By neo-Calvinist point of view, I refer to analysis focused on assessing the diagnostic significance of policy decisions, rather than establishing the direct consequences of decisions. Diagnostic significance refers to information revealed by an irrevocable act, information about some underlying collective value or belief. Decisions, especially those that break with precedent, can expose the tradeoffs between competing values, tradeoffs that society tacitly endorses. In these situations, the role of analysis might be not so much to recommend a course of action, but to clarify what is at stake, the “diagnostic risk” as it were, created by the choice.

Here I outline a simple framework for how this analysis might proceed. The core ideas are taken from an earlier self-signaling model of choice (Bodner and Prelec, 2002; Prelec and Bodner, 2003). That model postulated a distinction between two types of utility: utility that flows directly from the causal consequences of choice, and diagnostic utility, which is the pleasure or pain derived from learning something positive or negative about one’s own internal state, disposition, ability, or future prospects. People are presumed to be chronically uncertain about where they stand with respect to these broad attributes, which in turn makes their choices diagnostic. Anticipation of diagnostic reward, or fear of diagnostic pain promotes self-control and inhibits self-indulgence.

The new aspect considered this paper is that the self-signaling agent is taken to represent a unified collective, such as a nation, society or corporation. With self-signaling agents,
individual or collective, standard decision analysis methods are no longer appropriate. The question then becomes how these methods might be usefully modified or adapted.

I start by looking at the motivational paradox inherent in the Calvinist doctrine of predestination, and argue that Calvinism brings into sharp relief an otherwise universal mechanism for self-control and self-management. We treat this mechanism — the mechanism that sustains motivation even in the absence of any causal link between actions and outcomes — as a purely naturalistic, psychological phenomenon. I then review the self-signaling model, and conclude with a speculative discussion of how decision analysis might be adapted to this new setting.

The Calvinist solution to the problem of self-control

As is well known, a distinctive aspect of the Calvinist religious system is the doctrine of predestination, which asserts that God has divided mankind into two categories, the Elect, who will enjoy eternal life, and the rest, who will not:

“By an eternal and immutable counsel, God has once and for all determined, both whom he would admit to salvation and whom he would condemn to destruction… This counsel, as far as concerns the elect, is founded on his gratuitous mercy, totally irrespective of human merit (Calvin, Institutes III, 21, 7).”

The die has been cast before birth and nothing a person does during his lifetime can change the outcome. Effort, good deeds, fulfilling duties, and so on, will not help. One might ask how this fatalistic doctrine could compel such stringent uniformity of conduct and observance. If one were designing doctrines from scratch, and wanted to create a set of beliefs to encourage virtue, it is not likely one would incorporate a belief in predestination, which takes all power over the final outcome away from the individual. On the face of it, the message seems to be: relax, it’s all been decided anyway. But relaxation is the last thing associated with Calvin and the system erected in Geneva, a “most perfect theocracy” where even details of dress and haircut were regulated.

From the standpoint of rational behavior, the motivational power of the doctrine is mysterious. What are the incentives for virtue, if virtue cannot accomplish the one goal that a believer values above all others? Max Weber proposed a solution to this puzzle over a hundred years ago, in his class work On the Protestant Ethic and the Spirit of Capitalism:
“… however useless good works might be as a means of attaining salvation, nevertheless they are indispensable as a sign of election. They are the technical means, not of purchasing salvation, but of getting rid of the fear of damnation.

In practice this means that God helps those who help themselves. Thus the Calvinist, as it is sometimes put, himself creates his own salvation, or, as would be more correct, the conviction of it. But this creation cannot, as in Catholicism, consist in a gradual accumulation of individual good works to one’s credit, but rather in a systematic self-control which at every chosen moment stands before the inexorable alternative, chosen or damned.”

