Journal of Institutional Economics (2026), **22**, e33, 1–13
doi:10.1017/S1744137426100721

RESEARCH ARTICLE

Moral dissonance and judicial dynamics: a differential model of normative misalignment

Massimiliano Marletta

Department of Economics and Business, University of Catania, Italy
Email: massimilianomarletta682@gmail.com

(Received 22 October 2025 UTC; revised 14 June 2026 UTC; accepted 17 June 2026 UTC)

Summary
This article develops a dynamic model of judicial behaviour in which moral alignment operates as an anti-function of judicial utility. The model assumes that moral adherence can generate disutility whenever it conflicts with legal or reputational incentives. Judges maximise a time-dependent utility function driven by normative conformity, reputational rewards, and moral dissonance. Using a system of differential equations, the analysis examines how judicial and social morality co-evolve under competing institutional incentives. The model formalises moral dissonance as an endogenous friction that both destabilises and renews institutional order, transforming moral integrity into a catalyst for normative evolution. This framework provides a micro-dynamic explanation for recurrent moral-legal tensions within judicial institutions.

Keywords: differential institutional dynamics; judicial decision-making; law and morality; moral dissonance; social norms

Introduction

Judicial decision-making is commonly analysed as a form of constrained rational choice, shaped by institutional incentives such as reputation, career prospects, peer evaluation, and adherence to precedent. Within this framework, morality is typically treated as either an internalised preference that enters judicial utility positively or as a reputational asset that reinforces institutional coherence (Epstein and Knight, 1998; Landes and Posner, 1975). However, many salient judicial contexts contradict this assumption. Judges may experience a private cost from acting in accordance with their moral convictions when those convictions conflict with dominant legal doctrines, institutional expectations, or prevailing social norms. In such cases, moral integrity does not enhance utility but reduces it, exposing judges to reputational loss, professional marginalisation, or institutional sanctions. In this article, the term 'institution' is used in the sense common in institutional economics, namely as a system of shared rules, expectations, and incentive structures that shape behaviour within a given organisational environment (Aoki, 2001; North, 1990).

This tension is particularly visible in areas of law characterised by high normative salience, such as constitutional adjudication, rights-based litigation, and judicial activity under conditions of political or social polarisation. In these contexts, courts often move through phases in which judicial decisions appear morally discordant with prevailing social expectations, followed by periods of doctrinal consolidation or retrenchment. Historical and comparative observations suggest that these movements are not exceptional anomalies but recurrent features of legal systems that rely on judicial discretion. Yet standard economic models of judicial behaviour, which emphasise equilibrium and convergence,

© The Author(s), 2026. Published by Cambridge University Press on behalf of Millennium Economics Ltd. 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/S1744137426100721 Published online by Cambridge University Press

provide limited tools to analyse such recurrent patterns of divergence and realignment (Epstein and Knight, 1998; Landes and Posner, 1975).

The theoretical problem addressed in this article is therefore dynamic rather than static. The question is not whether judicial morality matters, but how legal institutions evolve when moral motivation systematically interacts with institutional incentives over time. When moral conviction and reputational rewards point in opposite directions, judicial behaviour generates feedback effects: individual decisions affect perceptions of legitimacy and normative expectations, which in turn reshape the incentive environment faced by future judges. This feedback structure suggests that legal systems may exhibit non-monotonic adjustment paths, including oscillations between moral divergence and institutional alignment, rather than simple convergence towards a stable equilibrium.

This article proposes a dynamic framework to formalise this interaction. The central mechanism is moral dissonance, defined as the disutility experienced by a judge when his or her moral convictions are misaligned with the normative environment that governs institutional rewards. Unlike approaches that treat morality as stabilising by construction, the framework allows moral motivation to generate short-run friction within institutions. These frictions need not imply institutional failure. Instead, under certain conditions, they can produce bounded and recurrent cycles of adjustment between judicial behaviour, institutional norms, and social moral expectations.

Methodologically, the analysis proceeds in two steps. First, the article introduces a simple discrete-time interaction between judicial choice and institutional norm updating. A judge chooses whether to conform to an institutional baseline or to deviate in the direction suggested by his or her moral position. Conformity yields reputational benefits, while deviation yields moral coherence but entails institutional costs. The institution updates its normative stance as an aggregate response to observed judicial behaviour. This discrete-time structure provides transparent intuition for why convergence is not guaranteed and why cycles may emerge when institutional adjustment is slow relative to judicial moral responsiveness.

Second, the article derives a continuous-time representation as a limiting case of this interaction. The resulting system of differential equations links the evolution of judicial moral alignment and institutional conformity through feedback parameters capturing moral responsiveness, institutional inertia, and reputational discipline. Local stability analysis characterises the conditions under which the system converges to alignment, exhibits dampened adjustment, or displays persistent bounded oscillations. These dynamic regimes are interpreted in institutional terms, highlighting how different configurations of judicial independence and institutional responsiveness shape the temporal structure of moral–legal interaction.

