Economics & Philosophy (2026), 1–23
doi:10.1017/S0266267126100728

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Abstraction as flexibility: the veil of evaluative uncertainty

Carlo Ludovico Cordasco

Alliance Manchester Business School, University of Manchester, Manchester, UK
Email: carlo.cordasco@manchester.ac.uk

(Received 20 September 2025; revised 08 January 2026; accepted 30 January 2026)

Abstract

I argue that evaluative uncertainty gives rational agents instrumental reasons to abstract from some of their salient preferences when bargaining about social institutions. Because agents cannot assume stability in their future evaluative outlook, it is rational to favour rules that preserve options that may become salient. Building on Kreps (1979), I show how flexibility-driven abstraction expands the bargaining set, enabling convergence on rules while preserving motivational continuity. Since options are endogenous, bargainers also have reason to deliberate about option-generating and option-filtering meta-rules that structure the emergence, appraisal and revision of options over time.

Keywords: social contract theory; evaluative uncertainty; preference for flexibility

1. Introduction

Social contract theories face a dilemma. If one grounds justification in an impersonal standpoint that abstracts from morally arbitrary features – as Nagel (1986), Barry (1989) and Rawls (1999) do in different ways – or by presupposing that deliberators are already motivated to find terms others could not reasonably reject, as Scanlon (1998) does, one secures the determinacy and fairness of the agreement but fails to deliver solutions that actual citizens can see as binding once they return to their actual circumstances. If, by contrast, one begins from agents' actual preferences, beliefs and constraints (Gauthier 1986; Sugden 1998; Binmore 2005; Gaus 2011), one obtains motivational continuity but at the cost of tractability: convergence is uncertain, the nature of the agreement is indeterminate, and outcomes are liable to appear unfair since they reflect the contingencies of bargaining power.

A third strategy appeals to uncertainty. Brennan and Buchanan (1985) ask individuals to imagine themselves at a constitutional stage behind a "veil of uncertainty", where their future social position is unknown. The aim is not to secure impartial justice but to elicit agreement on broad institutional rules that remain attractive when the distribution of advantage is unpredictable. Limited abstraction

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

of this kind, grounded in long-run self-interest, can discipline short-term opportunism by inducing support for stable frameworks of cooperation. Yet once citizens form confident expectations about their standing, the prudential symmetry that supported agreement dissolves, incentives to uphold the constraints fade, and the contract risks entrenching existing distributions rather than constraining them (Kogelmann 2018; Schaefer 2021).

In this paper, I offer a prudential account to ground abstraction that appeals to uncertainty, but I relocate its focus. For Buchanan the uncertainty is positional: one does not know where one will stand in the future distribution of power and resources. In my account, uncertainty is evaluative and concerns which values and projects one will come to regard as central. On this view, abstraction is rational because agents cannot assume that their present evaluative outlook will remain stable.

Kreps's (1979) account of flexibility is useful here. When agents are unsure what they will value, it is rational to prefer arrangements that keep several possibilities open rather than lock them into a single course of action. Flexibility has value because it hedges against shifts in one's evaluative standpoint. Institutions can be assessed on the same basis: they are attractive insofar as they preserve option sets one may later have reason to prize and maintain the routes of uptake through which such options can be realized. Flexibility-driven abstraction, I will show, does not entirely eliminate worries about motivational continuity or power-sensitivity, but it helps create the conditions for convergence in the context of pluralism.

The flexible set should include those futures that are *recognizable* as one's own under higher-order values already present in one's evaluative outlook, and reachable through feasible routes of uptake such as learning, training, socialization or entry into practices. Foreclosing such trajectories imposes a prudential loss, since their value can be grasped in outline and there are identifiable paths by which one could come to occupy them. By contrast, futures that cannot be made intelligible within one's present evaluative standpoint, and for which no feasible route of uptake exists, fall outside the scope of flexibility. Prudence cannot give action-guiding reasons to pay the costs of preserving them.

The option set itself is also uncertain. Many paths that could matter tomorrow are not yet visible today. The rationale for flexibility-driven abstraction therefore goes beyond preserving access to known possibilities and extends to meta-institutions: option-generating and option-filtering rules.

The paper proceeds as follows. First, I illustrate the tradeoff faced by different approaches to social contract theory and Buchanan's appeal to positional uncertainty. Then I develop an alternative account that grounds abstraction in evaluative rather than positional uncertainty, drawing on preferences for flexibility. Finally, I specify the scope of the flexible set and address uncertainty about the option set, deriving implications for institutional design.

2. Social Contract Theory: A Tale of Two Cities

One influential family of social-contract theories locates justificatory force in radical abstraction. The parties are not citizens as we find them but idealized

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

representatives stripped of knowledge about their identities, social positions and comprehensive commitments. Rawls's original position is the canonical device: behind the veil of ignorance, contractors lack information about class, natural endowments and conceptions of the good, so appeals to morally arbitrary advantage are excluded and a fair position of equality is modelled (Rawls 1999).

Nagel (1986) and Barry (1989) pursue related strategies. Barry casts justification as argument from a standpoint of impartiality that rules out reasons grounded in contingency. Nagel describes the moral standpoint as a "view from nowhere", detached from any particular perspective. The formulations differ, but the underlying commitment is shared: justification proceeds by abstracting from whatever features of persons and circumstances would license partiality.

This strategy has two well-known attractions: first, it renders the choice problem tractable. Without informational constraints, as Rawls notes, no definite theory of justice could be established (1999); by restricting admissible reasons, the space of disagreement narrows and convergence becomes possible. Second, it promises fairness. Because no one can tailor principles to her own endowments or attachments, outcomes are insulated from morally arbitrary contingencies. Barry is explicit: permitting appeals to contingent advantage would simply legitimate arbitrariness (1989).

Yet the very features that deliver these attractions generate familiar difficulties. The first concerns hypothetical force. Dworkin captured it bluntly: "a hypothetical contract is not simply a pale form of an actual contract; it is no contract at all" (1973: 501). The fact that idealized parties would agree does not by itself give actual citizens reasons to comply. Rawls's recourse to reflective equilibrium and to a "political, not metaphysical" justification is designed to address this gap, but compliance still depends on whether citizens come to see the constructed principles as continuous with their own reasons.

A second problem concerns orientation: Sandel (1982) argues that excluding conceptions of the good, social roles and attachments leaves parties without the evaluative bearings required for deliberation. Either choice becomes indeterminate or it covertly relies on substantive values that the device officially excluded. Rawls tries to provide parties with a thin evaluative standpoint through the list of primary goods: things "any rational person" would want whatever her plan of life (1999). But this solution is unstable. If the list is too thin, it cannot generate determinate principles; if thickened, it builds in substantive values and undermines the neutrality of the construction. Sandel presses that Rawls's list already presupposes a liberal egalitarian conception of the person oriented toward maximizing transferable resources (1982).

A third difficulty arises from Rawls's concern with stability: he requires that the parties anticipate the strains of commitment and avoid principles citizens could not realistically follow (1999). Grant that this can work for commitments already legible at the time of agreement: parties can see, in broad outline, that certain requirements would predictably burden identifiable groups. The deeper difficulty is structural. The informational and motivational thinness that makes the original position tractable also limits what it can track. Some of the commitments most likely to generate instability do not yet exist at the point of agreement, or only become intelligible through lived experience. A citizen who later undergoes a transformative

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

change in outlook, a religious conversion, the reordering of attachments that follows becoming a parent, the slow crystallization of a vocation, may find the agreed principles alien in ways that no prior modeling of strains could have captured. The stability threats that matter most are not always visible from behind the veil; some emerge only after the veil is lifted and actual lives unfold.