Weber not only identified the source of motivation, but also of its tendency to impose order, regularity, and pattern on a series of actions:

.. [The Calvinist] could not hope to atone for hours of weakness or of thoughtlessness by increased good will at other times. The God of Calvinism demanded of his believers not single good works, but a life of good works combined into a unified system…

The moral conduct of the average man was thus deprived of its planless and unsystematic character and subjected to a consistent method for conduct as a whole. It is no accident that the name of Methodists stuck to the participants in the last great revival of Puritan ideas in the 18th century, just as the term Precisians, which has the same meaning, was applied to their spiritual ancestors in the 17th century.”

On Weber’s account, predestination, the very aspect of doctrine that should have eroded self-control, paradoxically served to enhance it. Radical uncertainty about underlying conditions endowed even minor actions with great significance. As actions became more systematic and regimented (Precision, Method, Purity), deviations become easier to spot. Thus order and planning gave rise to even more order and planning, in a self-reinforcing cycle.

A parallel line of thought appears in George Ainslie’s discussion of rule governed behavior and compulsion (Ainslie, 1992). Theoretical approaches to self-control within economics have emphasized the temporal aspect of the problem, the conflict between the near and the far. This conflict, however, does not capture a key — perhaps the key — aspect of the self-control decision, which is the problem of scale. Let us take a standard example — smoking. Monterosso and Ainslie state the problem concisely:
Consider a smoker who is trying to quit, but who craves a cigarette. Suppose that an angel whispers to her that, regardless of whether or not she smokes the desired cigarette, she is destined to smoke a pack a day from tomorrow on. Given this uncertainty, she would have no incentive to turn down the cigarette — the effort would seem pointless. What if the angel whispers instead that she is destined never to smoke again after today, regardless of her current choice? Here, too, there seem to be little incentive to turn down the cigarette—it would be harmless.

Only if future smoking is in doubt does a current abstention seem worth the effort. But the importance of her current choice cannot come from any physical consequences for future choices; hence the conclusion that it matters as precedent (Monterosso and Ainslie, 1999).

Again, we have a pre-existing condition (one is either destined to smoke, or destined not to smoke), which is not caused but only revealed by the current action. What keeps the smoker from smoking is the immediate loss in expectations about long run health that would be triggered by a single cigarette; the physical consequences of smoking one cigarette are of course negligible.²

We may use the term ‘self-signaling’ to refer to actions taken in order to provide good news about some personal characteristic or state, even when such actions do not cause that state. The basic assumption is that certain deep characteristics — which could be values, abilities, or beliefs — cannot be deduced through introspection, but can only be revealed through actual choice. Such characteristics are “cognitively inaccessible” (Bodner and Prelec, 2002). In this setting the choice becomes potentially diagnostic, which in turn creates a motive to self-signal the desired characteristics through a virtuous choice.³

**Self-signaling by unified collectives**

Societies and other corporate collectives also live in condition of uncertainty about their deep characteristics. Are they generous, are they just, are they destined to survive or to perish? Political rhetoric tends to strike characteristic Calvinist notes when there is a conflict between

---

² However, being forced to smoke the cigarette at gunpoint, or doing so in order to achieve some worthwhile purpose (e.g., to please your host), would not lead to the same loss in expectations. Only free actions are informative, on this account.

³ See Ginossar and Trope (1987), Mijovic-Prelec and Prelec (2010), Quattrone and Tversky (1984), and Sanitioso et al. (1990), for laboratory psychological evidence supporting self-signaling.
expediency and broader values or self-esteem. For example, scanning a few recent headlines: “How we care for orphans as a society shows what kind of society we are,”4 “Are We a Nation Doomed to be Violent?”5 “Is Canada a Country That Hates Its Young?”6 By bringing pre-existing conditions into the picture, the rhetoric amplifies the stakes on what may be a minor or narrow issue. Granted, orphans are important, but the decision is not just about them. Rather, lack of care for orphans is symptomatic of a deeper disorder. The rhetoric invokes Calvinist intuitions about pre-existing conditions, which are revealed rather than caused by the current decision.