The contribution of the article is theoretical and deliberately circumscribed. It clarifies the dynamic conditions under which moral motivation can destabilise judicial equilibria in the short run while contributing to institutional adaptation in the long run, within the specific literature on judicial decision-making under reputational and institutional incentives. The article does not advance a general theory of institutional change; it offers a tractable model of dynamic judicial adjustment under moral dissonance. The final section outlines testable implications and suggests directions for future empirical work based on textual analysis of judicial opinions, citation networks, and longitudinal measures of doctrinal change.

Literature review

The economic analysis of judicial decision-making has traditionally emphasised the role of institutional incentives. Within the public choice and positive political economy traditions, judges are modelled as rational agents responding to reputational concerns, career incentives, political constraints, and hierarchical control (Epstein and Knight, 1998; Landes and Posner, 1975). In these frameworks, judicial behaviour tends to converge towards stable patterns because institutional incentives are designed to reward conformity to legal doctrine and precedent. Moral motivation, when acknowledged, is typically

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

treated as either an internalised preference or a reputational asset that reinforces institutional coherence rather than destabilising it.

Behavioural law and economics relaxed the assumption of perfect rationality by introducing cognitive limitations and systematic biases into judicial decision-making (Guthrie *et al.*, 2001; Jolls *et al.*, 1998). While this literature highlights deviations from strict optimisation, it largely preserves the idea that moral reasoning improves decision quality or enhances coherence. Moral reflection is rarely modelled as a source of disutility in itself. As a result, behavioural approaches focus on error correction and bounded rationality, rather than on persistent tension between moral conviction and institutional incentives.

Recent contributions in legal and political theory have also examined the role of moral motivations in shaping judicial reasoning and institutional behaviour. In particular, Pietrzyk (2025) emphasises how ethical commitments and normative identity can influence judicial decision-making beyond purely instrumental incentives. These approaches highlight that moral motivations may interact with institutional constraints in complex ways, reinforcing the importance of understanding how moral reasoning operates within judicial institutions.

Institutional economics provides a broader framework for analysing such tensions by emphasising feedback between individual cognition and institutional structures. Institutions are understood as systems of shared beliefs and expectations that evolve over time (Aoki, 2001; North, 1990). When individual beliefs diverge from institutional rules, institutions may either adapt or resist. This perspective highlights the judiciary as a key transmission mechanism between individual reasoning and institutional change. However, much of the institutional literature remains descriptive and offers limited formal treatment of how moral disagreement within the judiciary translates into dynamic patterns of institutional adjustment.

The literature on social norms complements this view by modelling norms as equilibria sustained by conditional beliefs about behaviour and approval (Bicchieri, 2006; Ellickson, 1991). Judicial decisions, due to their public and authoritative nature, play a distinctive role in shaping such beliefs. When judicial reasoning aligns with prevailing norms, it reinforces equilibrium; when it deviates, it introduces normative signals that may alter collective expectations. Existing models of norm evolution, however, rarely incorporate judicial actors explicitly or analyse how judicial moral deviation interacts with institutional incentives over time.

Evolutionary approaches to law offer partial insight into these dynamics. Hayek's conception of law as an evolved order and subsequent models of legal change driven by bounded rationality and precedent-based learning emphasise variation and selection as engines of institutional adaptation (Gennaioli and Shleifer, 2007; Hayek, 1973). In these accounts, systematic deviations from established rules generate diversity from which efficient norms may emerge. Moral motivation, however, is typically treated implicitly or conflated with cognitive bias, rather than modelled as an independent source of institutional friction.

In constitutional theory and political economy, the tension between legality and morality is often discussed through concepts such as judicial activism, counter-majoritarianism, and legitimacy. These contributions recognise that judicial deviation from dominant norms can be both destabilising and transformative. Moral disagreement, however, is usually treated as episodic or exceptional, rather than as a persistent feature of judicial behaviour embedded in institutional feedback loops.

Overall, the existing literature provides valuable insights into judicial incentives, cognitive limitations, institutional evolution, and norm dynamics, but it lacks an integrated framework capable of modelling how judicial morality can generate systematic disutility and produce recurrent, non-equilibrium dynamics within legal institutions. The present article contributes to this gap by formalising moral dissonance as an endogenous component of judicial decision-making and by analysing its dynamic interaction with institutional incentives. Rather than assuming convergence by construction, the framework allows for bounded oscillations between moral divergence and institutional alignment, offering a dynamic perspective on the evolution of legal norms.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

Theoretical model

The model constructs a dynamic system linking individual judicial behaviour, moral dissonance, and institutional adaptation. Its purpose is to formalise how an individual judge's moral conviction interacts with legal and reputational incentives to produce oscillatory patterns of alignment and divergence between law and social norms. The framework treats the judiciary as a cognitive-moral subsystem of society, continuously adjusting to moral and legal feedback through differential change.

To provide a robust microfoundation and clear intuition for the continuous dynamics, we first define the system in a discrete-time setting, detailing the judge's instantaneous utility and the institutional reaction function, as suggested by the literature on adaptive dynamics.