Scanlon's contractualism represents a different route to the same impartialist ambition. Rather than abstracting away from one's standpoint, he presupposes a standing commitment to finding principles others could not reasonably reject. His device of reasonable rejectability works only on this basis: principles are permissible if, and only if, no one could reasonably reject them, given that each deliberator is motivated to find terms others could not refuse (1998). The structure is thus inverted. Rawls requires parties to detach from their particular standpoints in order to generate impartial principles, hoping that citizens will later reconnect to them. Scanlon leaves citizens in their standpoints, but addresses them as deliberators already committed to mutual justifiability. Each device secures impartiality by filtering out parochial advantage, but each faces a corresponding difficulty: Rawls struggles to transmit reasons back to actual citizens, while Scanlon narrows his justificatory reach to those who already share the moral commitment.

Later works tacitly concede these pressures. In *Political Liberalism*, Rawls (1993) shifts the justificatory target to an overlapping consensus among reasonable doctrines. Cohen (1989) and Habermas (1996) likewise ground legitimacy in fair deliberation among actual citizens. Gutmann and Thompson (1990) argue that justification must be accessible to citizens who continue to disagree about comprehensive views. Nagel (1991) and Waldron (1999) stress that pluralism places sharp limits on convergence and that democratic procedures may supply the only reliable ground of legitimacy. In each case, the pull is the same: as abstraction increases, determinacy and fairness become tractable but motivational grip weakens; as abstraction decreases, motivational grip strengthens but convergence and impartiality are jeopardized.

A second tradition rejects radical abstraction altogether. It begins from the Hobbesian and Humean thought that justice and morality are instruments for solving recurrent problems of cooperation among agents with divergent ends. Rules of justice, on Hume's account, arise "by a kind of convention ... by a sense of interest, suppos'd to be common to all" (Hume 1978: 490). The justificatory task is not to model ideal agreement but to explain how actual citizens, with their existing preferences and constraints, can endorse rules that fit with and remain stable under their motivations.

Gauthier's *Morals by Agreement* (1986) is perhaps the most influential contemporary statement of this view. He begins from a *presumption against morality*, treating moral rules as constraints on straightforward maximization (1986). His central claim is that unrestrained maximization is collectively self-defeating, so rational agents have reason to adopt dispositions of constrained maximization that stabilize cooperation. The justificatory aim is not impartial justice but a demonstration that rational self-interest itself can support cooperative moral dispositions.

Game-theoretical approaches sharpen this Humean line. Vanderschraaf (2019) treats justice as an equilibrium solution to recurring bargaining problems, robust

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

against renegotiation. Sugden models conventions as self-enforcing coordination devices that allow agents with diverse ends to secure mutually beneficial outcomes; in *The Community of Advantage* (2018), he reframes the principle of mutual benefit as a recommendation to citizens that it is their common interest to accept. Binmore (2005, 2007) combines bargaining theory with evolutionary anthropology to argue that fairness norms function as focal points in equilibrium selection: heuristics for resolving disputes over cooperation and division. The contract, on his view, is not a tribunal of impartial reason but a device for stabilizing coordination.[1]

This instrumentalist strand reaches its clearest expression in Gaus's justificatory liberalism. An "order of public reason" exists when each citizen has sufficient reason of her own to endorse the authority of social morality (2011). Legitimacy, on this view, requires not impartiality but convergence from within each standpoint. Moehler (2018) develops the point under conditions of deep pluralism: when no shared ideals exist, a purely instrumental contractarianism can still yield a minimal morality of peace and cooperation, thin enough to sustain order among liberal, non-liberal and even non-moral agents.

These views from everywhere have obvious strengths: they secure motivational continuity, since inputs are drawn from agents' actual reasons and the step from agreement to compliance is seamless. Gauthier's constrained maximizers comply because the disposition is rational;[2] Sugden's principle of mutual benefit presents itself as a mutual recommendation; Binmore's fairness norms act as focal points that each recognizes; Gaus's order of public reason is stable because every citizen sees it as supported by her own standpoint; Moehler's minimal morality works because even antagonists find cooperation rational.

But the liabilities are structural, and the most serious concerns convergence itself. In deeply pluralistic societies, the intersection of rules that every standpoint can endorse may be vanishingly small or empty altogether. Gaus himself concedes that in some contexts public justification fails (2011). This is not merely the familiar point that many games have multiple equilibria; it is that the bargaining set (the region of mutually acceptable rule packages) may not exist at all, or may be so thin as to be normatively uninteresting. When it does exist, secondary issues arise: indeterminacy, because conventions alone do not select among eligible outcomes, and power-sensitivity, because justificatory inputs mirror actual endowments and outside options, so those without bargaining power risk exclusion or domination

[1]For further analysis of path-dependence in public reason theories, see Thrasher (2019). On the relationship between contractarian models and the Rawls–Harsanyi framework, see Gaus and Thrasher (2015).

[2]Gauthier's claim is dispositional: it can be rational to adopt a disposition of constrained maximization because doing so makes one an eligible partner for mutually beneficial cooperation that would otherwise be unavailable. The argument presupposes a modest informational condition: others must be able, through observed behaviour, reputation or repeated dealings, to form sufficiently reliable expectations about one's compliance disposition. If agents were fully opaque and mimicking costless, the incentive would be to feign compliance while defecting whenever profitable, and the route from "rational to adopt" to "stable compliance" would break. The mechanism is selection via access to cooperative surplus, not enforcement via sanctions. Formal enforcement could in principle substitute for some of this work, but enforcement itself requires an informational base (defections must be detectable) and faces bootstrapping problems: building enforcement is a cooperative project that cannot simply assume the compliance it is meant to secure.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

(Hampton 1991; Morris 1991). Gauthier's Lockean proviso, Binmore's focal fairness norms and Sugden's conventions all aim to offset these effects, but in doing so they reintroduce normative assumptions the approach officially eschews.

To be clear: the objection is not to indeterminacy as such. Gaus can accept that many rule packages belong in the eligible set and let path-dependence or cultural evolution select among them. The problem is that under deep pluralism the eligible set may be empty or negligibly thin in the first place, and when it does exist, its location is shaped by outside options in ways that track bargaining power.

The contrast is plain. Views from nowhere secure determinacy and fairness at the cost of motivational force. Views from everywhere secure motivational continuity at the cost of convergence, determinacy and perceived fairness. This dilemma motivates a third approach: one that relocates abstraction from impartial morality to prudence under uncertainty.

2.1 Buchanan's Veil of Uncertainty

In *The Reason of Rules*, Brennan and Buchanan (1985) develop the clearest articulation of such instrumentally rational abstraction. They invite individuals to imagine themselves at a constitutional stage, behind a "veil of uncertainty" that conceals their future social position while leaving intact their general preference for advantage. The device is not intended to model a moral standpoint, as in Rawls, but to protect individuals against their own shortsightedness and the exploitation of short-run opportunities. Behind this veil, where one's future position in the post-constitutional order is concealed, individuals are induced to recognize the long-term interest of all members of the community in rules that constrain day-to-day politics. In choosing rules under such conditions of uncertainty, citizens set aside immediate advantage in order to secure the superior payoffs associated with predictable and stable frameworks of cooperation. The outcome is not a substantive principle of justice but an agreement on institutional arrangements that all will find it in their long-term interest to accept. The force of the argument is entirely prudential: rules are chosen because they reduce the costs of collective action and promise mutual gains across time.

This emphasis on long-term prudence reflects Buchanan's central distinction between the constitutional level, where rules are selected, and the post-constitutional level, where rules are applied (Buchanan and Tullock 1962; Buchanan 1975; Brennan and Buchanan 1985). At the latter stage, agents predictably pursue their immediate interests; at the former, uncertainty about their future standing induces them to adopt more general rules. The social contract is not an appeal to moral truth but a calculus of constraint: individuals submit to binding rules that restrict short-run opportunism in exchange for the higher returns secured by durable cooperation.