It is plausible that the deep values that define a society are not fully accessible by collective introspection, whether informal or formalized through social science research. Political liberties, guarantees of due process, protection against threats and catastrophes, beliefs in national mission, uniqueness and destiny, are examples of such characteristics. No matter how much research is done, there remains a margin of uncertainty that is only resolved through an actual test. Indeed, one might suppose that a greater range of characteristics is inaccessible at the social level than at the individual one. Preferences and beliefs that are individually inaccessible surely remain so in the aggregate, and collective preferences and beliefs are distributed among the crowd. It is hard to tell how a collective would react in a new situation, faced with unexpected challenges and tests. Protections afforded by constitutional guarantees against imprisonment without trial or against torture, exist on paper, but will they hold up in exceptional times? Even if one could elicit information about deep characteristics, there is a matter of credibility. Official representatives, who are in position to see how decisions are made, also have strong incentives to disguise the real picture. Hence any direct expressions of core preferences or beliefs may be treated as indicators of the official ideology, not fact.

Analysis for a self-signaling decision maker

What might be the role of decision analysis in this setting, when underlying values and beliefs are inaccessible? Or, to put it differently, is there any way to exploit the objectivity and analytical rigor of decision analysis in situations where decisions have diagnostic rather than

5 http://indiancountrytodaymedianetwork.com/ict_sbc/are-we-a-nation-doomed-to-be-violent/
6 http://montrealsimon.blogspot.com/2011/05/is-canada-country-that-hates-its-young.html
causal significance?

It will be helpful to begin with a stylized account of decision analysis in the traditional mode, with a single decision maker representing a unified collective confronted with a choice between actions \( x, y, z, \) etc.. Prior to analysis, she is unclear about her values (preferences) and her beliefs (subjective probabilities over relevant events). Formally, one would say that the decision maker does not know her type, \( \theta, \) which is a catchall variable representing values, or beliefs, or both.\(^7\) Uncertainty about \( \theta \) translates into uncertainty about the actual utility function over actions, \( u(x, \theta), \) which the decision maker presumably wishes to maximize. If choosing without benefit of analysis, by gut feel, the decision maker would go with the action \( x \) that maximizes \( \sum_\theta p(\theta)u(x, \theta), \) where \( p(\theta) \) represents uncertainty about one’s type.

In the ideal scenario governing the traditional mode, the end result of decision analysis will be to clarify the decision maker’s actual type, \( \theta^*, \) which is to say, her beliefs, desires and abilities. This process may involve a combination of discussion and formal elicitation exercises, where the decision maker is presented with hypothetical choices, and reveals preferences and probabilities through these choices. Enlightened by the results of analysis, the decision maker will then choose to maximize the exact utility criterion \( u(x, \theta^*), \) instead of the imprecise criterion, \( \sum_\theta p(\theta)u(x, \theta). \)

Traditional decision does, therefore, acknowledge that the discovery of true values and beliefs is a complex process, and, indeed, that is the very reason why professional assistance is brought into play. But the difficulty in extracting underlying characteristics is purely cognitive, not motivational, and can be solved by entertaining hypothetical scenarios and choices.

For a self-signaling decision maker the underlying characteristics — preferences and beliefs — are radically inaccessible: they cannot be mapped out by any purely intellectual exercise. However, these characteristics may be revealed through actions. Exactly how they are revealed will depend on the underlying decision rule, mapping characteristics to actions. To move forward with analysis within the broad framework of utility maximization, one would need to specify the general form of the utility function that the self-signaling agent is trying to optimize.

\(^7\) In the case of belief, \( \theta \) would be probability vector \( \theta = (\theta_1, \ldots, \theta_n) \) over events \( S_1, \ldots, S_n. \)
We follow here the specification given in (Bodner and Prelec, 2002), which is arguably the simplest model of this phenomenon. As before, \( u(x, \theta') \) will denote the utility associated with action \( x \). This is the ‘deep desire’ for \( x \), cognitively inaccessible but nevertheless registering an impact on choice. As before, explicit beliefs about \( \theta \) are defined by a self-image distribution \( p(\theta) \). The value of this self-image is, in turn, determined by a separate utility function, \( V(\theta) \) which indicates how much utility would be gained by discovering the true \( \theta \). This function represents the only new element in the theoretical setup.

