderate alcohol consumption published in recent decades that would never have been discovered had the prohibition
movement employed potent nudge technologies to achieve (nearly) universal non-drinker status.

Heterogeneous belief formation, heterogeneous procedures for making inferences and the portfolio diversification benefits of heterogeneous behavioral profiles are all potentially at risk if New Paternalism succeeds at reducing individual heterogeneity. Given the role that inductive versus deductive reasoning plays in creative endeavors and Philip Kitcher’s observations about holding inconsistent thoughts as a common pre-cursor to scientific discovery, the New Paternalist’s goal of reducing inconsistency of beliefs and behavior writ large may carry unanticipated risks and costs. The species-level biological benefit of heterogeneity itself is analyzed by Bookstaber and Langsam (1985).

A close variant of the benefits of heterogeneity are the vulnerabilities of monocultures (i.e. reduced robustness). Although it may seem frustrating that experts are not able to persuade higher rates of conformity, we might also appreciate that resistance to monoculture and maintenance of heterogeneity thanks to individuals or social groups that do not uniformly adopt current views about “best practices.” Not only does heterogeneity afford greater rates of discovery about new understandings of what constitutes best practices, it also helps avoid harm through portfolio diversification whenever current best practices are found out to be wrong.4

2.5 Risk (e): Reduced transmission of valuable information

See the model of expert and non-expert communication presented below.

4Do we really wish that potent nudge policies such as ‘Save More Tomorrow’ had been put in place in Japan to induce greater exposure to risky equity markets among Japanese savers during the 1980s along the long run-up to its peak in 1989 (trading in recent months of 2017-2018 at a little over 50% of its 1989 peak)? Given the frequency of reversals of experts (e.g., health, nutritional and financial recommendations), might we not benefit from a decentralized approach for transmission of valuable information—without admonishing and lamenting high population variability in following expert recommendations, but rather valuing it as veritable social good?
2.6 Risk (f): Reduced payoffs from reduced transmission of information

Reduced information flow and reduced payoffs from reduced information flow are conceptually distinct (although equivalent for many types of analysis).

2.7 Risk (g): Increased cognitive burden from increased complexity of the communication game with government

As shown below, nudging introduces complexity and shifts the communication game between non-experts and government to one with higher-dimensional, more sophisticated communicative action sets.

2.8 Risk (h): Reduced psychological wellbeing from increased complexity

“Increased complexity” here refers to the communication games that a nudging government forces everyone to play. Greater complexity requires more effort in filtering and interpreting information that governments now bundle together with implicit or explicit communication strategies designed to exert influence over the population’s behavior. It is hardly outlandish to imagine new problems that this could lead to. One such problem is newfound skepticism (or increased skepticism) about the intent and strategic thinking underlying the government’s transmission of information (which perhaps was previously more straightforward to interpret).
2.9 Risk (i): Greater license to actively influence behavior may raise the likelihood that a nudging government becomes tyrannical or otherwise abuses its “expert” status

2.10 Risk (j): Dignity forgone

While acknowledging arguments that Conly makes against the charge that nudges diminish human dignity, I think that as long as we observe people who do perceive insult, indignity and estrangement from a government that does not value their conception of an individual’s responsibility for his or her beliefs and actions the way the individual would like, that such harms should also at least be acknowledged; some people simply want to be left alone, which is a disposition that welfare economics could try to take more seriously, although implementing the desire “to be left alone” using a Bergson-Samuelson social welfare function may prove challenging for economic modellers insofar as process-dependent preferences over social allocations and policies are invoked (i.e., cannot be easily modelled using consequentialist payoff functions);

2.11 Risk (k): Autonomy forgone

Reduced objective level of individual autonomy in government’s eyes and, as a result of its paternalistic policies, conditioning individuals to become less responsible, shifting cultural norms away from individualistic views toward a view of government responsibility in which their own decision-making faculties have been judged to be pathological. Perhaps social isolation from harbouring a minority view critical of government’s paternalistic policies becomes more painful when the government takes on new, wide-ranging paternalistic initiatives.
2.12 Risk (l): Nudges weaken the signaling value of good behavior, crowd out social reward, and weaken social structure founded on volunteerism

If everyone is pressured to achieve a more rigorous profile of “good behavior,” then the social reward (and social meanings) from autonomously having chosen those good behaviors will change and likely diminish. We may lose valuable information about prospective partners, friends or employees by shrinking channels through which voluntary behavior sends important signals to others and strengthens their social networks. If everyone were forced to be vegetarian, then one can imagine that social forces which unite vegetarian communities would likely change or diminish. If everyone were forced to listen to classical music, then one can imagine life-long opera fans and hobbyists looking for new ways to distinguish their intrinsic appreciation. Or perhaps just as likely, such paternalistic proposals—albeit far-fetched—could be expected to dissolve some important communities and social structure as a function of reduction in the range of population-wide heterogeneity (i.e., conformity with the government’s paternalistic behavioral target).

2.13 Nudging generates potentially harmful complexity

A brief explanation of the problem of increased dimensionality of the decision problems that both information transmitters and information receivers face if experts give themselves license to nudge is as follows. Suppose the expert possesses a scalar-valued piece of private information \( \omega \) that she is asked to transmit. If she feels bound to either share verbatim or not, then she would face a binary choice (to reveal \( \omega \) or not). Whether she views her decision problem as binary choice, discrete choice from a finite list (of subsets of the continuous random variable’s partition), or choosing a scalar value from the real line, the decision
problem—without nudging—is one-dimensional. (In Crawford and Sobel’s setup, she chooses a (coarse) subset of the partition of the support of $\omega$ without revealing its value precisely.)

Now suppose the expert ascribes to nudge theory. She seeks to paternalistically frame her report with other pieces of information (private or public). If she possesses a $K$-vector of private information $\omega$, then she now must think strategically about which value of $\omega$ to report and which other pieces of information (and values of those other variables) to report. Her action set becomes more complex. Complexity increases in the precise sense of dimensionality of the expert’s action set or choice set because nudging implies she uses more variables to reason about strategic communication.

Even if the expert binds herself to reporting only verbatim values of any elements of $\omega$ that she chooses to report, her decision about which subset to report causes her choice menu to expand to become the power set of the elements of $\omega$ with cardinality $2^K$. This set of all subsets may not be easy for her to rank (i.e., may contain non-comparable alternative framings).

The non-expert, realizing that experts strategically choose framings, is thrust into a new communication game with substantially larger action sets and more complexity. Reasoning about the expert’s more elaborate action set that involves choice over framings adds complexity, cognitive burden and, no doubt, gives rise to skepticism about experts’ intentions and interests when trying to influence others. It might then be reasonable for non-experts to apply more complex informational filters to help them interpret both the information in the expert’s possession—which they wish to have and act upon—and the intent of the expert. Due to increasing complexity, skepticism about the expert’s paternalistic intentions, or a combination of these, the non-expert may rationally choose to ignore an expert’s advice altogether.
2.14 Negative externalities from nudging that follow from Risks (a) through (l)

Considering the benefits and costs of nudges in public policy, this risk of easily doing inadvertent damage to the transmission of valuable information by introducing strategic reasoning into the government’s communication strategy should be considered. As an example, consider easy-to-observe communication patterns between the Transportation Security Administration (TSA) and travellers using US airports. The TSA communicates frequent warnings and admonitions, including looped recordings that repeatedly remind travellers to report unattended bags and that carrying liquids or using mobile devices onto airplanes may jeopardize flight safety. These warnings are ignored by many travellers or viewed with skepticism, perhaps rationally, given scientific debate over the empirical evidence justifying TSA restrictions and the fact that some countries with excellent flight-safety records have chosen not to implement similar restrictions.