Buchanan's constitutional contractarianism has drawn sustained criticism. One common objection is that because constitutional change requires unanimity, existing arrangements are disproportionately privileged. The status quo becomes entrenched, not because it is just, but because it is already in place. The effect is procedural inertia, which may preserve institutions that are manifestly unfair (Rae 1975; Vanberg 2004; Munger and Vanberg 2023; but cf. Müller 1998). Rules may be

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

judged acceptable not because they are genuinely defensible, but because they fit with what people already acquiesce in. The result is that arrangements produced by domination or exclusion can appear legitimate simply by virtue of persistence. Constitutional contractarianism, on this reading, threatens to become less a justification of rules than a rationalization of whichever equilibria happen to endure.

A second critique targets the fragility of the motivational foundation: as agents form confident expectations about their positions, the justificatory force of the veil weakens and consensus unravels (Kogelmann 2018; Schaefer 2021). Moreover, critics argue that mutual-advantage contractarianism secures at best a *modus vivendi*, a fragile equilibrium that the worst-off lack reason to regard as binding (Rawls 1993; Barry 1996).

The force of these criticisms is not that Buchanan's project fails on its own terms. As a model of how prudential reasoning under uncertainty can discipline rent-seeking and short-term opportunism, it captures a genuine insight. But the limitations are structural. A justificatory device that rests entirely on long-run mutual advantage delivers stability only so long as uncertainty persists and existing inequalities do not overwhelm the bargaining set. Once either condition fails, the normative core dissolves into acceptance of the status quo.[3]

Buchanan's device can be read in two ways: as a claim about what actual constitutional choosers would do under genuine uncertainty, or as a hypothetical construction in which agents are modelled as uncertain regardless of their actual epistemic state (see Kogelmann 2018 on this ambiguity). On the first reading, the standard worry is immediate: as individuals form confident expectations about their likely position, the prudential symmetry that supported agreement weakens. On the second reading, the device may be less vulnerable to erosion, but it becomes unclear what justificatory work the stipulation is doing – why model agents as positionally uncertain when they are not, and why should actual citizens regard the output as binding?

3. Abstraction as Flexibility

I want to propose an alternative ground for prudential abstraction: uncertainty not about one's future position, but about one's future evaluative outlook. These two sources of uncertainty differ in resilience. Uncertainty about position is, in ordinary circumstances, highly sensitive to information: as knowledge accumulates about endowments, prospects and social location, the rationale for setting aside positional considerations becomes correspondingly harder to sustain. Evaluative uncertainty is different. Even an agent who knows where she stands cannot be confident which projects, attachments and higher-order priorities will become central to her

[3]Two further Buchanan texts sharpen the contrast. In "Positive Economics, Welfare Economics, and Political Economy", Buchanan notes that even Pareto-improving rules can be blocked by opportunistic agents who anticipate local gains from resisting constraints; this illustrates a structural vulnerability for mutual-advantage devices once strategic incentives are made explicit. In *The Limits of Liberty*, he sketches a more "flexible" contract in a different sense: constitutional terms are revisable as relative power and bargaining positions shift. That is flexibility driven by changes in position and power, whereas the flexibility defended here is procedural and prudentially grounded in evaluative uncertainty and the value of preserving routes of uptake across a life course.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

conception of the good, especially across predictable life transitions. This is not a demanding idealization but a familiar feature of human agency. Shifting the focus from positional to evaluative uncertainty therefore relocates prudential abstraction to a ground that tends to remain salient across the life course rather than fading as positional information is acquired.

The comparative claim does not require that evaluative uncertainty be universal. Some agents may be stubbornly confident that their current priorities will persist. What the argument requires is that evaluative uncertainty be sufficiently common, sufficiently persistent across predictable life transitions, and sufficiently action-guiding to serve as a stable basis for a contract device. The formal model developed below accommodates stubborn agents as boundary conditions: those with negligible probability of future change do not contribute to the prudential rationale for flexibility. But these cases are made visible as limits rather than silently assumed away.[4]

The underlying idea, then, is that *qua* individuals we cannot assume stability in our own evaluative outlooks. Life projects evolve, attachments transform, and what now appears peripheral may later become central to one's conception of the good. This gives agents prudential reasons to value institutional flexibility: by preserving a sufficiently wide range of options, institutions insure agents against the risk of precluding opportunities they may come to find salient. The shift also addresses a limitation noted earlier: because evaluative change is hard to forecast *ex ante*, agreement should not only screen for anticipated strains but also preserve routes of uptake against the possibility that what becomes central later was not legible at the point of choice.

Kreps (1979) offers a precise representation of this prudential logic. Rather than choosing a single option, an agent may prefer to choose a set of feasible options (a menu) from which a later choice will be made once her "tastes" are revealed. Menus have value under evaluative uncertainty because they protect against precluding options that one's future self may regard as essential. Importantly, this is not ordinary risk aversion over outcomes (Lipsey and Lancaster 1956), it is prudence directed at one's own evaluative trajectory: even if preferences are well defined today, they may shift as projects, relationships and information change. For this reason, option preservation carries value in its own right.

Suppose Betty must decide today where to dine tomorrow. At present she prefers Italian, but she recognizes that her tastes may shift: tomorrow she might instead want sushi. Faced with this uncertainty, she chooses a restaurant with a broad menu, even if this comes at the cost of slightly higher prices or less authentic dishes. The choice is rational not because she is risk-averse about food quality, but because she values the flexibility to adapt to her future evaluative outlook.

[4]Moral commitments are a natural source of "stubbornness" in the present sense (I thank John O'Neill for pressing on this point). They are often far more resistant to revision than ordinary preferences, and for some agents they may be treated as effectively fixed. This does not undermine the case for procedural flexibility. Even a highly confident agent typically lacks full certainty that her moral outlook will never shift, and she can therefore have reason to support meta-rules that permit peaceful contestation and lawful revision over time, not a requirement that she preserve for herself the option of acting against her current convictions.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Placed in a two agents setting, the logic of menu choice illuminates how flexibility-driven abstraction can support convergence. Suppose Alf and Betty have to book a restaurant for tomorrow. Alf currently prefers meat while Betty prefers fish. If each were confident their tastes would remain fixed, they would fail to coordinate: each would book a restaurant serving only their preferred dish. But if Alf and Betty acknowledge some probability that their preferences may change, each has reason to prefer a third restaurant that offers both meat and fish, even if there are some tradeoffs attached (e.g. lower quality or higher prices).

A simple model makes the structure of the prudential argument precise. Suppose the choice set is $X = \{m, f\}$, representing two incompatible courses of action. An agent faces two possible evaluative states. In state M she prefers m, in state F she prefers f. Let p be the probability of finding herself in state F when the moment of choice arrives.

If she commits now to $\{m\}$, her expected value is

$$V(\{m\}) = (1-p)\cdot 1 + p\cdot(1-\gamma),$$

where $\gamma > 0$ represents the cost of being locked into m should her future evaluative outlook shift to F. If she commits to $\{f\}$, the expected value is

$$V(\{f\}) = (1-p)\cdot(1-\delta) + p\cdot 1,$$

where $\delta > 0$ is the cost of being locked into f when she would later prefer m. If instead she chooses the menu $\{m, f\}$, she can defer commitment until her evaluative state is revealed. The expected value is then

$$V(\{m,f\}) = (1-p)\cdot 1 + p\cdot 1 - k = 1 - k,$$

where $k > 0$ captures the potential tradeoffs attached to the menu (e.g. lower quality, higher prices).

It follows that the menu strictly dominates $\{m\}$ whenever $p\gamma > k$, and it strictly dominates $\{f\}$ whenever $(1-p)\delta > k$. Together these yield a simple condition:

$$\frac{k}{\gamma} < p < 1 - \frac{k}{\delta}.$$

Essentially, when Alf and Betty exhibit *enough* uncertainty about their future tastes, it is rational for them to converge on a larger menu despite potential tradeoffs attached.