Variables: interpretation and measurement scale

All state variables are normalised to the interval [0, 1], representing the position of a judge or institution along a unidimensional normative spectrum anchored at two substantive poles. The value 0 corresponds to the most conservative or deferential position on a given normative dimension (full institutional conformity, rejection of expansive rights-based reasoning, minimal departure from precedent), while 1 corresponds to the most progressive or morally activist pole (full alignment with social moral expectations, maximum departure from formal legal constraints). This scaling is not probabilistic and does not represent subjective uncertainty in the sense of Bayesian decision theory; it captures the degree of normative commitment on a given interpretive axis.

Formally, let J denote the set of judges. Each judge $j \in J$ is characterised at each point in time by two state variables:

- Moral stance $m_{jt} \in [0, 1]$: The judge's internal ethical conviction on the normative dimension under consideration. A value of $m_{jt} = 0$ indicates that the judge holds a position fully aligned with the existing legal-institutional order and assigns no weight to extra-legal moral considerations. A value of $m_{jt} = 1$ indicates that the judge's moral position coincides entirely with the most expansive reading of social moral expectations, regardless of formal legal constraints. Intermediate values represent partial alignment with each pole.
- Legal alignment $n_{jt} \in [0, 1]$: The judge's observable degree of conformity with formal law and institutional expectations, including procedural compliance, adherence to precedent, and professional norms. A value of $n_{jt} = 0$ indicates full institutional conformity; $n_{jt} = 1$ indicates maximal departure from institutional baselines in the direction of moral activism.

The distinction between m_{jt} and n_{jt} is theoretically essential: it captures the classic gap between private conviction and publicly observable behaviour that is central to the study of judicial discretion. A judge may hold strong moral convictions (m_{jt} close to 1) while behaviourally conforming to institutional expectations (n_{jt} close to 0), precisely because reputational and career costs penalise visible deviation. Moral dissonance arises endogenously from the divergence $(m_{jt} - n_{jt})$ and from the divergence of the moral stance from the social environment $(m_{jt} - s_t)$.
Two reference variables are defined at the aggregate level:

- Social morality $s_t \in [0, 1]$: The prevailing normative orientation of society at time t. A value of $s_t = 0$ indicates a fully conservative social environment; $s_t = 1$ indicates a maximally progressive one. This variable evolves exogenously in the model, representing slow-moving sociocultural change.
- Institutional legal baseline $i_t \in [0, 1]$: The normative position encoded in prevailing legal doctrine, precedent, and institutional expectations at time t. A value of $i_t = 0$ corresponds to maximally conservative formal law; $i_t = 1$ to maximally expansive legal doctrine. Unlike s_t, the baseline i_t is partially endogenous: it responds to aggregate judicial behaviour over time.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

The four variables m, n, s, i are distinct and jointly necessary for the model. Social morality s and the legal baseline i are not collapsed into a single variable because legal doctrine and social norms evolve at different speeds and respond to different mechanisms. Similarly, m and n are not collapsed because the judicial incentive structure explicitly penalises the gap between private conviction and public conformity. Collapsing any pair would eliminate the central feedback mechanism that generates the oscillatory dynamics the model is designed to explain.

Discrete-time foundation: utility maximisation and adaptation

Time is discrete, $t = 0, 1, 2, \ldots$. Each judge $j \in J$ maximises a time-dependent utility function $u_{j,t}$ at each period by choosing $(m_{j,t}, n_{j,t})$ subject to the constraints imposed by the normative environment (s_t, i_t).

Judicial utility function and maximisation problem

The instantaneous utility of judge j at time t is:

$$u_{j,t} = \rho_j n_{j,t} - \theta_{1,j}(m_{j,t} - s_t)^2 - \theta_{2,j}(m_{j,t} - n_{j,t})^2 \tag{1}$$

where all parameters are strictly positive: ρ_j, $\theta_{1,j}$, $\theta_{2,j} > 0$. The three components of (1) have the following interpretations:

- $\rho_j\, n_{j,t}$: Activist reputational payoff. Since $n_{j,t} = 1$ denotes maximal visible departure from the institutional baseline in the direction of moral activism, and $n_{j,t} = 0$ denotes full institutional conformity (consistently with the scaling defined in Section Variables: interpretation and measurement scale), the term $\rho_j\, n_{j,t}$ enters utility positively: the judge derives reputational benefit from being publicly perceived as morally progressive. Note that higher n corresponds to greater visible activism, not greater conformity. The coefficient $\rho_j > 0$ captures the intensity of this reputational incentive.
- $\theta_{1,j}(m_{j,t} - s_t)^2$: Social dissonance cost. This quadratic term penalises divergence between the judge's moral stance and prevailing social morality. It captures the psychological and reputational discomfort of being perceived as morally out of step with society.
- $\theta_{2,j}(m_{j,t} - n_{j,t})^2$: Internal (legal) dissonance cost. This quadratic term penalises divergence between moral conviction and observable legal behaviour. It formalises the disutility of hypocrisy: the judge experiences cognitive friction when acting differently from his or her moral convictions.