It is an open question whether frequent warnings and unusually severe restrictions have improved transmission of information. The TSA is suspected by some fliers of exaggerating threats, even by those who regard its objectives as legitimate, accept that TSA does indeed possess an informational advantage, and would benefit from receiving timely warnings containing valuable pieces of the TSA’s private information. But if travellers believe that the TSA exaggerates and begin to filter out incessant TSA warnings with frequencies they regard as unrepresentative of the objective risks, then travellers are likely to become conditioned to discount other messages about airport security which they might have benefited from had both parties maintained a good communication equilibrium (i.e., transparent, without deliberate distortion by experts or discounting by non-experts).

A second negative externality when nudges are used in government policymaking, which is related to information loss yet conceptually distinct from it, is the behavioral change and
lost payoffs associated with the loss of information. One may object that we should only care about lost information if it translates into lost payoffs. My reason for maintaining the distinction between lost information and lost payoffs from that lost information is perhaps similar to Sugden’s (2018) normative analysis using the opportunity criterion. Sugden shows that removing strictly dominated future choice sets can make the person worse off, not in the payoff sense, but in the sense of having lost the opportunity to change one’s mind (i.e., the lost opportunity to be a responsible agent with maximal choice sets along different continuation paths which permits option value for future preference change). Insofar as nudging policies risk causing a bad communication outcome with reduced transmission of information (described in previous paragraphs), then I claim we should measure this loss by at least two metrics. The first is some measure of information loss in appropriate units (e.g. 1 minus a normalized coefficient that non-experts apply to experts endorsed by the government, which could range in magnitude from 0 to 1; signal-to-noise ratios; or other information-theoretic metrics). A second and distinct metric would capture the economic loss attributable to information loss in appropriate units of payoffs or opportunity.

A third potentially negative externality from governments adopting nudge policies—which behavioral economists have commented on surprisingly little, if at all (as far as I am aware)—is the increased cognitive burden imposed on all agents in such communication games whenever the action sets in that game are enlarged by introducing sophistication. Sophistication (i.e., more complex strategic reasoning by experts in the formulation of nudges) among experts, in turn, demands greater sophistication (enlarged action sets) by non-experts to figure out how information should be interpreted or filtered. I think a likely response to this increase in complexity is more frequent ignoring, more frequent transmission failure of vitally important information, and a noisier information field when information transmissions from governments are more frequently bundled with paternalistic intention to influence choices.
the individual makes leading to greater cognitive burden, less information and perhaps even greater psychological stress. Although there is a range of opinion in the research literature on paternalistic nudges as to whether they can be construed as an assault on the individual’s dignity and autonomy, the observation that there are many of us who find the idea of governments implementing nudges unpleasant should, one would think, at least be integrated into social welfare function analysis. Often times, the nudges based on experimental studies that document logical inconsistency of the mean subject (even in between-person studies by Kahneman in which no single individual was ever observed to behave inconsistently) are sold on the premise that an established decision-making bias can be undone. The presumed welfare improvements to be achieved by inducing more logical consistency through nudges draw on scant (or altogether absent) evidence that logical inconsistency generates any meaningful economic costs, individual, social or otherwise (Berg, 2003, 2014a, 2014b; Berg and Gigerenzer, 2007, 2010; Berg, Biele and Gigerenzer, 2016). See Jolls, Sunstein and Thaler (1998) on the law and economics of de-biasing allegedly biased individuals, or Sunstein and Vermeule (2008) writing approvingly of proposals that government infiltrate and sabotage online communities to groups which express skepticism toward information provided by the government—endorsing that governments sometimes disseminate false information.

Rather than condemning skeptics and labeling them conspiracy theorists as Sunstein (2016) does, the ecological rationality approach to normative economics—which sees a potentially beneficial role for heterogeneity (in beliefs and behaviors)—implies no need to intervene and force conformity to a single set of norms. In fact one-size-fits-all thinking leads to many potential problems including the replacement of heterocultures in the policy space with monocultures, thereby losing information (as previously mentioned) and discoveries afforded by novel beliefs and behaviors (Berg and Maital, 2007).

Ecological rationality explains why individual decision rules that deviate from axiomatic
rationality can perform to a sufficiently high level of performance in a particular class of decision domains (Gigerenzer and Goldstein, 1996; Gigerenzer and Selten, 2001; Gigerenzer, Todd and The ABC Group, 1999; Berg and Gigerenzer, 2006, 2007, 2010; Berg, Biele and Gigerenzer, 2016; Berg, 2010, 2014a, 2014b, 2015, 2017). It also explains less-is-more effects in regulation and institutional design (Bennis et al, 2012)—for example, the virtues of simple legal codes (Epstein, 1995); light regulation of ‘tragedy of the commons’ problems in public goods (Berg and Kim, 2015); beneficial price discrimination in public healthcare (Berg and Kim, forthcoming a); gains from decentralization of social assistance (Berg and Gabel, 2015, 2017a, 2017b); meritocratic education institutions based on simplicity (Berg, 2009; Berg and Nelson, 2016); name recognition and scientific networks (Berg and Faria, 2008); decentralized information sharing (Finin et al, 2009); problems caused by adding new protected classes under anti-discrimination law (Berg and Lien, 2002, 2006, 2009); decentralized solutions to Schelling’s neighbourhood segregation problem (Berg, Hoffrage and Abramczuk, 2010; Berg, Abramczuk and Hoffrage, 2013); government valuation of real estate (Berg, 2006a); labor market policy (Berg, 2006b); pitfalls of anti-defamation law and regulation of expression from an information economics point of view (Berg and Kim, forthcoming b); the rational function of premiums paid for financial services from Islamic Banks (Berg and Kim, 2014, 2016; Berg, El-Komi and Kim, 2016); and pitfalls of financial market regulation based on claims of irrationality from behavioral economics (Berg and Lien, 2003, 2005).