This way of grounding agreement differs from both Rawls and Buchanan: it does not posit a standpoint of impartial reason, nor does it rely on ignorance of one's social position. Instead, it proceeds from the recognition that one's evaluative standpoint is itself subject to change. Institutions are justified because they preserve option sets that agents prudently expect their future selves may value.

Consider disputes over land use. At first sight, NIMBY and pro-abundance positions appear irreconcilable. Preservationists treat restrictions on new construction as essential to safeguarding neighbourhood character, while advocates of abundance see such restrictions as the principal obstacle to affordability and growth. If each party deliberates only from the standpoint of its current salient preferences, the bargaining set is empty: the preservation of existing

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

built form is non-compossible with the large-scale expansion of housing supply. Yet once we shift the focus from present commitments to evaluative trajectories, convergence becomes possible. The preservationist recognizes that she may one day value affordability, whether for her children, for the viability of local services, or for her own prospects of downsizing. The abundance advocate recognizes that she may later prize heritage, green space or the quiet of a stable community. Precisely because neither can be certain which outlook she will come to endorse, it becomes prudentially rational to preserve both options.

What would such rules look like? The key is to build flexibility into procedures rather than trying to settle everything at once. Consider a regime in which upzoning is time-limited and geographically targeted: density increases staged over defined intervals, renewal conditioned on publicly specified metrics such as affordability or infrastructure capacity. Permissions, on this model, are provisional: sunset clauses and mandatory review dates build in the expectation that rules may be revised. This is not the fantasy of "undoing" a building once constructed; what remains reversible is the authority to build more. Future projects can be paused, scaled back or redirected even if past construction remains. The point is to keep both paths live, development and preservation, without requiring either side to concede defeat at the outset.

Other devices can redistribute burdens. Transferable development rights allow density to migrate toward willing locations while shielding protected areas; inclusionary requirements can be paired with heritage or green-space protections. None of this settles which urban form is best. What it does is build review, bounded commitment and clean exit conditions into the authorization process, allowing some experimentation while making it easier to change course. Each side can endorse these procedural rules, not because they agree on outcomes, but because the rules preserve paths into futures that may later matter to them, while limiting the downside of experimenting now.

Positions that look irreconcilable converge once uncertainty is acknowledged. Each party recognizes that her commitments may shift, and so each has prudential reason to support rules that preserve flexibility. The basis of agreement is not impartiality but prudence: the recognition that one's future self may embrace what one now rejects.

4. What Belongs in the Flexible Set?

Agents often have prudential reason to preserve options that their future selves may come to value, and institutions can be justified, in part, insofar as they sustain such option preservation under evaluative uncertainty. But the claim requires inclusion criteria. Which futures count as admissible members of the flexible set for a given agent at a given time, and which do not? If we answer too narrowly, we defeat the very purpose of prudential abstraction, which is to ensure the tractability of a social contract bargaining procedure among agents who exhibit diverse conceptions of the good. If we answer too broadly, we may model convergence on sets of rules that citizens will later fail to recognize as authoritative.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Callard's (2018) account of aspiration bears on this issue. On her view, some of our most important projects are undertaken not because we already value their objects, but because we seek to learn how to value them. Aspiration differs both from ordinary choice and from passive change of taste: it is an active process of rational self-transformation aimed at goods we only partially grasp. She describes this as being guided by "proleptic reasons": reasons borrowed from the value we anticipate coming to appreciate more fully in the future. For instance, a novice drawn to classical music may not yet experience the rich aesthetic satisfaction that seasoned listeners report. Still, she has reason to attend concerts, read about the tradition and practice attentive listening, because she expects that through such engagement she will come to appreciate the music in the way that experts do. In this sense, she can rationally commit herself to an evaluative outlook she does not yet possess, treating her future, more knowledgeable self as a source of reasons for action.

One of the many interesting implications that I find emerging from Callard's account is that the flexible set cannot be limited to the outlooks that line up with our current tastes. It must also include futures that are aspirationally reachable, goods that we can only learn to value by taking the first steps toward them, even before we see their full worth. But even this seems too restrictive. For example, five years ago I had no aspiration to learn more about game theory. It figured nowhere in my evaluative outlook, and nothing in my pattern of interests suggested that it might come to matter. Yet through a series of exposures, it has now become a central aspiration. I now regard understanding game-theoretic reasoning as indispensable, not only to my research but to how I conceive of practical rationality more broadly. The lesson, I believe, is that we cannot reduce the flexible set to current aspirations, but should allow for a wider range of possible futures.[5]

A second useful framework is provided by L.A. Paul's analysis of transformative experiences (2014). Paul argues that some decisions change not only what we value but also the very standpoint from which we deliberate, such that their subjective value cannot be known in advance. Her most prominent example is the decision to become a parent. Someone who has never had a child cannot prospectively grasp what the experience will be like, since it involves a transformation in preferences, phenomenology and identity. From the standpoint of expected-utility theory, such decisions are intractable: without a basis for assigning values to outcomes, there seems to be no rational way to compare parenthood with remaining childfree. Paul generalizes this point to other life-altering choices, such as undergoing a radical religious conversion or pursuing a vocation that reshapes one's sense of self. The challenge, then, is whether futures of this kind should figure in the flexible set.

On a strict reading, prudence provides no reason to preserve such future options, since they cannot be prospectively evaluated. This conclusion, however, is too strong. While their phenomenology cannot be grasped in advance, many such options can nonetheless be recognized under higher-order values such as love,

[5]Kogelmann (2025) develops a related point about how economic dynamism can undermine the conditions necessary for autonomous life planning, particularly what he calls the "Stability Condition," which requires that options remain available once pursued. His analysis reinforces the importance of institutional conditions that keep uptake routes open.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

devotion, excellence or service, that already figure in one's evaluative outlook. Where that recognizability condition is met, we may still exhibit instrumental reasons to preserve these future options.

It is not my goal here to provide a rich account of what those higher-order values might be. It is up to the individual to articulate which higher-order headings structure her outlook and to say, under those headings, which futures count as recognizable and which uptake routes are real. In Rawls's terms, these headings function as elements of a conception of the good: they order ends and guide a plan of life. I use higher-order descriptions only to fix recognizability for prudential purposes; I do not require that they be reasonable in Rawls's political sense or shared across persons. The standard is first-personal and indexical: what matters is that the agent can sincerely locate a future under the headings she now avows and that there exist uptake routes by which she could come to inhabit it. In this respect, higher-order values track Rawls's second moral power: the capacity to form, revise and rationally pursue a conception of the good. But they do so in order to ground a preference for flexibility rather than to supply a public basis of justification.

Taken together, these accounts motivate a criterion. A future belongs in an agent's flexible set at a time if it is *recognizable* as hers under a higher-order description already embedded in her evaluative outlook, and if there is a feasible *route of uptake* by which she could come to occupy it. The recognizability condition ensures that prudence is not inflated into an indiscriminate concern for alien lives. The uptake condition ensures that the set does not contain merely notional possibilities without accessible paths.

Recognizability should not be conflated with endorsement or phenomenological familiarity. To be clear about what recognizability is *not*: it is not "I can imagine enjoying it", and it is not "I would endorse it now". It is "I can situate it under a value-description that already governs my self-understanding". Parenthood, artistic excellence, or civic service may be intelligible in this way even if their lived experience remains opaque. Futures that cannot be placed within any of one's evaluative categories are excluded without prudential loss. This preserves the distinctiveness of prudence as concern for one's own futures, not for all possible lives.

The uptake condition prevents the flexible set from expanding indiscriminately. The value of preserving an option does not attach to every conceivable future, but only to those that remain in principle attainable if institutions do not foreclose them. Attainability depends on the existence of intelligible pathways of development through which one can come to inhabit the relevant outlook. When these pathways are not available, the flexible set is reduced. This reduction does not show that aspiration is incoherent, but rather that some possible lives are rendered practically inaccessible. Prudence therefore directs us not only to preserve futures that are recognizable in evaluative terms, but also to maintain the institutional and social conditions that make uptake feasible.