The judge's problem is to choose $m_{j,t}$ and $n_{j,t}$ to maximise $u_{j,t}$, taking the other parameters as given. The first-order conditions are:

$$\partial u_{j,t}/\partial m_{j,t} = -2\theta_{1,j}(m_{j,t} - s_t) - 2\theta_{2,j}(m_{j,t} - n_{j,t}) = 0 \tag{2}$$

$$\partial u_{j,t}/\partial n_{j,t} = \rho_j + 2\theta_{2,j}(m_{j,t} - n_{j,t}) = 0 \tag{3}$$

From (3), the static optimum requires:

$$m^*_{j,t} - n^*_{j,t} = -\rho_j/(2\theta_{2,j}) \tag{4}$$

This implies that at the static optimum, the judge maintains a negative moral-behavioural gap: observable activist behaviour $n^*_{j,t}$ exceeds private moral conviction $m^*_{j,t}$ by an amount proportional to the reputational return ρ_j and inversely proportional to the internal dissonance sensitivity $\theta_{2,j}$. Since higher n corresponds to greater visible activism, this means the judge performs more activism publicly than his or her private moral stance strictly warrants – a form of strategic moral amplification driven by reputational incentives.Substituting (4) into (2) yields:

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

$$m^*_{j,t} = s_t + \rho_j/(2\theta_{1,j}) \tag{5}$$

which is *strictly above* s_t for all positive parameter values. This result clarifies the mechanism of *moral overclaiming*: when reputational incentives reward visible activism ($\rho_j > 0$), judges optimally set their private moral stance above the prevailing social morality. The intuition is that, since the activist reputational payoff $\rho_j\, n_{j,t}$ creates pressure to perform moral activism in public behaviour (n is pushed above m by the gap condition (4)), the judge compensates by inflating the private moral stance m above s_t in order to limit the social dissonance cost $\theta_{1,j}(m - s_t)^2$ that would otherwise be incurred. At the optimum, these two forces balance: the judge claims a moral stance above social morality, thereby reducing the cost of the gap between m and n that reputational incentives generate. The static optimum is a global maximum because the objective function is strictly concave (the Hessian is negative definite by inspection).

Institutional updating

The institutional legal baseline i_t responds to the aggregate moral orientation of the judiciary, reflecting how collective judicial signalling gradually shapes precedent and statutory interpretation:

$$i_{t+1} = G(\acute{m}_t), \acute{m}_t = \left(1/|J|\right) \sum\nolimits_j m_{j,t} \tag{6}$$

where G: [0, 1] → [0, 1] is monotonically increasing, continuously differentiable, with $G(0) = 0$ and $G(1) = 1$. This ensures that the institution's evolution is bounded and driven by the collective moral state of its agents. Because $i_{t+1} = G(\acute{m}_t)$, the steady-state institutional baseline satisfies $i^* = G(\acute{m}^*)$: any shift in aggregate moral stance $\acute{m}$ produces a corresponding shift in i, which in turn feeds back into each judge's legal alignment $n_{j,t}$ through the adjustment rule below. This feedback channel is incorporated explicitly in the continuous-time analysis in Section Continuous-time dynamics.

A natural alternative is to let i track observable judicial behaviour rather than private morality, setting $i_{t+1} = G(\acute{n}_t)$. Under this specification the institutional baseline responds to what the judiciary visibly does rather than what it privately believes – arguably more consistent with how precedent and doctrine are actually shaped by recorded decisions. This alternative updating rule shifts the eigenstructure of the system towards a saddle configuration under the article's benchmark parameters. Both specifications are analytically coherent; the choice between them reflects a substantive assumption about the transmission mechanism between judicial cognition and institutional change, and is flagged here as an avenue for future empirical discrimination.

Discrete adjustment rules

Because judges cannot jump instantaneously to the static optimum – due to adjustment costs, institutional inertia, and cognitive friction – they revise their moral stance and legal alignment gradually. The discrete transition laws are:

$$m_{j,t+1} - m_{j,t} = \alpha_j(s_t - m_{j,t}) + \beta_j(n_{j,t} - m_{j,t}) \tag{D1}$$

$$n_{j,t+1} - n_{j,t} = \gamma_j(i_t - n_{j,t}) + \delta_j(m_{j,t} - n_{j,t}) \tag{D2}$$

The four parameters α_j, β_j, γ_j, δ_j are judge-specific and have precise economic interpretations:

- $\alpha_j > 0$: Social moral responsiveness. The speed at which judge j adjusts his or her moral stance towards prevailing social morality s_t. A high α_j characterises a judge whose moral convictions are strongly shaped by societal expectations – an 'assimilationist' judge. A low α_j characterises a judge with stable, internally grounded moral convictions that resist social influence.
- $\beta_j > 0$: Internal coherence drive. The strength of the judge's tendency to bring moral conviction into alignment with his or her legal behaviour $n_{j,t}$. A high β_j reflects strong internal consistency motivation: the judge is uncomfortable with hypocrisy and seeks to reduce the gap $(m_{j,t} - n_{j,t})$.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

- $\gamma_j > 0$: Institutional conformity speed. The rate at which the judge's legal behaviour adjusts towards the institutional baseline i_t. This captures responsiveness to hierarchical pressure, precedent, and career incentives. A high γ_j reflects a judge who rapidly internalises changes in formal doctrine.
- $\delta_j \in \mathbb{R}$: Moral expressiveness. The extent to which moral conviction $m_{j,t}$ shapes observable legal alignment $n_{j,t}$. When $\delta_j > 0$, the judge translates moral convictions into legal behaviour: a judge whose moral stance exceeds current practice will push legal alignment upwards. When $\delta_j < 0$, the judge strategically suppresses moral expression, exhibiting greater legal conservatism than his or her private beliefs would imply.