Consider the burden on cognitive load and the psychological costs of a shift from non-nudging government speech toward expanded use of nudges. (Let us assume for now that non-nudging government speech is possible, which I admit may not always be the case when, as Thaler and Sunstein (2008) observe, there is no neutral description of the menu). Suppose that the information transmitted by government offices (e.g., CPI data from the BLS, regulatory guidance from financial and bank regulators, environmental regulators, health
regulators, the FDA, the TSA, etc.) formerly consisted of unframed and un-nudged reporting of "objective facts" which I (as non-expert) could previously accept at face value without much filtering or deliberation about underlying intent. But then a new generation of nudge-inspired policy makers enter government. Panels of behavioral economists are brought in to "help people perform better by their own metric of performance." Defaults, framings, cooling-off periods, opportunity-constraining commitment devices, sin taxes, bans, etc., are implemented based on a growing enumeration of alleged biases in the behavioral economics literature. Because government’s information transmissions are now bundled with strategic decisions aimed at influencing the non-expert’s choice behavior, the task of aggregating and acting on this information may be substantially more complex, burdened by new skepticism about the government’s presentation of facts, and more cognitive requirements—possibly resulting in less net utilization of the available information, more stress and reduced wellbeing.

2.15 Tobacco sin tax in New Zealand

Several of the papers at this conference on Behavioral Economics and New Paternalism, a number of papers have used examples of government policies toward tobacco and smoking behavior. A useful example of some unintended consequences of sin taxes as advocated by some behavioral economists concerns tobacco taxes in New Zealand. In 2010, a Select

\footnote{Sin taxes do not necessarily fit under the heading of “New Paternalism” because they do materially change the decision maker’s choice set and perhaps should therefore be excluded from critical analysis of New Paternalism. Thaler writes in a 2017 tweet: “We don’t consider a tax to be a nudge.” It does not stretch the imagination too far, however, to think of behavioral economists advocating revenue-neutral sin taxes as a legitimate component of their nudge-inspired behavioral policy toolkit (where revenue raised is returned as subsidies for smoking cessation services, vegetables or copies of \textit{Nudge} and therefore interpreted as retaining the mean consumer’s bundle as feasible in the post-policy choice set although consumption possibilities spaces for all consumers are materially altered thanks to modified relative prices owing to sin taxes and subsidies). I can already hear someone proposing a Kaldor-Hicks criterion defending revenue-neutral sin taxes as somehow libertarian or choice preserving (for the mean consumer)—that losers from relative prices due to sin taxes can be compensated by winners, and that both will be “better off anyway”, once their logical, deliberative self realizes long-run improvements in wellbeing from a fractional extra year of expected lifespan or Quality Adjusted Years of Life (QALYs).}
Committee in Parliament proposed a policy of New Zealand being smoke free by 2025, which drew widespread support from both the governing coalition and opposition parties. Parliament then implemented a series of tax increases on tobacco. The current cost of a pack of 20 cigarettes is $25 to $30 (78% federal tax) with plans to continue increasing sin tax rates in the range of 10 to 40% per year, with a target of pricing a pack of cigarettes at more than $100 by 2025. For low-income households that do not respond to price increases the way paternalistic policy makers want, a damaging substitution effect has now been uncovered in low-income NZ households with smokers reallocating limited disposable income away from food and into tobacco expenditure (without necessarily increasing physical consumption of tobacco) and, as a result, reducing nutritional wellbeing for smokers and their families members (especially children). This tax now accounts for more than 2% of the national government’s revenue and some not-at-all libertarian Members of Parliament who previously supported sharp sin taxes are changing their minds, saying that the policy has become a tax grab that harms low-income families to an unacceptable extent.

2.16 Rational choice model of preference change implies high likelihood of unintended behavioral responses to sin taxes

Berg, Berg and Ungor (2017) consider a rational-choice model of reasoned, deliberative preference change. The model is a textbook budget allocation problem between effort that raises the mean level of future expected hedonic consumption versus effort expenditure on a preference-change technology that lowers the minimum level of hedonic consumption required to avoid disappointment with one’s position in life. The time/effort/resource budget can be allocated only to these two endeavors: raising the mean level of hedonic consumption or preference-change-technology that reduces the amount of hedonic consumption required to satisﬁce. Solving the expected utility objective for an agent who deliberatively reasons about
trade-offs of having “more” versus “being satisfied with less” (which requires effort expended into an exogenously given preference change technology), a demand function for preference moderation is derived as a function of its relative price. A numerical analysis of the models parameter space shows that upward-sloping demand curves for preference-change technology are rather common. Nothing unusual or pathological is required. The implication is that sin taxes that make “unhealthy” eating or “excessive” borrowing more expensive and subsidies that “incentivize” people to eat healthier or improve their financial literacy can lead to the opposite of the intended behavioral change. In our rational choice model of demand for effortful or costly preference change, income effects are so strong that subsidies do the wrong thing. For more instances of allegedly irrational decision-making procedures that nonetheless perform well in particular classes of investment decisions, see Berg Also see Berg and Lien (2003, 2005), Berg (2014b), Berg and Kim (2014), Berg, Prakhya and Ranganathan (2017), Goldstein and Gigerenzer (2009), Berg, El-Komi and Kim (2016), Monti et al (2012, 2014) and Akhtaruzzaman, Berg and Hajzler (2017).

2.17 Methodological note on New Paternalists’ mis-measurement of the individual’s objective function “by her own standard”

Suppose we grant that New Paternalists have identified settings in which a nudge or intervention could, in theory, succeed at helping people “by their own standard,” there is a basic technical challenge of measurement error that forces a trade-off between the benefits of payoff gains that paternalistic intervention can achieve and imprecision in the behavioral target that the intervention aims to achieve. New Paternalists would, I think, readily admit that when we write a model representing the imperfect agent’s payoff function, we undoubtedly miss many aspects or competing objectives that are important to many individuals whose preferences the model is intended to represent.
Suppose we assume that a representative agent \((i)\) drawn from the population under consideration is to take action \(x_i\) which New Paternalists believe is frequently chosen sub-optimally. The New Paternalist assumes that the phrase "by their own standard" is subsumed (subsumable) by a scalar valued objective function \(u_i(x_i)\) that represents \(i\)’s preferences (while admitting, perhaps over drinks after the seminar presentation is finished, that \(u_i(x_i)\) is not perfectly observable). If we are willing to assume that \(i\)’s objectives are commensurable and can be represented by a utility function, then we could posit the existence of the true but imperfectly observable objective function and represent it as:

\[
v_i = u_i + e_i,
\]

where \(e_i\) captures “individual heterogeneity” that was abstracted away from when constructing the model \(u_i = u(x_i)\). The New Paternalist faces the challenge of the measurement error represented by \(e_i\).

In the rhetoric of Samuelson-inspired neoclassical modelling, \(u_i\) (it is hoped by the modeller) is a reasonable first-order approximation. It somehow “captures” the first-derivative effects on wellbeing of changes in action vector \(x_i\) while the effects of all the missing variables or imperfection in the mapping from \(x_i\) to \(v_i\) absorbed in error term \(e_i\) are “second-order.”

Following this logic, we could say that there is an approximation cost represented by some metric of \(e_i\), represented for example by absolute error, \(|e_i|\), or a percentage deviation—either \(|e_i/v_i|\) or \(|e_i/u_i|\) (assuming they are scaled in such a way that they exist) could serve the purpose. But how to measure measurement error in the representation of an individual’s preferences when used for purposes of paternalistic government policy?