Imagine Betty recognizes "medical care" under the higher-order value of service. She can intelligibly see herself as the kind of person whose work consists of treating patients, and this recognition is stable. The option enters her flexible set only if there is uptake. In her case, the pathway requires preparatory study, an entrance examination, a period of full-time medical training, supervised practice and licensing. Given her present age and obligations, even the fastest route would extend

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

well beyond the horizon within which she could reasonably complete training. In this setting, recognizability holds, but uptake fails. The medical career drops out of Betty's flexible set not because her aspiration is incoherent with her higher-order values, but because the social institutions and temporal constraints make that future unavailable to her.

A natural objection arises: if uptake is tied to presently available pathways, the test simply reproduces the exclusions imposed by existing rules. Path dependence would then arbitrarily contract the flexible set. My reply is to index uptake to the routes available under the rule package we have prudential reason to endorse when we come to the collective bargaining table, not only to the routes available under the status quo. Call this prospective uptake. An option is admissible if there is a real path either under present rules or under rules that agents can now jointly adopt at reasonable cost and that they have independent prudential reason to endorse.

This last constraint is essential: it prevents the account from becoming vacuous. Without it, a sceptic could say that anything is uptake-feasible if we allow enough reform. The enabling reforms must themselves be jointly endorsable at reasonable cost, that is, they must pass a test analogous to the one applied to first-order options. It is not enough that a path could be imagined; there must be a feasible reform that we can authorize and maintain without imposing costs that other agents have reason to reject.

Under current regulations, Betty's medical route is closed. If, however, there is a credible mid-career bridge that can be instituted at modest cost and that many similarly situated agents would use, then we have reason to adopt that enabling rule. Conditional on its adoption, Betty's medical future satisfies uptake and can enter the flexible set. If the only way to open the path requires measures that fail the prudential balance, the option remains excluded.

A natural worry, however, is that even this more permissive picture remains too restrictive. Derek Parfit's view of personal identity (1984), for instance, appears to support a more expansive standard (see also Korsgaard 1989; Brink 1997 for similar approaches). On his account, a concern is owed to any future self with whom I stand in relations of psychological continuity and connectedness, irrespective of how foreign her evaluative outlook may be. A prudential concern, on this view, is decoupled from the particular contents of my present evaluative standpoint: it extends as far as the relations of memory, character or experience that stretch forward in time.[6] This position rightly highlights that prudence cannot be equated with the protection of current tastes. My reasons of self-concern must reach beyond the horizon of what I presently endorse.[7]

[6]Importantly, neither Parfit, nor Korsgaard, nor Brink frame concerns for future selves in a prudential fashion. I am using their theory of identity in such a way.

[7]See, for instance, Brink (1997): "If . . . the separateness or diversity of persons is not so fundamental . . . Insofar as distinct individuals are psychologically connected and continuous, each can and should view the other as one who extends her own interests in the same sort of way that her own future self extends her own interests. . . . This means that each should regard the good of those to whom she stands in such relationships as a constituent part of her overall good, just as she should regard the good of her own future self as a constituent part of her overall good. . . . Interpersonal connections and continuity can be found among intimates who interact on a regular basis and help shape each other's mental life; in such relationships, the experiences, beliefs, desires, ideals and actions of each depend in significant part upon those of the others" (Brink 1997: 141–143).

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Yet if this proposal is to be taken at face value, it risks making the flexible set implausibly overinclusive. Psychological continuity comes in degrees, and thin chains of connection can extend to futures in which my outlook is so transformed that I can no longer recognize it as mine in any meaningful sense. If the flexible set is defined via this route, it will expand until it encompasses outlooks utterly alien to me, ones I cannot even place within my conception of the good under a higher-order description. At that point, the distinction between prudence and altruism collapses: if I have reason to preserve futures so different from my own, then I seem equally bound to take account of the good of other persons. Prudential reasoning, which begins as a concern for my future self, dissolves into impartial morality, bringing back the worries about motivation.

Parfit's view can hardly be sustained in a prudential fashion. Instrumentally rational abstraction requires us to stick with the idea that futures must be recognizable as ours. This condition does not require that I already endorse their value, or that I can prospectively grasp their phenomenology. It requires only that I can locate them within the evaluative categories that currently structure my conception of the good. Parenthood or artistic excellence are all intelligible in this way, even if their lived experience is opaque to me now. By contrast, futures that fall entirely outside my evaluative vocabulary can be excluded without prudential loss. In this respect, the flexible set has to be narrower than in Parfit's account, but precisely because it acknowledges the inescapable role of our present "lenses" in thinking about our own futures.

The resulting view occupies a distinctive intermediate position between rational egoism and impartialism.[8] It is broader than egoism since it demands that I insure against futures that do not line up with my present desires but which I can still recognize as possible lives of mine. But it is narrower than impartialism, since it does not require that I weigh all possible interests equally, or treat every imaginable future as my own. The flexible set is thick enough to capture the genuine uncertainty of evaluative change, yet bounded enough to avoid collapsing into an impartial concern for all interests.

Beyond recognizability and uptake, there is the question of what it takes, here and now, to keep an option available. Sometimes the answer is "almost nothing". Sometimes it ties up resources such as money, attention or space. Sometimes one can keep the door open for an option only by pushing others aside. This is the *opportunity cost of preserving an option* (Arrow and Fisher 1974; Henry 1974; Dixit and Pindyck 1994). When that cost is light and non-displacing, inclusion is easy. When it is heavy or it requires sacrificing other options, inclusion requires stronger reasons.

I often cannot tell, at the outset, whether a path that looks meaningful on paper will actually fit me, or whether the route of uptake I imagine is real for me. The only way to find out is to keep the option live for a time and test: do I in fact take steps toward it, can I sustain them, do the goods I expected show up for me, and what costs do I encounter? In this sense, keeping an option available has a learning role: it gives me room to check my expectations against my own practice.

[8]I rely on Shaver's (1999) account of rational egoism.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Because of that learning role, how I keep an option open matters. If I can keep it live in ways I can scale back or stop at low cost, then inclusion is unproblematic. I can include the option now, watch my own response, and revise if needed. If, instead, keeping it live requires huge investments or steps that are hard to undo, then the stakes are higher.

4.1 Uncertainty about the Option-set

So far I have followed Kreps in treating the menu of options as fixed. But a further complication arises: the menu itself is uncertain. Options that may become salient tomorrow are largely unknown today. Uncertainty attaches not only to the evaluation of known options but to the very composition of the menu.

To illustrate: imagine Betty is going to a restaurant thinking she will have either pasta or pizza for dinner. As she reads through the menu, she learns about an option she had not considered, eggplant parmigiana, and now prefers it to her original choices. Reading the menu does not just reveal her taste; it creates recognizability and a feasible uptake path for an item that was not in her deliberation at all.

This differs from the transformative experiences Paul describes. There, one knows what the options are but cannot assess their value because pursuing them would transform oneself. Here, the difficulty is that the menu is open-ended: institutional change, technological innovation and shifting practices generate futures not yet available for assessment.

This is closer to the radical or Knightian uncertainty discussed by Keynes (1921) and Knight (1921), where agents cannot assume an exhaustive prior description of the relevant state space. Importantly, the case for flexibility-driven abstraction does not turn on resolving how far this departure from probabilistic uncertainty goes. Even under radical uncertainty, the structural asymmetry remains: the costs of a bounded, reversible trial can be contained by institutional design, whereas the costs of prematurely foreclosing an option that would later have entered one's flexible set can be large and hard to anticipate *ex ante*. When we cannot assign confident probabilities, we can still see that one type of error is systematically more containable than the other. This is enough to support a preference for reversible handling and staged commitment as procedural features, without leaning on expected-utility reasoning in any strong sense.