Heterogeneity and the representative-agent interpretation. The parameters $(\alpha_j, \beta_j, \gamma_j, \delta_j)$ are explicitly judge-specific. Different judges may occupy different dynamic regimes simultaneously: a judge with $\alpha_j > \beta_j$ follows the conformity regime while a colleague with $\beta_k > \alpha_k$ follows the counter-majoritarian regime. The aggregate institutional dynamics are governed by the distribution of these parameters across J, mediated by the institutional reaction function $G(\dot{m}_t)$, which aggregates individual moral stances into the collective baseline i_t. The representative-agent interpretation is a limiting case applicable when the judiciary is sufficiently homogeneous or when the analysis focuses on the behaviour of a pivotal judge (such as the median member of a collegial court). In this general case, the model should be understood as describing a distribution of dynamic trajectories across judges, with aggregate behaviour determined by $G(\dot{m}_t)$.

Continuous-time dynamics

Taking the limit $\Delta t \to 0$ from the discrete adjustment rules (D1) and (D2) yields the continuous-time system of coupled differential equations:

$$\dot{m}_j(t) = \alpha_j\big(s(t) - m_j(t)\big) + \beta_j\big(n_j(t) - m_j(t)\big) \tag{7}$$

$$\dot{n}_j(t) = \gamma_j\big(i(t) - n_j(t)\big) + \delta_j\big(m_j(t) - n_j(t)\big) \tag{8}$$

These equations jointly describe the continuous co-evolution of the judge's moral stance and legal alignment. The institutional baseline $i(t) = G(\dot{m}(t))$ enters (8) as a function of the aggregate moral stance, introducing an additional feedback channel: changes in m_j propagate through $\dot{m}$ to i, and thence to $\dot{n}_j$. This indirect feedback becomes active along transition paths and enriches the dynamics of the system.

Dynamic regimes

The relative magnitude of the adjustment parameters determines the qualitative behaviour of each judge's trajectory:

- Conformity regime ($\alpha_j > \beta_j$): Social influence dominates the internal coherence drive. Moral alignment m_j tracks social morality s more closely than legal alignment n. Judges in this regime exhibit adaptive moral behaviour, adjusting their convictions to match the prevailing normative environment.
- Counter-majoritarian regime ($\beta_j > \alpha_j$): The drive to reduce internal dissonance ($m_j - n_j$) dominates social influence. The judge's moral stance is pulled towards legal alignment rather than towards social morality. At the steady state, $m^\star_j$ may stabilise at a level persistently below $s^\star$, characterising a judge who maintains principled independence from social moral trends. This regime formally models judicial counter-majoritarianism as an equilibrium outcome of the individual incentive structure rather than an exogenous institutional choice.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

Steady states and local stability analysis

A steady state (m^*, n^*) is defined by $\dot{m}_j(t) = 0$ and $\dot{n}_j(t) = 0$. Setting $s(t) = s^*$ and $i(t) = i^*$ (so that $i^* = G(m^*)$ under the representative-agent approximation), the steady state satisfies:

$$\alpha_j(s^* - m^*) + \beta_j(n^* - m^*) = 0 \tag{9}$$

$$\gamma_j(i^* - n^*) + \delta_j(m^* - n^*) = 0 \tag{10}$$

Linearising the system (7)–(8) around (m^*, n^*) and evaluating the Jacobian (denoted A to distinguish it from the set of judges J) yields:

$$A = [-(\alpha_j + \beta_j)\beta_j; \delta_j - (\gamma_j + \delta_j)] \tag{11}$$

The trace and determinant of A are:

$$tr(A) = -(\alpha_j + \beta_j + \gamma_j + \delta_j) \tag{12}$$

$$det(A) = (\alpha_j + \beta_j)(\gamma_j + \delta_j) - \beta_j\delta_j \tag{13}$$

Local stability conditions. Let α_j, β_j, $\gamma_j > 0$ and $\delta_j \in \mathbb{R}$. Then:

1. $tr(A) < 0$ always holds when $\delta_j > -(\alpha_j + \beta_j + \gamma_j)$, so the steady state is never a source under empirically plausible parameterisations.
2. The steady state is a stable node or stable focus if and only if $det(A) > 0$, which requires: $\alpha_j\gamma_j + \alpha_j\delta_j + \beta_j\gamma_j > 0$.
3. Oscillatory dynamics (stable focus) occur when $tr(A)^2 < 4\ det(A)$, i.e. when:

$$(\alpha_j + \beta_j + \gamma_j + \delta_j)^2 < 4\left[(\alpha_j + \beta_j)(\gamma_j + \delta_j) - \beta_j\,\delta_j\right]$$

Algebraically, this condition reduces to $(\alpha_j+\beta_j-\gamma_j-\delta_j)^2 < -4\beta_j\delta_j$, which requires $\beta_j\delta_j < 0$. Since $\beta_j > 0$, the stable-focus (oscillatory) regime therefore requires $\delta_j < 0$, i.e., moral suppression. The benchmark numerical illustration below employs this parameterisation.