This seemingly routine technical challenge points to numerous methodological problems and opens the door to myriad instances of discretionary methodological choice by economists charged with designing such policies. There is no domain-general solution to the problem
of choosing an appropriate loss function to analyze the costs of mis-representing individual preferences in paternalistic policy design.

If \( i \) were balancing on a tight wire in a circus performance without a net, then the loss function for a continuously valued \( x_i \) and \( u_i \) might look a binary function. What about comparing different ways of dying? Would this require somehow ranking different ‘negative infinities’? It is not terribly challenging to imagine real-life situations where paternalists already involve themselves where individual a bad outcome the paternalist is seeking to avoid—or a pleasure that the \( i \) wants to pursue—is regarded by the individual as incommensurable with other choices she makes. This would mean that the New Paternalist’s method had already veered off course by assuming the scalar-valued objective function \( u_i \) exists.

A more radical rejection of the scalar-objective function approach to modelling an agent’s domain-general ”by their own standard” is the idea that reasonable and rational people in a complex and fundamentally uncertain world will, in general, have no way to subsume all domain-specific objectives affected by their actions \( x_i \). Suppose that when I’m thinking about time trade-offs, I can come up with a most preferred sequence of work-consumption plans, \( x_i \). But when thinking about leaving my job, accepting a pay cut and starting a business, now a different \( x'_i \) looks best. When I fall in love and decide to quit my job to move to a far-away place in my girlfriend’s hometown, yet another decision \( x''_i \) looks rational to me. Should I be condemned as irrational for being open to transcendent religious conversion, to artistic endeavour, to unanticipated entrepreneurial opportunity, and to myriad reasons—including some good ones—to change my mind? Even if New Paternalists put their name-calling aside, the regularity by which successful people (by their own standard) do in fact change their objectives inflicts a mortal methodological wound on the intellectual underpinnings in the behavioral economics literature that New Paternalists want to claim provide justification for new paternalistic policies.
Rational people should generally expect their objectives—that is, “their own standard” for evaluating feasible decisions $x_i$—to change through time and across context even at what is virtually a single point in time. The sole requirements for rationality in the new paternalism world are the axioms of internal logical consistency mentioned above, which are technical requirements for the utility representation theorems that enable them to model “by their own standard” with a scalar-valued objective function (whether it be veridical $u_i$ or approximation value $v_i$). Either way, it might be that, while such abstractions provide valuable insights about surprising mechanisms that generate some forms of observed economic behavior, the paternalism program’s dependence on these same assumptions required for scalar-valued representation already imposes far too much potentially damaging misspecified structure on rationality. In the context of paternalistic policy making, the stakes rise. Technical requirements that morphed into behavioral economics’ normative standard of rational behavior results in a dogmatic and narrow conception of “by their own standard.”

Although it is dogmatic and narrow, it is also too loose! New Paternalists’ standard of rationality provides no upper or lower bounds on rationality. It rules out many high-performing decision-making procedures that successful and healthy people use, which quite reasonably vary by context, domain and over time. Their standard of rationality also admits as rational (read “perfectly internally consistent”) that are commonly regarded disciplines outside economics (and probably most people) as pathological.
3 Model of communication equilibria between experts and non-experts

3.1 Motivation for communication games between experts and non-experts

Many institutions (both formal and informal) have a structure in which, by design, expert information is meant to influence non-experts. Such institutions would include expert panels of behavioral economists giving policy recommendations that advocate paternalistic nudges. Institutions involving transmission of information from experts to non-experts would also include other kinds of authorities (i.e., not behavioral economists) on various subject matters seeking (or invited) to influence lawmakers, regulators, the citizenry or private clients. Expert testimony before law-making and regulatory bodies, expert witnesses in court, educational institutions where expert information possessed by advisors is disseminated to students, and parents transmitting information to children all share this structure. Importantly, private markets for information, advising and consulting services feature transmission of information from experts to non-experts as well.

Even when recipients of advice understand that experts providing said advice indeed possess valuable and actionable information from which they could benefit, it is not uncommon to observe genuine experts struggle to be taken seriously by those to whom they transmit information. Sincere academic advisors—even those with sophisticated communication skills and pedagogical talent—frequently fail at the task of transmitting information they possess (e.g., career advice, the availability of scholarships, the economic and humanistic value of formal education, etc.). Apparently, it is substantially easier to become an esteemed expert than it is to become one who can reliably transmit information generated by that expertise.
3.2 Rational ignoring

When non-experts ignore what experts say and advise, it is, once again, easy for behavioral economists to interpret this as a “foible” (Hausman, 2018), “bias” (Gigerenzer, 2018), or “irrational” information updating (ignoring of the kind that hundreds of papers on non-Bayesian updating have documented and evaluated normatively as pathological belief updating). In previous work, I have demonstrated that ignoring payoff-relevant information can be consistent with expected payoff maximization (Berg and Hoffrage, 2008)\(^6\) and that non-Bayesian belief updating can correlate with improved (not inferior) objective accuracy (Berg, Biele and Gigerenzer, 2017). The purpose of the forgoing model with strict neoclassical assumptions is to demonstrate (i) rational ignoring of experts in a strategic setting and (ii) the potential costliness of bad communication equilibria. The link to new paternalism based on behavioral economics is, I think, straightforward in this context (Rizzo and Whitman, 2009; Kapeliushnikov, 2015). When experts cite behavioral economics as a rationale for influencing non-expert behavior and give themselves license to nudge (Thaler and Sunstein, 2008), the dimensionality of the strategic communication increases and the likelihood of failing to coordinate on a good communication equilibrium (informationally efficient) increases substantially.

3.3 Identifying good experts as contractarian dissemination agents

The non-expert reasons about the following questions in the simple model presented below. Does the expert seek to help me maximize my objective function? Or does the expert seek to influence me to maximize her view what my objective function should be? I will refer to a *good expert* (or contractarian expert in line with Sugden’s (2018) adaptation of Gauthie’s

terminology) as one whose objective function is identical to the non-expert’s but with the informational advantage of knowing the non-expert’s objective function more precisely than the non-expert knows his own objective function. Alternatively a bad expert is one who seeks to influence the non-expert to maximize an objective that is different from the non-expert’s own objective. 7

Note that, by this reductionist binary definition I have put forward, the expert may be bad even when she paternalistically influences the non-expert based on benevolent, sincere and even loving intentions to make the non-expert better off in her view.

A caveat: Although I shall continue focusing on how valuable information can be accurately disseminated by experts, sometimes what we want from experts is prescriptive opinion and expert judgement. In those cases, we may regard an expert’s recommended action itself as a piece of private information—transmitted by making it public, and then weighted, filtered or interpreted by non-experts.