By intentionally choosing one outcome over others, the decision maker potentially learns something about the inaccessible \( \theta \). Hence, a deliberate action leads to an updating of the self-image, from \( p(\theta) \) to \( p(\theta \mid x) \). The updated self-image generates a second form of utility, called diagnostic utility: \( \sum_{\theta} p(\theta \mid x) V(\theta) \), which is computed by replacing the prior self-image \( p(\theta) \) with the posterior \( p(\theta \mid x) \). Diagnostic utility captures the extent to which choices provide good or bad news about \( \theta \):

**Total utility = Outcome utility + Diagnostic utility**

\[
U(x, \theta') = u(x, \theta') + \sum_{\theta} p(\theta \mid x) V(\theta)
\]

Whether an action provides information about \( \theta \) will depend on the inferential process that specifies how the updated self-image \( p(\theta \mid x) \) is computed from the choice and from prior beliefs \( p(\theta) \). There are two endogenous approaches to such inferences, i.e., approaches requiring no new parameters beyond the ones already given, viz. \( u(x, \theta), V(\theta) \)and \( p(\theta) \) (Prelec & Bodner 2003). A rational inference rule assumes full awareness about the self-signaling criterion. This means that \( p(\theta \mid x) \) must properly discount the signaling value of an ostensibly virtuous action for the fact that the action is partly motivated by anticipated favorable inferences. This carries to a logical conclusion the basic idea in self-perception theory (Bem 1972), namely, that the process

---

8 For related models of signaling between temporally differentiated multiple-selves, see Benabou and Tirole (2004), and Bernheim and Thomadsen (2005).

9 It would be more correct to write: \( u(x, \theta') + \sum_{\theta} (p(\theta \mid x) - p(\theta)) V(\theta) \), but to simplify, we omit the constant term \( \sum_{\theta} p(\theta) V(\theta) \) representing expected utility of the prior self-image.
of inferring hidden beliefs and desires from overt behavior is the same irrespective of whether
the inferences pertain to someone else or to ourselves. Just as we might discount someone else’s
good behavior as being due only to a desire to impress, so too we could discount our own
behavior for ulterior motives. The second, and perhaps more plausible inferential rule assumes
no awareness of diagnostic motivation, and hence all actions are treated as the result of simple
maximization of outcome-utility. There is no discounting for diagnostic motivation. In that
case, a good action is taken at face value, resulting in an overly optimistic self-image.
Descriptively, this may be the more realistic model of self-signaling. However, a normative
analysis, as pursued here, would require the rational inference rule.

To complete the model in either variant one imposes a consistency requirement between
choice of \( x \) as function of \( \theta \), and inference about \( \theta \) as a function of observed \( x \). This means that
the updated self-image \( p(\theta | x) \) places positive probability only on those characteristics \( \theta \) that
maximize utility in light of \( p(\theta | x) \) (total utility for the rational variant, or outcome utility for the
face-value variant). The formal requirements for equilibrium are that if action \( x \) is taken with
positive probability for some \( \theta^* \), then for all \( y \), \( U(x, \theta^*) \geq U(y, \theta^*) \). Conversely, \( p(\theta^* | x) > 0 \)
implies: for all \( y \), \( U(x, \theta^*) \geq U(y, \theta^*) \).