4. When $\delta_j < 0$ (moral suppression), the sign of $det(A)$ depends on parameter magnitudes, and instability becomes possible.

Parameter space analysis and the empirical relevance of the stable focus

The claim that the stable focus is the empirically most relevant regime requires substantiation beyond the benchmark parameterisation. To assess this, we analyse the full parameter space of the model.

The stable-focus condition $\beta_j\delta_j < 0$ requires $\delta_j < 0$ (given $\beta_j > 0$), that is, it requires moral suppression. Holding β and γ at the benchmark values ($\beta = 0.80$, $\gamma = 1.20$), a grid search over (α, δ) with agnostic uniform priors reveals that the stable focus occupies approximately 27% of the relevant parameter space, while the stable node accounts for approximately 71%, and the saddle for approximately 2%.

The stable focus is therefore not the modal regime in an unconstrained parameter space. However, it is the institutionally most relevant regime for the following reasons. First, moral suppression ($\delta_j < 0$) is the empirically natural condition for judges operating under reputational discipline: strategic underexpression of moral conviction is a well-documented feature of judicial behaviour in high-salience cases, where visible deviation from institutional baselines carries career costs. Second, within the moral-suppression half-space ($\delta < 0$), the stable-focus region corresponds precisely to parameter combinations in which both α (social moral responsiveness) and the absolute value of δ (suppression intensity) are moderate – conditions characteristic of judiciaries that combine meaningful independence with institutional accountability.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

Third, the stable node, while dominant in raw parameter-space share, is institutionally associated with tight hierarchical control and rapid norm absorption – conditions that characterise dependent rather than independent judiciaries, which are not the primary object of the analysis. For these reasons, the stable-focus regime, though not the most frequent in an agnostic prior, is the most theoretically and institutionally motivated parameterisation for the class of judiciaries this article aims to model.

Economic interpretation of dynamic regimes

The local dynamics implied by the trace and determinant of A admit the following institutional interpretations in terms of the underlying parameters:

- Stable node ($det(A) > 0$, $tr(A)^2 \geq 4\ det(A)$): Monotonic convergence to the steady state. This occurs when γ_j (institutional conformity speed) is large relative to δ_j (moral expressiveness). Strong hierarchical discipline and rapid institutional updating quickly absorb moral deviations, leaving no room for oscillatory adjustment. This corresponds to a judiciary under tight institutional control.
- Stable focus ($det(A) > 0$, $tr(A)^2 < 4\ det(A)$): Damped oscillations converging to the steady state. This is the empirically most relevant regime for independent judiciaries, as argued in Section Parameter space analysis and the empirical relevance of the stable focus. It arises when $\delta_j < 0$ – the judge strategically suppresses visible moral expression – while $\beta_j > 0$ maintains internal coherence pressure. The interaction between suppressed moral expression and internal dissonance generates repeated crossings around the equilibrium. The amplitude of oscillations decays over time, reflecting the institutional capacity to metabolise moral conflict into adaptive learning (Figure 1).
- Unstable focus or saddle ($det(A) < 0$ or $tr(A) > 0$): Divergence from equilibrium. This pathological regime arises when moral suppression is so strong that institutional corrective mechanisms cannot compensate. Institutionally, this corresponds to a judiciary in which suppressed moral dissonance accumulates without correction, eventually producing a discontinuous normative break rather than smooth oscillatory adjustment.

Analytical discussion

The dynamic model developed in the previous section characterises judicial behaviour as the outcome of continuous adjustment between moral conviction and institutional incentives. The central analytical implication is that alignment between law and morality need not emerge as a stable equilibrium. Instead, depending on the relative strength of moral responsiveness and institutional inertia, the system may display dampened convergence, persistent bounded oscillations, or instability. This section discusses these regimes and clarifies their institutional interpretation.

Dynamic alignment and bounded oscillations. When reputational discipline and institutional adjustment dominate moral responsiveness, the system converges towards a stable configuration in which judicial moral alignment and institutional conformity coincide. In this regime, deviations from institutional norms are gradually corrected, and moral disagreement plays a limited role in shaping long-run outcomes. This corresponds to settings in which judicial discretion is tightly constrained and institutional feedback is rapid.

By contrast, when moral responsiveness is sufficiently strong relative to institutional inertia, the model generates damped or persistent oscillations. In this regime, moral deviation is neither fully suppressed nor explosively destabilising. Instead, judicial decisions periodically diverge from institutional baselines, induce partial institutional adjustment, and are then followed by renewed pressure towards conformity. The resulting trajectory is characterised by repeated crossings between moral divergence and institutional realignment. Importantly, these oscillations are bounded: the system does not drift indefinitely away from alignment, but neither does it settle into a static equilibrium.