Once this further uncertainty is introduced, the rationale for flexibility-driven abstraction deepens. Our interest is not merely to uphold rules that protect options we currently deem recognizable, but extends to the social and institutional conditions under which new, recognizable options are likely to emerge. This is why many central constitutional guarantees (freedom of inquiry, open competition, mobility rights, the protection of diverse associations) can be seen as meta-institutions.

With an endogenous option set, the main risk is not carrying an option that later proves worthless, but precluding options that would later belong in my flexible set; the latter loss is typically larger.[9] Hence, many might exhibit a prudential preference toward option-generating meta-institutions.

[9]This is well illustrated by Arrow and Fisher (1974) and Henry (1974).

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

A natural worry is that openness admits futures some will find alien; since recognizability constrains us only once an option is visible, present consent can reach such futures only procedurally. This particular aspect brings in a problem about the extension of solutions to social contract bargaining procedures. Schmidtz (1990) famously argued that actual consent in collective bargaining decision procedures carries stronger force than it does in invisible-hand justifications. Invisible-hand justifications appeal to the idea that a wider social order can be deemed legitimate because it arises from voluntary exchanges. Their weakness, Schmidtz notes, is that while each exchange is consented to, no one ever consents to the emergent order as such.[10]

In contractarian bargaining decision procedures, by contrast, consent is not fragmented across small-scale transactions but gathered into a collective agreement. Bargainers intentionally converge on a common scheme, and this shared act of agreement is taken to show that the arrangement is mutually advantageous, even if its eventual outcomes are uncertain. Unlike emergent processes, where a larger order emerges as an unintended by-product of decentralized choices, contractarian bargaining is purposive: it yields a social order that participants can see as the object of their joint authorization.[11]

Although I agree with Schmidtz that actual consent in collective bargaining procedures carries more justificatory force than actual consent in emergent justifications, I doubt that such consent can reach futures that no one could even describe at the time of agreement. The only exception, I suggest, is where the agreement includes an explicit proviso for yet unknown options: that is, a proviso for the meta-rules that generate and filter them. The upshot can be stated as a crisp thesis: consent cannot reach unknowable downstream options except via consent to the meta-rules of generation, trial and withdrawal. Such a proviso extends consent in a purely procedural sense: what is fixed *ex ante* is the framework for coping with menu change, not the acceptability of downstream distributions once those new options emerge.

Abstraction therefore has to operate upstream, identifying which options and option-generating rules belong in the flexible set. Each agent first fixes a flexible set

[10]See Schmidtz (1990: 100): "The problem is that what people consent to are individual transactions, rather than to the order that spontaneously emerges from them. In other words, that an outcome arose by consent does not entail that people consented to it."

[11]See Schmidtz (1990: 98–99): "Consider contractarianism as a theory about how emergent justification might work. In a contractarian bargaining process, members of a large group seek a collective agreement. Consent to the agreement is taken as a sign that the agreement is mutually advantageous. It is by no means a guarantee, however. (At least, it does not guarantee ex post advantage, which is presumably what bargainers really care about.) People enter the agreement without the benefit of hindsight. Nor does actual consent presuppose rationality in the idealized way that hypothetical consent does. But actual consent carries emergent force regardless, so long as, for example, failures of foresight are not due to fraud. … Contractarian accounts of the state's emergence are distinguishable, at least in a rough sense, from invisible hand accounts. In contractarian models, intentional collective actions leads to an intended and mutually agreeable result. In invisible hand models, bargaining occurs among shifting and relatively small subsets of the collective. The larger scheme of stable society evolves through a series of relatively small-scale exchanges and is an unintended result of such exchanges. There are various agreements between individuals, but there is nothing resembling an agreement to create the emerging social order. The social order emerges spontaneously."

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

that can contain two kinds of admissible items: (i) currently visible options and (ii) option-generating rules (meta-institutions governing the entry, trial and review of yet unknown options). An item enters the set only if it satisfies the inclusion rule already stated: recognizability within the agent's evaluative outlook, attainability through feasible uptake, and a positive prudential balance once preservation costs are counted.

Agents' flexible sets provide the inputs to the social contract bargaining procedure. The object of agreement is not the options themselves but the rules that preserve them. In particular, two sets of rules are needed. First, we deliberate about rules that protect options already present within citizens' flexible sets. Second, we deliberate about option-generating and option-filtering rules. And such rules will do two main things: first, they will affect how many new options appear. Where transaction costs of forming, entering and experimenting are low, more options tend to be generated; where those costs are high, fewer appear.

Second, they will fix how seriously reversibility is taken. Arrangements that admit novelties provisionally, set time limits and review dates, and specify clear exit conditions treat reversibility as a constraint. Arrangements that grant indefinite permissions or embed changes in hard-to-undo structures treat reversibility lightly.

5. Flexibility-Driven Abstraction and Convergence

Flexibility helps convergence because, for any agent who allows that some future options may become *recognizable* and *uptake*-feasible, a rule that permits provisional, reversible trials with review is almost never worse than a rule that excludes such options in advance. The downside of trying and then withdrawing can be contained; the downside of foreclosing an option one would later include in one's flexible set cannot be comparably contained. This makes reversible handling acceptable from within very different evaluative outlooks.

The point does not depend on agreement about values. Agents may disagree sharply about which options they now favour. What they share, other things equal, is (i) a non-zero chance of later recognizability and uptake and (ii) the fact that learning from a bounded, reversible trial has non-negative value. Because these features do not presuppose common ends, a common structure becomes acceptable: allow some generation of new options; admit them provisionally; review against recognizability, uptake, and updated costs; withdraw cleanly if they do not meet these requirements.

Disagreement then moves to levels, not structure. People face different *transaction costs* (which affects how many options are generated) and different *irreversibility costs* (which affects how strict reversibility and review should be). So they will set different thresholds for "how many" and "how strict". But because each person can accept a range of generation and review settings, there is typically a non-empty overlap across agents. Bargaining can select from that overlap without settling the ranking of particular options now.

Only in boundary cases does this fail (e.g. if someone is certain that no new option could ever become recognizable for them, or if reversible trial is impossible or unbounded in cost). Outside such cases, flexibility-driven abstraction widens

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

common ground: it supports convergence on rules for generation and handling while leaving substantive acceptance of particular options to later stages, when those options are actually visible.

The two formal illustrations do different work. The earlier menu model (with options m and f) isolates the core logic: under evaluative uncertainty, deferring commitment by keeping a larger menu has positive value. The model below translates that logic into institutional variables that look like policy levers: how much novelty to allow and how reversibly to handle what emerges. This second model is a tractable special case under probabilistic uncertainty, but as noted earlier in the discussion of radical uncertainty, the mechanism is robust even when the probabilistic structure is underdetermined: the structural asymmetry between bounded reversible trials and potentially unbounded foreclosure losses does not depend on agents being able to assign confident probabilities.

Formally, consider rule packages that set two features of institutional design: (i) how many novel options are generated and (ii) how those options are handled once visible. Let $m \geq 0$ measure the intensity with which new options are generated (other things equal, lower transaction costs correspond to higher m). Let $s \in [0, 1]$ measure the strength of reversible handling (higher s corresponds to lower irreversibility costs: shorter trials, clearer review, cleaner withdrawal).

For each agent i, when a candidate option appears it will be recognizable with uptake for that agent with probability $q_i \in (0, 1]$. If so, following it yields net value $U_i > 0$ for the relevant horizon. Admitting a candidate provisionally has a bounded cost $k(s)$ (trial, screening, carrying), with $k'(s) < 0$. Generating options carries an agent-specific background cost $C_i(m)$, increasing in m and capturing the (direct or indirect) transaction costs that the agent bears.

The expected prudential payoff to agent i from new options under (m, s) is

$$V_i(m,) \quad = \quad m\big(q_i U_i - k(s)\big) \quad - \quad C_i(m),$$

with exclusion baseline $V_i(0, s) = 0$.