7In a model where the communication action sets do involve framings or nudges, I think my (overly) simple binary taxonomy of expert types could be refined to reflect suggestions of McKenzie and Sher (2018) about how nudges could be carried out in ways that ”may lessen skepticism about covert manipulation” (p20) and better preserve the decision maker’s (referring her to the non-expert’s) autonomy. By transparently informing the non-expert that a nudge is being employed on the theory that the non-expert’s construction of his otherwise undefined or incomplete preferences in rational, McKenzie and Sher observe the possibility that nudging (e.g., in cases where there exists no neutral framing or description of the menu) can ”respect her dignity as an agent and thus, in the long term, to preserve her trust” (p20) (e.g. by informing whether defaults are expert recommendations or descriptions of modal behavior; and informing non-experts if items included in the framing are objectively representative or non-representative of most decision problems in the class of decisions the non-expert faces). Insofar as nudges that strategically communicate context can be re-cast as transmitting information or sending a signal, the ”general idea is to empower, rather than steer, DMs. Given that uncertainty seems to be a common factor in many of the phenomena reviewed here, uncertainty reduction may be a simple and effective means of improving decision behavior” (McKenzie and Sher, 2018, p23). But if we recast nudges into strategic information transmission, then all the challenges identified in the model presented below stand. Good-faith information transmission where covert nudges are made overt along the lines that McKenzie and Sher suggest may in fact help reduce the likelihood of a bad communication equilibrium in which information transmission is completely lost. But the problems of coordinating on a good, informationally efficient communication equilibrium are formidable and may become more severe simply by the fact of the expert’s communication choice space becoming complex once a given communication culture gives itself license to engage in paternalistic nudges.
3.4 Tough love and bad communication equilibria

Consider the problem a student faces of wanting to know how many hours he should study to gain entry to an elite university or to get a job offer at a particular firm or in a particular industry. Expert advisors, teachers and parents frequently exaggerate their recommended intensity of study and therefore deliberately provide distorted information based on the benevolent motive of instilling greater discipline, or on the theory that the student tends to put too little weight on expert advice. In Amy Chua’s (2011) autobiographical account of conversations with her daughter in *Battle Hymn of the Tiger Mother*, Chua recounts her frustration at being stuck in a bad communication outcome as an advice-giving parent trying to inform her daughter about how best to prepare for university entrance exams. Chua’s sentiments will be familiar to many parents and advisors who have endured the frustration of seeing advisees ignore valuable information. In frustration over having had previous messages about study behavior ignored by her daughter, Chua chooses to strategically transmit obviously exaggerated admonitions hoping that they would help undo some of the biased down-weighting of advice used as an information filter by her daughter: “You must study at least five hours every night after coming home from school; otherwise, there is zero chance you will be admitted to (the) university (you want to attend)!” There is a similarly exaggerated maxim well-known among Korean parents and teachers that is often dispensed to high school students as they prepare for university entrance exams: “If you sleep less than four hours, you will pass the college entrance examination; if you sleep more than four hours, then you will fail.”

Could we not ascribe some rationality even to young students who rationally down-weight or altogether ignore deliberate well-intentioned distortions of information? Might we find better ways to instill discipline in our students without deliberately distorting information we provide?
3.5 Private versus government speech

Although the example of information transmission from academic advisors to their students is private communication, I assert that similar mechanisms can take place in communication games with governments in the role of expert. Although the mechanism I describe is common to communication games between private parties or between private parties and government (in the role of informationally advantaged expert), the risk that covert nudges followed by rational down-weighting, filtering, or outright ignoring—leading to multiple negative externalities—are potentially more damaging when government is the expert. I return to this comparative analysis of greater harms in communication games with governments after describing the mechanism I have in mind that is common to private and government communication games alike.

Insofar as the rise of strategic nudges and more complex communication strategies lead people toward skepticism, greater uncertainty about which messages are worth listening to (and how to interpret them), all else equal, implies less likelihood of achieving informationally efficient coordination on a good communication equilibrium. I characterize a good communication equilibrium as one in which valuable information is transmitted accurately or truthfully from expert to non-expert, believed by the non-expert, and then acted upon. I claim that nudges make coordination more challenging.

The fragility of coordination that achieves efficient information transmission becomes apparent once one appreciates how easily disruption of information flows can occur (in environments with substantially asymmetric distributions of information) even in models where agents are endowed with superhuman (i.e. unrealistic) completeness and consistency of preferences and cognitive resources to optimize. If nudges threaten good communication equilibrium even under heroic rationality assumptions, then it would seem that nudges could undermine information flows even more easily in extended versions of the model with more
human attributes.

3.6 Model

Borrowing from Berg and Kim’s (2017) model, consider the following three-stage strategic communication (i.e., signalling) game with two players: an expert $E$ and non-expert $N$. Following Berg and Kim, suppose that $N$ has a decision to make, $x$, representing a real-valued quantity decision such as how much to study, which career path to take, how intensely to restrict cholesterol, how much to exercise, how much sun exposure to accumulate, how much income to allocate to a retirement savings account, what quantity of vegetables to eat, expenditures on protecting one’s private data online, etc. For simplicity, suppose that the objective function of $N$ contains a random variable unknown to $N$ but known to $E$. This expert knowledge should be interpreted as an estimate of the objective returns on the costly investment that $N$ must decide on measured in payoff units that $N$ accepts as the relevant metric of performance for this particular decision or domain (class of decision problems). If $N$ could be sure that he has received an undistorted copy of $E$’s expert information, then $N$ could immediately and effortlessly solve his constrained optimization problem and choose $x$ optimally. If the interests of $N$ and $E$ differ even slightly, however, then $N$ can expect that $E$’s message will contain some distortion because of $E$’s strategic motive to paternalistically influence $N$’s choice of $x$. But if their interests coincide exactly and this fact of perfectly aligned motives is known to $N$, then the message can be both truthful (i.e., undistorted by $E$) and believed (i.e., undiscounted by $N$). In the model endowed with standard neoclassical assumptions of complete and well-ordered preferences, my definition of good expert can be interpreted as a special case of Sugden’s (2018) analysis of opportunity (with fewer assumptions about completeness or consistency of preference) as following from Gauthie-inspired contractarianism as follows. $N$ would like to believe the message that $E$
transmits; and he will believe her if he believes that \( E \) cares solely about \( N \)'s objective in other words, if \( N \) believes that \( E \)'s preference coincides exactly with his own. In my setup,

From the seminal work of Crawford and Sobel (1982), it is well-known in the so-called cheap talk literature that truthful messages can be transmitted as an equilibrium outcome only if the objective functions of the sender and the receiver coincide and this fact is common knowledge.\(^8\) The Berg and Kim (2017) model is an extension of their paper in which it is assumed that the receiver does not know the preference of the sender\(^9\) while the preference of the sender is known to the receiver in Crawford and Sobel (2012). In addition to a costless signal in cheap talk models such as Crawford and Sobel’s, a sender can choose to use a costly signal, which Berg and Kim (2017) refer to as the \( E \)'s sacrifice—so that \( E \)'s true preference can be inferred by \( N \). Others have included both costly and costless signals (Austen-Smith and Banks, 2000; Kartik, 2007). The main difference is that both signals in their models directly influence \( N \)'s payoff function whereas, in the model in the present paper, \( N \) derives no direct benefit from \( E \)'s costly sacrifice and is influenced by it only in updating his belief about \( E \)'s type (good versus bad expert).