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

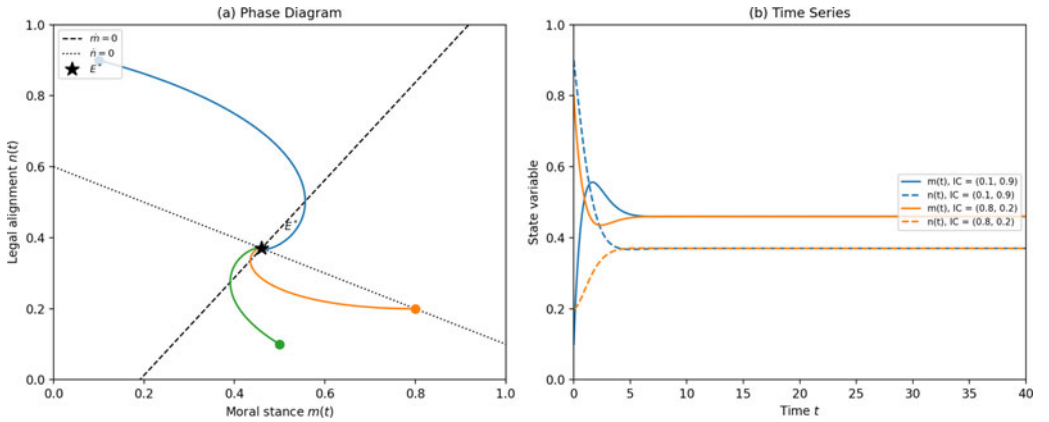


Figure 1. Simulated dynamics of the moral–legal system under a stable-focus parameterisation ($\alpha = 0.30$, $\beta = 0.80$, $\gamma = 1.20$, $\delta = -0.40$, $s^\star = 0.70$, $i^\star = 0.40$; equilibrium $E^\star = (0.46,\ 0.37)$). Panel (a): Phase diagram in the (m, n) plane. Trajectories from three initial conditions spiral inward towards $E^\star$, illustrating damped oscillatory convergence. Panel (b): Time series of $m(t)$ (solid) and $n(t)$ (dashed) for two initial conditions confirming the stable-focus prediction.

Institutional interpretation of oscillatory dynamics. From an institutional perspective, bounded oscillations capture a recurrent pattern in which legal systems alternate between phases of normative contestation and phases of consolidation. Moral deviation by judges functions as a source of variation within the institutional environment, while reputational and hierarchical mechanisms operate as stabilising forces. The interaction between these elements produces a dynamic balance rather than a fixed point.

The model does not assume that moral deviation is inherently desirable or normatively superior. Rather, it shows that moral disagreement can persist endogenously even in the absence of exogenous shocks, provided that institutional adjustment is sufficiently gradual. In this sense, oscillatory dynamics reflect a structural feature of judicial institutions that combine discretion with reputational discipline. Whether such dynamics are beneficial or costly depends on the broader institutional context and lies beyond the scope of the model.

Judicial independence and institutional responsiveness. The parameters governing institutional inertia and reputational enforcement can be interpreted as capturing different degrees of judicial independence. Lower institutional responsiveness increases the amplitude and persistence of oscillations by allowing moral deviation to accumulate before corrective feedback takes effect. Higher responsiveness dampens oscillations and accelerates convergence, but at the cost of reducing the scope for sustained moral disagreement.

The model therefore highlights a trade-off rather than identifying an optimal institutional design. Judicial independence expands the dynamic range of moral–legal interaction, while institutional discipline constrains it. The analysis does not imply that one configuration dominates the other, but rather that different institutional settings correspond to different dynamic regimes. This comparative interpretation aligns with observed variation across legal systems without requiring system-specific assumptions.

Scope and limits of the analytical framework. The analytical framework is deliberately minimal. Social morality and institutional baselines are treated as slowly moving reference points rather than fully endogenous variables. This choice isolates the feedback between individual judicial behaviour and institutional incentives, but it also limits the model's explanatory reach. The analysis does not claim to capture the full co-evolution of law and society, nor does it account for political shocks, strategic interaction among judges, or external enforcement mechanisms.

Within these limits, the model provides a coherent explanation for why judicial morality may generate recurrent, non-equilibrium dynamics rather than monotonic convergence. It clarifies how

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

moral dissonance can persist as an endogenous feature of judicial institutions, shaping the temporal structure of legal adaptation without presupposing either permanent conflict or inevitable harmony.

Case study: the evolution of the 'separate but equal' doctrine

To ground the formal model in historical institutional dynamics, we apply the framework to the transition from *Plessy v. Ferguson* (1896) to *Brown v. Board of Education* (1954). This case provides a canonical example of how judicial systems move from a locked equilibrium to non-linear doctrinal shifts through the accumulation of latent moral dissonance. Following the *Plessy* decision, the institutional baseline was firmly anchored to the segregationist principle of 'separate but equal'. During this initial phase, the system exhibited high institutional rigidity. Even if individual judges harboured private moral objections, the reputational and career costs of departing from such a foundational precedent – which at the time was reinforced by dominant social norms and political structures – were prohibitive. The system resided in a stable equilibrium where judicial output remained closely aligned with the institutional baseline, effectively suppressing judicial moral integrity in favour of systemic predictability.