Proposition 1 (Common ground near zero). *If there exists s^* such that $k(s^*) < q_i U_i$ for every i, and $C_{i'}(0) = 0$ (or sufficiently small), then there is $\varepsilon > 0$ with*

$$V_i(m, s^*) \quad \geq \quad 0 \quad \text{for all } i \text{ and all } m \in (0, \varepsilon].$$

Proof sketch. The right derivative at $m = 0$ is $q_i U_i - k(s^*) - C_{i'}(0)$, which is non-negative for each i. By continuity, small positive m weakly benefits everyone.

Interpretation. If reversible handling keeps trial losses below the expected value of a genuinely recognizable option, then allowing some generation with review weakly improves on exclusion for all agents. Institutionally, "small m" does not mean abundant novelty at no cost; it means a modest pilot margin: enough room to test a few new options without forcing anyone to bear serious displacement costs. The first step is cheap enough to be jointly acceptable; learning and review do the rest.

Proposition 2 (Convergence without determinacy). *For each i, the acceptable set $A_i = \{(m, s) : V_i(m, s) \geq 0\}$ is non-empty and contains $(0, s^*)$ and, by Proposition 1, some (m, s^*) with $m > 0$. Heterogeneity in q_i, U_i, C_i makes the A_i differ across agents,*

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

but for sufficiently high s (so k(s) is small) their intersection $\cap_i A_i$ *is non-empty (at least for small* $m > 0$*). Thus flexibility yields convergence on a family of rule packages; it does not select a unique settlement.*

Interpretation. "Convergence" here has a specific, modest meaning: the intersection of acceptable rule packages across agents is non-empty. This is convergence in the contractarian sense of finding jointly endorsable rules, not convergence of ideal points or final rankings of ends. What convergence has to accomplish, in a contractarian setting, is to secure the existence of some jointly endorsable rule package that can serve as the object of agreement, even if many packages remain contested; the claim is about existence, not uniqueness, and existence is already a substantive improvement over models where the overlap can vanish. Agents may continue to disagree, perhaps sharply, about which (m, s) pair is best. What the result establishes is that evaluative uncertainty, together with sufficiently strong reversible handling, can *expand* the region of mutual acceptability relative to models in which agents have fixed tastes and no stake in preserving options they do not currently favour. In particular, under the assumptions of Proposition 1, there is guaranteed common ground in a neighbourhood of the exclusion baseline: for s high enough and $m > 0$ small enough, every agent weakly prefers some option generation with review to outright exclusion. Disagreement then concerns *levels*, not structure. Agents can endorse the same procedural shape – permit some option generation; admit provisionally; review against recognizability, uptake, and updated costs; and revise or discontinue authorizations cleanly – while reasonably selecting different thresholds for m (given transaction costs) and s (given irreversibility costs). Because each agent accepts a *range* of (m, s), the acceptable regions can overlap even when ideal points do not coincide; Proposition 1 shows that overlap is guaranteed at least near $m = 0$ when s is sufficiently high.

Lemma (Reversible trial weakly dominates exclusion, one shot). *For any agent with* $q_i U_i > 0$*, admitting a single candidate option provisionally with reversibility level s weakly dominates exclusion whenever* $k(s) \leq q_i U_i$*: exclusion yields* 0*, whereas reversible trial yields* $q_i U_i - k(s) \geq 0$.

Boundary cases. Convergence can fail only at the edges: (i) some agent assigns $q_i = 0$ (they are certain no future option will be recognizable with uptake); (ii) reversible handling cannot bound losses ($\inf_s k(s)$ is large); or (iii) even tiny increases in generation are already costly for someone (very high $C_{i'}(0)$). Outside such cases, *evaluative uncertainty* together with the *recognizability* and *uptake* tests enlarges the overlap of acceptable rules for how many options are generated (m) and how they are handled (s), while leaving the ranking of particular options to later stages, when those options are actually visible.

Asymmetry and refusal to converge. The preceding claims concern the existence of overlap in acceptable procedural packages; they do not imply that agreement is inevitable. One predictable failure mode arises when flexibility is asymmetric. A party open to a wide range of packages can be held up by a party who is rigid. Sometimes this is straightforwardly instrumental: if one side's rigidity shifts the feasible region so that the flexible party bears most of the costs, abstaining can be rational given the outside option. But there is a second mechanism. Even when the flexible party would gain from accepting, she may reject in order not to reward a

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

posture that exploits her openness, or because the other party's rigidity expresses a refusal of reciprocity. Refusal then functions less as ordinary prudence and more as resistance to perceived domination.

If strategic rigidity is a live possibility, then expanding overlap is only part of the task. The design problem is to specify meta-rules that reduce the returns to strategic inflexibility and protect flexible bargainers from one-way exposure: procedural symmetry, mutual exit rights, review rules that bind all parties.

6. Conclusion

Evaluative uncertainty supplies a prudential ground for limited abstraction in social contract bargaining. The focus shifts from Buchanan's "positional" uncertainty – ignorance about where one will sit in the future distribution of power and resources – to the recognition that we cannot safely assume our future selves will care about exactly what we care about now. I also treat the menu of options as shaped by institutions. Some rules make options visible and reachable; others hide them or close them off. On this picture, abstraction is rational when it helps us keep the options we may later have reason to value.

A future belongs in a person's flexible set when it can already be made intelligible under one's higher-order values, when there is a real path by which one could grow into it, either under present rules or under modest reforms we could sensibly adopt, and when the expected gain from keeping it available outweighs the opportunity cost of doing so. This keeps prudence first-personal without inflating it into an indiscriminate concern for futures that prudence cannot give action-guiding reasons to preserve.

Procedures that allow provisional admission with review and clean withdrawal are, other things equal, better than foreclosing in advance. If there is any genuine chance that a newly visible option will later meet the recognizability-and-uptake test, the losses from a bounded trial are contained, while the losses from premature exclusion can be large and irreversible. That asymmetry creates room for agreement on how widely to open entry and how seriously to enforce reversibility and review, even when people continue to disagree about which options they ultimately prefer. Convergence targets procedures rather than final rankings of ends.

Because recognizability and uptake depend on background conditions, the view also points toward meta-rules that lower entry frictions and stage learning: freedoms that let options surface, and procedures that pilot, assess and unwind without entrenchment. In settings where the menu itself evolves, these rules help citizens learn which possibilities genuinely belong in their flexible sets while protecting them against lock-in.

The result is thinner than Rawlsian impartialism and less fragile than Buchanan's positional device. It does not require a standpoint "from nowhere", and it does not collapse once people learn their place in the social order, because evaluative uncertainty persists even when position is known. It preserves motivational continuity by grounding agreement in reasons each person can recognize as her own, while recovering tractability by shifting the object of agreement from comprehensive outcomes to the procedures that govern generation, trial and

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

withdrawal. This approach does not neutralize bargaining power or ensure fairness in any strong sense; evaluative uncertainty by itself leaves outside options and endowments in place. What it can do is expand the bargaining set and reduce the stakes of early defeat by keeping routes of uptake open for those who do not prevail initially.

Two limits bear on the account. First, across the life cycle, evaluative uncertainty typically narrows. New projects become less likely, routes of uptake shorten and the expected gain from keeping additional options live falls relative to their carrying costs. Flexible sets contract and, with them, the private incentive to support option-generating rules.

Second, the intergenerational scope is imperfect. Recognizability and uptake are keyed to the outlooks and opportunities of present citizens; closures today can shrink what future people will be able to see or reach. A purely prudential agreement among people within a generation may therefore overlook options that would matter to those not yet represented. Within this framework, some of the risk can be reduced by rules that keep paths open: broad education and research freedoms, mobility and associative protections, and reversible implementation where feasible. These measures do not settle distributive questions between generations, but they lower the chance that present settlements preclude lives later citizens could intelligibly claim as their own.