\( N \) chooses action \( x \in [0, \infty) \). Payoff functions of both \( N \) and \( E \) depend on \( N \)'s choice of \( x \) and a parameter \( \theta \) which is known only to the \( E \). Because \( E \) knows \( \theta \) and \( N \) does not, \( \theta \) represents \( E \)'s private expert information. \( E \)'s choice variable \( y \in [0, \infty) \) represents her intensity of sacrifice.

Assume that \( \theta \) is uniformly distributed over \( \Theta \equiv [0, 1] \). Payoff functions of \( N \) and \( E \) are

\(^8\) Battaglini (2002) demonstrates that cheap talk by two senders can fully reveal two-dimensional information even if the conflict of interest is arbitrarily large. Chakraborty and Harbaugh (2010) also consider two-dimensional cheap talk (by one sender) and show that such comparative cheap talk can be credible even if the conflict of interests is large because there of trade-offs whereby the sender gains in some dimensions but loses in others. Although this possibility

\(^9\) Li (2003) also assumes uncertainty on the part of the receiver about the sender’s bias. He compares outcomes when the bias of the sender is disclosed and when it is not disclosed instead of considering the option that \( E \) can send a costly signal in the first round before sending a cheap-talk signal in the second round.
denoted by $U_N$ and $U_E$, respectively, and take on the following functional forms:

$$U_N(x, \theta) = -(x - \theta)^2,$$  \hspace{1cm} (1)

$$U_E(x, y, \theta; b) = -(x - \theta - b)^2 - c(y; b),$$ \hspace{1cm} (2)

where the $E$’s cost-of-sacrifice function $c(y; b)$ is increasing in $y$ ($c'(y) \geq 0$), unbounded ($\lim_{y \to \infty} c(y; b) = \infty$) for all $b$, and increasing in $b$: $\partial c(y; b)/\partial b > 0$ for all $y$ and $b$, and $c(0; b) = 0$ for all $b$. Note that the parameter $b$ (from Crawford and Sobel, 1982) measures distance in some sense between the two players’ preferences. The key assumption is that, for $N$, $b$ is a second unknown parameter in this game and, for $E$, is a second kind of private information (i.e., she knows her own type but $N$ does not). I denote $E$’s gross payoff before netting out the costs of sacrifice as $V_E(x, \theta, b) = -(x - \theta - b)^2$, which can be interpreted as $E$’s intrinsic utility of $E$.

By the quadratic construction of these payoff functions, both $N$ and $E$ have bliss points that identify their respective views about how $N$ should choose $x$. Their preferences can be said to differ by $b$. If $b = 0$, then $N$ and $E$’s interests coincide; $E$ is a perfect agent that $N$ could employ to inform him about $\theta$ (with message $m$) and guide choice of $x$ to achieve $N$’s view of his optimal choice. In this case, $E$ cares about $N$ as if they were the same person regarding how $N$ is to choose $x$ (even if $E$ would make decisions differently for her own version of the $x$ variable, which is not in the model).

If $b > 0$, then $E$ wants to influence $N$ to choose a greater value of $x$ than $N$ himself views as ideal (under the counterfactual that $\theta$ would be known to both players.) I assume that $b$ is also private information observable only to $E$ and that, for simplicity, it takes on one of two values: $b_I$ or $b_D$ (where subscripts $I$ and $D$ denote gaps corresponding to “identical”
(\(b_I = 0\)) or “different” preferences). It is assumed throughout that \(b_I = 0\) and \(b_D > \frac{1}{4}\),\(^{10}\) (i.e., that \(b_D\) is not too close to \(b_I\)). \(N\)’s prior belief about the probability that \(b = 0\) is denoted by the non-degenerate probability \(\lambda \in (0, 1)\).

The game proceeds as follows. \(E\) moves first by choosing \(y\) (trying to influence \(N\) to update his belief upward from \(\lambda\) toward 1, but which otherwise has no direct effect on \(N\)’s payoff). Although \(U_E\) is strictly decreasing in \(y\) holding the \(N\)’s later choice of \(x\) constant, \(E\) is forward looking and chooses \(y\) considering its potentially positive effect as a signal by increasing the chance that \(N\) chooses values of \(x\) that \(E\) prefers. Therefore, there is a meaningful trade-off that \(E\) faces between the direct cost of sacrifice \(y\) and its beneficial effect (from \(E\)’s perspective) on the probability distribution describing the \(N\)’s subsequent choice of \(x\).

After observing \(y\), \(N\) updates his posterior belief, denoted \(\hat{\lambda}\), that \(E\)’s objective coincides with his. Next, the value of \(\theta\) is realized to \(E\) (but not to \(N\)), and \(E\) chooses a cheap talk message \(m\) (from a message space \(M \equiv \Theta\)) that is (by definition of cheap talk and the specification of \(U_E\) and \(U_N\)) payoff-irrelevant. Based on this message, \(N\) updates his belief about \(\theta\) and finally chooses action \(x\), which is directly payoff-relevant to both \(N\) and \(E\).

This model introduces uncertainty about \(b\) (the gap between the sender’s and receiver’s preferences) and action \(y\) (which is costly to the sender) before the cheap talk game of Crawford and Sobel (1982) begins. If \(b\) were common knowledge, then the game would become identical to Crawford and Sobel’s model, which is therefore nested as a special case. Because the game has a sequential structure (as shown in Figure 1), the analysis of equilibrium proceeds, as usual, by backward induction. The natural solution concept is weak Perfect Bayesian Equilibrium (wPBE).

---

\(^{10}\)This inequality is well-known as a condition for a cheap talk equilibrium to have only one equilibrium message (Crawford and Sobel, 1982). Without a sufficiently large separation between the \(E\) and \(N\)’s views of ideal \(x\), the analysis becomes more complicated without adding much in the way of new insight.
First, consider the cheap talk stage after $E$ has learned the value of $\theta$. It is well known that cheap talk games have multiple equilibria including a so-called babbling equilibrium in which the sender randomizes over all possible messages. I focus on the most informative equilibrium and denote equilibrium decisions with $(\cdot)^*$ superscripts. If it is common knowledge that $b = 0$, then there is no conflict of interest and it is well known that the fully revealing equilibrium exists (in which $m^* = \theta$ and $x^* = \theta$) among a continuum of equilibria. In this special case of perfectly aligned interests, $E$ always sends the truthful message $m^* = \theta$ and $N$ believes whatever $E$ says, choosing $x^* = m^*(\theta) = \theta$. In other words, $E$ can induce her preferred outcome, which reflects her view of $N$’s ideal behavior, by transmitting this view truthfully (i.e., without exaggeration or distortion).

If it is known that $b > \frac{1}{4}$, then there exists no meaningfully communicative equilibrium. Only the babbling equilibrium exists in which no message conveys meaningful information. In the uninformative babbling equilibrium, $N$ ignores any message from $E$ and chooses $x^* = \arg\min_x E[x - \theta] = \frac{1}{2}$. $E$ and $N$ are simply too far apart in their views of $N$’s ideal behavior for meaningful communication to take place.