Here is how the model might interpret the smoker’s dilemma, as presented by
Monterosso and Ainslie. The deep characteristics in play are the smoker’s beliefs about future
smoking: Is he destined to smoke always or never? Let us suppose that the angel did indeed
whisper something, but the whisper is cognitively inaccessible. The hand that might or might not
reach for the cigarette has heard the message but the conscious mind has not. Furthermore, no
amount of assisted armchair analysis will reveal the contents of the message. What decision
analysis might do, however, is to establish how the utility parameters depend on the message.
To be concrete, let us suppose that these parameters indicate that the diagnostic utility of finding
out that one will never smoke is +2 units irrespective of beliefs about future smoking. This is the
value of good news. Let us also suppose that the desire to smoke this single cigarette increases as
function of deep belief about future smoking: the desire is +1 if the angel whispered: “you will

\[ 10 \] There remains the case of out-of-equilibrium beliefs. One possibility is to adapt the D1
criterion: If \( x \) fails to maximize \( U(x, \theta) \) for any \( \theta \), then \( p(\theta | x) > 0 \) only for characteristics \( \theta \) for
which the loss associated with \( x \) relative to any other action \( y \) (hence measured by
\( \text{Max}_y \{ U(y, \theta) - U(x, \theta) \} \)) is minimized (see Cho and Sobel, 1990).
never again smoke” and +3 if: “you will smoke forever.” A positive interaction between deep beliefs and current desire for a cigarette is psychologically plausible: Resignation that one will smoke forever is a likely incentive cue, triggering a craving to smoke now; alternatively, deep belief in future smoking and momentary desire might both be linked to a chronic appetite for smoking, and so on.

We have then a situation where the desire for the cigarette will fall short of diagnostic utility if the smoker deep down, truly believes that he will never again smoke, and will exceed diagnostic utility if he deep down, truly believes that he will smoke forever. Thus, the decision analyst can state with confidence that the smoker will smoke this cigarette if and only if deep down he believes that he will smoke a pack a day from tomorrow on. The cigarette taken now exposes an underlying lack of conviction about the long run. Moreover, conviction, or lack thereof, is fully warranted: the smoker will smoke this cigarette if and only if he is destined to smoke forever. This last statement is all that the analyst would convey to the smoker, leaving the final decision up to him.

This is a hugely stylized example, but it conveys the essence of an analysis conducted in a neo-Calvinist spirit. The starting position is a radical skepticism about our ability to introspect desires and beliefs, either as individuals or as organized collectives. Verbal affirmations, e.g., of intentions to stop smoking, cannot be trusted. The standard procedures of decision analysis are likewise of no help, because they also rest on purely hypothetical scenarios and choices. Since access to deep characteristics is barred, the main function of traditional policy analysis — to rationally arrive at an optimal course of action — cannot be carried out. Instead, the goal of analysis should be to clarify the diagnostic significance of each possible action and present that information to the decision maker.

The first step in this process would be to define the deep characteristics (θ) that are at diagnostic risk. Such characteristics would have to be (a) pre-existing, i.e., not created by the decision, (b) uncertain, (c) inaccessible by armchair analysis, and (d) potentially relevant, that is, exerting influence on the action taken. In the policy domain, prominent categories of characteristics would include collective values (expressed as tradeoffs between competing

---

11 The story could be fleshed out in other ways of course. One could hypothesize that the unknown type parameter refers to how much one cares about future health (rather than belief about future smoking); in that case, the decision to smoke the current cigarette would expose an underlying lack of concern about future health.
objectives), collective abilities (knowledge, competence, determination), and collective beliefs about future prospects (destiny, survival, national prestige).

To link these characteristics to the self-signaling model, one would express them in terms of outcome utilities, \( u(x, \theta) \). From the standpoint of the model, a characteristic is nothing more than the associated outcome utility function. The tools for assessing \( u(x, \theta) \) could be taken from the standard decision analytic arsenal — choices between lotteries and trade-off analysis — except that the questions would be thoroughly depersonalized. The decision maker would be asked to consider a hypothetical scenario in which society deep down holds values or beliefs \( \theta \), and then, acting as an agent on behalf of that society, express preferences between different \( x, y \) that would be most consistent with the specified values and beliefs. This would require separate elicitation exercises for each possible \( \theta \), a process that could be simplified by assuming parametric forms for \( u(x, \theta) \).