These limitations should be weighed against several considerations that, for many agents, push toward supporting greater flexibility in levels as well as in structure. Where agents stand in stable relations of care, mentorship or legacy to members of future generations, the futures of those persons become recognizable under higher-order values the agent already avows, extending the horizon over which option-preservation matters. Even absent such proximity, many projects are future-dependent: their point presupposes the continued existence of practices, institutions and a shared social world beyond one's own lifetime.[12]

Radical uncertainty about which options and evaluative categories will become salient also plausibly increases the expected cost of premature foreclosure, since the losses from excluding options that would later have entered one's flexible set are hard to anticipate and potentially large. And open societies tend to generate new practices, technologies and forms of association, a tendency toward liberal dynamism.[13] This dynamism raises the likelihood that novel options will become recognizable with feasible uptake, strengthening the prudential case for option-generating rules. These considerations tend to raise the perceived downside of low option-generation and weak reversibility for many agents. They do not, however, eliminate heterogeneity: agents still differ in their probabilities of future recognizability, in the value they place on flexibility and in their cost structures. The framework does not yield a single optimal degree of flexibility. What it offers is a structural shape on which convergence is possible, together with an explanation of why, for many, the pressures point toward more rather than less openness.

[12]Scheffler (2013) develops this thought in a different but related context.

[13]See Muldoon (2016) and Schaefer (2023) on the dynamism-enhancing properties of complex societies and liberal rights.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Acknowledgements. I am grateful to my PhD supervisors, Matt Sleat and Alasdair Cochrane, and to my external examiner, Mark Pennington, for careful comments on dissertation chapters that eventually developed into this paper. I am especially grateful to Dora Xu for detailed feedback and generous discussion. I also thank participants in a pre-read seminar at the Manchester Political Theory Group for perceptive questions and suggestions, and two anonymous reviewers for exceptionally helpful reports.

References

Arrow K.J. and A.C. Fisher 1974. Environmental preservation, uncertainty, and irreversibility. *Quarterly Journal of Economics* **88**, 312–319.

Barry B. 1989. *Theories of Justice*. Berkeley: University of California Press.

Barry B. 1996. *Justice as Impartiality*. Oxford: Oxford University Press.

Binmore K. 2005. *Natural Justice*. Oxford: Oxford University Press.

Binmore K. 2007. *Playing for Real: A Text on Game Theory*. Oxford: Oxford University Press.

Brennan G. and J.M. Buchanan 1985. *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.

Brink D. 1997. Self-love and altruism. *Social Philosophy and Policy* **14**, 122–157.

Buchanan J.M. 1975. *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Buchanan J.M. and G. Tullock 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Callard A. 2018. *Aspiration: The Agency of Becoming*. Oxford: Oxford University Press.

Cohen J. 1989. Deliberation and democratic legitimacy. In *The Good Polity: Normative Analysis of the State*, eds. A. Hamlin and P. Pettit, 17–34. Oxford: Blackwell.

Dixit A.K. and R.S. Pindyck 1994. *Investment under Uncertainty*. Princeton: Princeton University Press.

Dworkin R. 1973. The original position. *University of Chicago Law Review* **40**, 500–533.

Gaus G. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.

Gaus G. and J. Thrasher 2015. Rational choice and the original position: the (many) models of Rawls and Harsanyi. In *The Original Position*, ed. T. Hinton, 39–58. Cambridge: Cambridge University Press.

Gauthier D. 1986. *Morals by Agreement*. Oxford: Oxford University Press.

Gutmann A. and D. Thompson 1990. Moral conflict and political consensus. *Ethics* **101**, 64–88.

Habermas J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. W. Rehg. Cambridge, MA: MIT Press.

Hampton J. 1991. Two faces of contractarian thought. In *Contractarianism and Rational Choice*, ed. P. Vallentyne, 31–55. Cambridge: Cambridge University Press.

Henry C. 1974. Investment decisions under uncertainty: the irreversibility effect. *American Economic Review* **64**, 1006–1012.

Hume D. 1978 [1739–40]. *A Treatise of Human Nature*, eds. L.A. Selby-Bigge and P.H. Nidditch. Oxford: Clarendon Press.

Keynes J.M. 1921. *A Treatise on Probability*. London: Macmillan.

Knight F.H. 1921. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

Kogelmann B. 2018. Rawls, Buchanan, and the search for a better social contract. In *Exploring the Political Economy and Social Philosophy of James M. Buchanan*, eds. P.D. Aligica, C.J. Coyne and S. Haeffele, 17–38. London: Rowman & Littlefield.

Kogelmann B. 2025. Creative destruction and the autonomous life. *Journal of Business Ethics* **197**, 659–671.

Korsgaard C.M. 1989. Personal identity and the unity of agency: a Kantian response to Parfit. *Philosophy & Public Affairs* **18**, 101–132.

Kreps D.M. 1979. A representation theorem for "preference for flexibility". *Econometrica* **47**, 565–577.

Lipsey R.G. and K. Lancaster 1956. The general theory of second best. *Review of Economic Studies* **24**, 11–32.

Moehler M. 2018. *Minimal Morality: A Multilevel Social Contract Theory*. Oxford: Oxford University Press.

Morris C.W. 1991. The idea of moral economy. In *Contractarianism and Rational Choice*, ed. P. Vallentyne, 209–236. Cambridge: Cambridge University Press.

Muldoon R. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance*. London: Routledge.

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press

Müller C. 1998. The veil of uncertainty unveiled. *Constitutional Political Economy* **9**, 5–17.
Munger M.C. and G. Vanberg 2023. Contractarianism, constitutionalism, and the status quo. *Public Choice* **195**, 323–339.
Nagel T. 1986. *The View from Nowhere*. New York: Oxford University Press.
Nagel T. 1991. *Equality and Partiality*. New York: Oxford University Press.
Parfit D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Paul L.A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Rae D. 1975. The limits of consensual decision. *American Political Science Review* **69**, 1270–1294.
Rawls J. 1993. *Political Liberalism*. New York: Columbia University Press.
Rawls J. 1999. *A Theory of Justice*, rev. ed. Cambridge, MA: Harvard University Press.
Sandel M.J. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
Scanlon T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
Schaefer A. 2021. Rationality, uncertainty, and unanimity: an epistemic critique of contractarianism. *Economics and Philosophy* **37**, 82–117.
Schaefer A. 2023. Is justice a fixed point? *American Journal of Political Science* **67**, 277–290.
Scheffler S. 2013. *Death and the Afterlife*. Oxford: Oxford University Press.
Schmidtz D. 1990. Justifying the state. *Ethics* **101**, 89–102.
Shaver R. 1999. *Rational Egoism: A Selective and Critical History*. Cambridge: Cambridge University Press.
Sugden R. 1998. Normative expectations: the simultaneous evolution of institutions and norms. In *Economics, Values, and Organization*, eds. A. Ben-Ner and L. Putterman, 73–100. Cambridge: Cambridge University Press.
Sugden R. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.
Thrasher J. 2019. Constructivism, representation, and stability: path-dependence in public reason theories of justice. *Synthese* **196**, 429–450.
Vanberg V.J. 2004. The status quo in contractarian-constitutionalist perspective. *Constitutional Political Economy* **15**, 153–170.
Vanderschraaf P. 2019. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford: Oxford University Press.
Waldron J. 1999. *Law and Disagreement*. Oxford: Oxford University Press.

Carlo Ludovico Cordasco is Lecturer in Management and Organisation Studies at Alliance Manchester Business School, University of Manchester. His research spans normative ethics and political philosophy, with a focus on moral uncertainty and its implications for normative theorizing and institutional design.

Cite this article: Cordasco CL. Abstraction as flexibility: the veil of evaluative uncertainty. *Economics and Philosophy*. https://doi.org/10.1017/S0266267126100728

https://doi.org/10.1017/S0266267126100728 Published online by Cambridge University Press