Now, consider $E$’s choice of $y$ when the value of $b$ is unknown to $N$. Interest focuses on a separating equilibrium in which an expert of type $b_I$ chooses $y^*(I)$ and an expert of type $b_D$ chooses $y^*(D)$, where $y^*(I) \neq y^*(D)$. In this separating equilibrium, the $b_D$-type expert’s choice of $y^*(D)$ fully reveals $E$’s type (i.e., the value of $b$), just as the $b_I$-type expert’s choice of $y^*(I)$ does. The difference is that the $b_D$-type expert’s choice, $y^*(D)$, induces a babbling equilibrium in which $N$ chooses $x^* = \frac{1}{2}$ regardless of the expert’s message $m$, while the $b_I$-type expert’s choice, $y^*(I)$, induces a fully communicative equilibrium in which the true value of $\theta$ is fully revealed by the expert’s message $m$.

In this separating equilibrium, the expected intrinsic utility of an $I$-type expert is $V_{E}(\theta, b_I) = -(\theta - \theta)^2 = 0$. If $E$ is an $D$-type, then her expected intrinsic utility is $E[V_{E}(\frac{1}{2}, b_D)] = \ldots$
\[- \frac{1}{2} \int_0^1 \left( \frac{1}{2} - \theta - b_D \right)^2 d\theta = - \left( b_D^2 + \frac{1}{12} \right). \]

How can N infer from observing a particular action by E that she is a good contractarian expert (i.e., an I-type rather than D-type)? First, N understands that the D-type expert will always choose \( y^*(D) = 0 \). To guarantee that N can make inferences about the expert’s type conditional on observing \( y \), the equilibrium sacrifice strategy of an I-type expert, \( y^*(I) \), must satisfy two conditions:

\[ V_E(\theta, b_I) - c(y^*(I), b_I) \geq E[V_E\left( \frac{1}{2}, b_I \right)], \]  

(3)

\[ E[V_E\left( \frac{1}{2}, b_D \right)] \geq E[V_E(m^*(\theta; b_D), b_D)] - c(y^*(I), b_D). \]  

(4)

In other words, the I-type expert’s sacrifice \( y^*(I) \) must be: (i) small enough that it is not too costly for an I-type expert to incur this cost, and (ii) large enough that an D-type expert has no incentive to imitate an I-type. Details of the derivation and proof of the following proposition can be found in Berg and Kim (2017).

**Proposition 1** Assume \( b_D > \frac{1}{4} \). If \( c(y, b) = (b + \epsilon)y^n \) where \( n > 0 \) for some small \( \epsilon > 0 \), there exists a separating equilibrium in which (i) an I-type expert chooses \( y^*(I) > 0 \) and an D-type expert chooses \( y^*(D) = 0 \), and (ii) if \( y = y^*(I) \), then \( \hat{\lambda}(y) = 1 \) is formed and full revelation follows (\( m^* = \theta \) and \( x^* = \theta \)), and if \( y \neq y^*(I) \), then \( \hat{\lambda}(y) = 0 \) and a babbling equilibrium follows (any \( m^* \) and \( x^* = \frac{1}{2} \)).

It is not the case that Proposition 1 holds for any increasing cost function satisfying the assumptions made in Section 2. The main reason is that for a large value of \( y^*(I) \) satisfying D-type’s incentive compatibility, i.e., \( \Delta(D) < c(y^*(I), b_D) \), the possibility cannot be ruled out that I-type’s incentive compatibility condition is violated, i.e., \( c(y^*(I), b_I) > \Delta(I) \). The \(^{11}\)If \( y^*(D) > 0 \), then the D-type expert can always increase her payoffs by deviating to \( y = 0 \), regardless of N’s posterior belief.
cost function form given by \( c(y; b) = (b + \epsilon)y^n \) prevents this possibility by maintaining \( I \)-type’s signaling cost of a large sacrifice very low, i.e., \( c(y^*(I), b_I = 0) \approx 0 \).

It is worth emphasizing that the expert’s sacrifice need not bring any direct benefit to the non-expert for effective signaling to occur, because \( y \) does not directly enter \( N \)’s utility function. The sacrifice might simply (and perhaps absurdly) consist of burning money or giving money to others (not to \( N \)), because it reduces the utility of the expert without increasing the utility of \( N \). Nonetheless, the cost of the sacrifice must differ across types. Therefore, one could hardly expect such an activity as money burning does nevertheless serve to help gain \( N \)’s trust. What is important in this model is not the amount of money forgone but the difference between the affordable sacrifice and the actual sacrifice made. This difference is the key determinant of how effective the signal is in our model.

4 The right level of skepticism toward government speech?

In *Skepticism and Freedom*, Epstein (2003) points to social benefits from skepticism while also rejecting relativistic arguments that all ideas deserve equal consideration. Juxtaposed against this subtle interpretation that accounts for multiple kinds of skepticism and different, sometimes non-monotonic, payoffs from skepticism, I wanted to engage in a brief though experiment. Instances of skepticism are sometimes non-comparable and should not be elided into a single dimension. That said, suppose we hypothesize a continuum of degrees of skepticism, denoted \( s \), that takes on values in the unit interval ranging from minimal to maximal skepticism toward government. Skepticism about the integrity of judicial action would surely be damaging from a rule-of-law perspective. Therefore maximal skepticism, \( s = 1 \), cannot be optimal (social-welfare maximizing). And if the skepticism among the governed provides a natural check that limits abuses of power and perhaps generates competitive pressure se-
lecting for good governance (e.g., Tiebout competition among local or state governments), then minimal skepticism, \( s = 0 \), also cannot be optimal. If one were to plot social welfare as a function of \( s \), what would it look like? Would it be a smooth function with local minima at \( s = 0, 1 \) and an interior optimum at some \( s^* \in (0, 1) \)?

If governments employ more strategic reasoning about the way they communicate (e.g., nudging) and grant themselves the option of lying (Junker, 2011; Sunstein and Vermeule, 2008), then surely an optimized \( s^* \) should respond by increasing. Skepticism is of course costly although it generates benefits, too. If we focus on the value of a judiciary whose interpretations of the law are widely accepted without skepticism about their integrity, then it becomes clear that nudging policies which increase \( s^* \) overall as a rational response is riskier and costlier than its proponents realize.

5 Ecologies of heterogeneous beliefs and behaviors

A heterogeneous ecology of conflicting beliefs and behaviors (inter- and intra-personally) enables discovery of best practices where best can be directly linked to a clearly defined performance metric: Who is achieving long lifespans? Who is doing well financially? Who is happy (measured by an intuition of psychological balance, the number of friends, counts of instances of anxiety or depression, or any other apropos metric for an observer in a particular context)? Nudging explicitly seeks to reduce heterogeneity.