The second step would be to assess the diagnosticity of different possible actions, that is, to pin down the updated self-image distribution \( p(\theta | x) \) for each possible \( x \). One might think that this could be done by direct elicitation of \( p(\theta | x) \), as a straightforward probability inference conditional on action \( x \). The direct approach has two drawbacks, however. First, the probabilities may not be credible, because if the decision maker is inclined to choose some particular \( x^* \), that may lead him to rationalize the choice by providing an excessively favorable assessment of \( p(\theta | x^*) \). Second, the distribution \( p(\theta | x) \) should in principle be derived from the equilibrium condition, and it is unreasonable to expect a respondent to be able to compute an equilibrium.

The better approach, instead, would be to formally derive \( p(\theta | x) \) from \( V(\theta), p(\theta) \), and the already extracted \( u(x, \theta) \). One could attempt to elicit \( V(\theta) \) and \( p(\theta) \) by treating \( \theta \) as “certain knowledge that \( \theta = \theta^* \),” and presenting the decision maker with choices between lotteries on \( \theta \), and between lotteries and modifications to the status quo. The exercise would challenge the decision maker to make tradeoffs between levels of deep characteristics, and to judge whether certain levels are superior to the current state of affairs. These tradeoff choices would reveal the official, public values, and official current self-image, but, of course, not the actual underlying values and beliefs. The exercise would map out the official ideology of the unified collective.

The final step of the analysis is purely deductive, namely, to compute equilibria consistent with the extracted parameters, \( u(x, \theta), V(\theta) \) and \( p(\theta) \). The result would be both an action rule that specifies \( x \) as a function of \( \theta \), and a diagnostic inference rule that specifies how
the updated self-image $p(\theta|x)$ depends on $x$. These two rules would comprise the bulk of the report conveyed back to the decision maker.

Previously I have said that the goal of analysis should only be to clarify the diagnostic significance of each possible action, and not to select one particular action as best. An apparent exception to this arises whenever analysis yields a unique pooling action as the only equilibrium consistent with the extracted parameters. The exception is only apparent, because even a slight deviation from the pooling action would trigger a strong negative inference, on the usual refinement criteria (Cho and Sobel, 1990). The pooling equilibrium is essentially a rule, stating that something must be done or not done irrespective of deep desires and beliefs. Pooling would be expected in situations where diagnostic utility looms much larger than any outcome utilities at stake, for any type. In that case, the message sent to the decision maker would be simple indeed: Whatever temptations are in play, you should cut deliberations short and do the right thing; the diagnostic risk to self-image is just too high.

In a pooling equilibrium, the diagnostic value of actions is highly asymmetrical. Doing wrong, which is to say, deviating from the pooling action, leads to a severe loss in self-image; however, doing right, which is to say, sticking with the pooling action, leaves the self-image exactly where it was before. This fits well with the traditional interpretation of Calvinist inference: A life of virtue does not guarantee salvation, but a dissolute life is proof of damnation.

Summary

I have outlined here one way in which diagnostic utility could be introduced into policy analysis. Diagnostic considerations are a staple of political argument, as shown by rhetorical phrases such as “Are we a society that allows, tolerates, cares about X?” It is as if the underlying levels of commitment by society to stated values — the true collective preferences — are unknown. In these circumstances, a decision with even minor consequences, concerning due process, say, may be tremendously informative about a larger issue, such as whether the liberties guaranteed by the Constitution would survive under stress. A full cost-benefit for policy analysis should then take into account these diagnostic implications. None of this is intended to minimize the formidable difficulties in carrying out this project in an actual policy dilemma. Even if it is true, as I have implied, that everyone, including policymakers, is to some degree a genetic
Calvinist, it is far from evident that people in authority would accept the diagnostic interpretations as served up by a technical, outside agency.
References