Unanimously held views on nutrition, medical practice, and monetary policy (to name only a few important domains where expert opinion influences government policy) are overturned with some regularity (perhaps with increasing frequency?). This overturning of orthodoxy can be viewed generally as a positive symptom of a healthy competitive marketplace for ideas. We no doubt benefit from transmission and utilization of expert opinion and
recommendations and will benefit from these being debated openly and vigorously without endorsement by government wherever government can refrain from causing informational lock-in by endorsing a static set of norms as orthodoxy. By doing less in the way of actively influencing non-experts to follow the static orthodoxy at one point in time (i.e., by nudging), the government can do far more good in ensuring open debate and heterogeneous adherence to orthodoxy. Although some deviations from orthodox expert recommendation are indeed costly and harm from those deviations could—at one point in time—possibly be reduced by nudging, it would seem that the potential negatives of nudging for future information flows speak against nudging as a general approach to public policy from a benefit-cost perspective. And the frequent cases in which orthodox recommendations turn out to be wrong or incomplete should, once again, give pause to those advocating a nudge policy program.

6 Thaler’s fortuitous inconsistency and unfortunate consistency

There was a major shift—i.e., inconsistency—from 1991 (Thaler’s *Quasi Rational Economics*) to 2008 (Thaler and Sunstein’s *Nudge*) regarding the question of whether behavioral economics was descriptive or normative. In 1991, behavioral economics was purely descriptive with no policy implications. In 2008, behavioral economics was rich with “policy implications.”


12Thaler (1991, p. 138) asserts: “A demonstration that human choices often violate the axioms of rationality does not necessarily imply any criticism of the axioms of rational choice as a normative idea. Rather, the research is simply intended to show that for descriptive purposes, alternative models are sometimes necessary.”
presents a conundrum. [T]he key findings are positioned to be orthogonal, i.e. unrelated, to nearly every interesting policy debate on which the question of economic behavior may bear. In this anti-normative appraisal, behavioral economics, although it accepts a broader vision of how firms and consumers actually behave, does not actually challenge homo economicus as the proper ideal for assessing how legal and cultural institutions should be designed.”

In Thaler and Sunstein 2008 (and earlier), however, Thaler switches gears and tell us to use nudges to induce greater conformity with the orthodox rationality axioms or greater consistency with what the individual’s ‘higher self’ wants (according to Thaler or someone’s tacit normative evaluation)—without providing evidence that deviations from those consistency norms are causing any economic harm or that any economic benefits are to be gained by achieving relatively greater conformity with the axioms of internal logical consistency.

Unfortunately, the normative standard of rationality based on axiomatic internal logical consistency was the constant thread through Thaler (1991), Thaler and Sunstein (2008) and a large number of works cited in both volumes. The normative standard of perfect rationality is therefore nearly identical for both neoclassical and behavioral economists (Berg and Gigerenzer, 2010; Berg, 2014a). A methodological observation: Behavioral economics could instead pursue genuine methodological pluralism in a competitive marketplace of ideas. A heteroculture among practitioners (rather than de facto monoculture regarding normative standards of rationality) could perhaps have nurtured more intense competition for better economic theories (e.g., evaluated by well-defined metrics of prediction). Then the relevant normative questions can be more fully investigated: Is the primary normative target for nudgers to induce greater conformity with the consistency axioms? Or something else? What follows is an enumeration of risks from the New Paternalism approach to government policy and its potential harms.
7 Decentralization Mislaid... and the Rule of Law?

Epstein (1998) argues for decentralization and skepticism directed toward excessive concentration of power:

The twin impulses of skepticism and dogmatism, then, easily lead to the special pleading that is the most insistent enemy of a free society. The only way to buck that trend is to resort to a strategy that gives a very different twist to skepticism and dogmatism. The one sure dogma—that human knowledge and human plans are inevitably limited—leads to a principled skepticism about putting first our own personal and group interests. Accordingly, a sound legal order is one that responds to the fragility of knowledge by giving no one absolute control and power. It seeks the dispersion of power across individuals and social groups. Yet even this note of caution leaves much work to be done. The celebration of individual rights and the decentralization of social power does not explain how these twin objectives should be achieved.

Epstein (1998) argues that individual autonomy should hold sway over pleading, even by experts, over what others’ preferences should be, while suggesting both pragmatic and principled approaches to addressing external costs that may arise from the exercise of those preferences. New paternalists appear to argue that theoretical future or other selves should be regarded as bearing external costs generated by exercise of a short-sided, biased or irrational self in some class of decisions. The implication is that the unit of analysis of the individual is no longer rich enough to handle legal theorizing and social welfare analysis about nudging. It is this methodological breach that no longer identifies the individual as the natural unit of analysis and instead proposes bundles of selves with conflicting interests that threatens well-functioning rule of law (Epstein, 2011, 2014).
Nudging gives experts and governments new reasons to rationalize restricting individual freedom. It opens up numerous new potential mechanisms for rent-seeking and special interest politics. The discretion the nudging program gives governments to trump individual freedom and use law and regulation to influence behavior by mechanisms that are less straightforward than impartial judicial and regulatory review undermines the rule of law by bundling policies with a tacit threat to intensify efforts to induce desired choice behavior for those who deviate. I think we want a transactional law that clearly says: “You can choose $A$, $B$ or $C$ and here are here are the consequences as costs: $c_A, c_B$ or $c_C$, ($c_A < c_B < c_C$).” Instead, nudging says: “You can choose $A$, $B$ or $C$ and we believe you should choose $A$ (although for the time being, you could choose $B$ or $C$.” Rather than making government policy transactional and straightforward, nudging makes it discretionary and more uncertain. Effective communication requires strategic consideration. If non-experts doubt the motives of experts, then non-experts rationally discount or ignore the advice they receive. Valuable information is lost in such bad communication outcomes and, consequently, the quality of education suffers. In this paper, I argue that an expert’s sacrifice (or some alternative means of signaling non-paternalistic intention, or impartiality) is an essential virtue of a good expert for efficient communication between experts and a non-experts. Communication between experts and non-experts is modeled as a signaling game. A separating equilibrium is found in which a good expert (whose objective function truly coincides with $N$’s own objective function) makes a costly sacrifice that causes $N$ to believe what the expert says. In contrast, a non-good expert (whose objective does not coincide with $N$’s) chooses not to make the costly sacrifice and, consequently, $N$ rationally discounts or completely ignores the expert. The model demonstrates the importance of making non-experts aware of those aspects of the expert’s objectives that non-experts may not realize are closely aligned with their own (e.g., the extent to which an expert cares about non-experts as evaluated by non-experts
themselves). It also suggests that any deliberate distortion of information will not occur in equilibrium (if the expert is good) or, if any, necessarily damage pedagogical communication, no matter how well-intentioned (if the expert is not good).

References


Li, M., 2003, To Disclose or Not to Disclose: Cheap Talk with Uncertain Biases, University of Wisconsin at Madison


Popov, S., D. Bernhardt, 2013, University Competition, Grading Standards, and Grade Inflation, Economic Inquiry 51, 1764-1778


$E$ chooses costly sacrifice $y$

$N$ forms belief $\hat{\lambda}(y)$ about $E$’s type

$E$ learns state of the world $\theta$

$E$ sends cheap talk message $m$

$N$ updates belief about $\theta$

$N$ chooses effort $x$

Figure 1. The Sequence of Events