However, as social morality began to shift – driven by changing normative climates and the post-WWII civil rights movement – the gap between the legal baseline and the external moral reality widened. This period represents a steady increase in the moral dissonance experienced by the judiciary. Judges were no longer merely applying a socially accepted rule; they were managing an increasing utility deficit caused by the friction between their institutional role and evolving normative expectations. The model captures this as a latent pressure where the variables move towards a tipping point: the moment when the 'cost of hypocrisy' begins to outweigh the stability provided by the existing institutional incentives.

The eventual shift in *Brown* (1954) represents this critical threshold. The Supreme Court's unanimous ruling, which explicitly stated that 'separate educational facilities are inherently unequal', was not a marginal adjustment but a bounded oscillation towards a new regime. By overturning decades of precedent, the Court rapidly realigned judicial output with the now-dominant social morality, thereby minimising the collective moral dissonance of the judiciary. Post-1954, the new baseline internalised the anti-segregationist norm via the institutional reaction function, demonstrating the feedback loop where judicial moral conviction reshapes the incentive environment for future judges. This historical pivot confirms that judicial change is rarely a linear progression; rather, it is a dynamic interaction between the rigidity of legal structures and the eventual, non-linear eruption of suppressed moral conviction.

Institutional and policy implications

The dynamic framework developed in this article yields a set of institutional implications that are conditional on the structure of judicial incentives and institutional responsiveness. Rather than identifying optimal rules or prescribing specific reforms, the analysis clarifies how alternative institutional configurations shape the temporal interaction between moral conviction and legal conformity. In this sense, the model contributes to policy analysis by framing institutional design as a problem of managing dynamic adjustment paths rather than achieving static alignment.

A first implication concerns the speed at which institutional baselines respond to judicial behaviour. When institutional updating is rapid, deviations from established norms are quickly absorbed, and moral–legal misalignment tends to be short-lived. Such configurations promote predictability and reduce uncertainty, but they also compress the time window in which moral disagreement can be expressed through judicial reasoning. By contrast, slower institutional adjustment allows deviations to persist and interact, increasing the likelihood of oscillatory dynamics. Institutional inertia thus does not merely delay convergence; it alters the qualitative nature of institutional change by permitting recurrent phases of contestation before consolidation occurs.

A second implication relates to the structure of reputational and career incentives faced by judges. Strong reputational discipline increases the immediate cost of moral deviation and biases behaviour

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

towards conformity. In dynamic terms, this dampens oscillations and accelerates convergence towards institutional alignment. Weaker reputational mechanisms reduce the marginal cost of deviation and allow a wider range of moral positions to be articulated in judicial decisions. The model suggests that reputational structures influence not only individual incentives but also the collective temporal pattern of legal adaptation, shaping whether adjustment proceeds smoothly or through cycles of divergence and correction.

Judicial independence can be interpreted within the framework as a particular combination of limited reputational sanctioning and gradual institutional updating. Such configurations increase the probability of bounded oscillations by allowing moral disagreement to persist without triggering immediate institutional correction. Importantly, the model does not imply that greater independence is inherently beneficial or harmful. Instead, it highlights that independence affects the dynamic regime of moral–legal interaction, expanding the scope for sustained moral deviation while also increasing the duration of adjustment processes. The institutional consequences of independence therefore depend on how it interacts with other elements of the incentive structure.

The framework also has implications for the interpretation of institutional stability. Stability is often equated with convergence towards a fixed legal–moral alignment. The model suggests an alternative perspective, according to which stability may consist in the containment of oscillations within bounded ranges rather than in their elimination. Institutions that combine discretion with reputational discipline may remain stable precisely because moral deviation is neither fully suppressed nor allowed to escalate uncontrollably. From this viewpoint, recurrent phases of divergence and realignment are not necessarily signs of institutional weakness but may reflect a stable dynamic pattern.

From a policy perspective, this insight cautions against institutional reforms aimed exclusively at minimising deviation or maximising short-run coherence. Designs that fully suppress moral disagreement may achieve immediate predictability but risk reducing adaptive capacity by eliminating internal sources of normative variation. Conversely, designs that tolerate extensive moral divergence without sufficient stabilising mechanisms may increase volatility and undermine confidence in legal outcomes. The model frames institutional design as a trade-off between stability and adaptability, mediated by the dynamic interaction between moral motivation and institutional incentives.

Finally, the analysis underscores the importance of context in evaluating institutional arrangements. The same incentive structure may generate different dynamic outcomes depending on the degree of social polarisation, the salience of moral issues, and the broader political environment. As a result, the framework does not support uniform policy prescriptions across legal systems. Instead, it provides a conceptual tool for assessing how changes in judicial incentives or institutional responsiveness are likely to affect the temporal pattern of moral–legal interaction in specific settings. In this way, the model contributes to institutional policy analysis by shifting attention from static outcomes to dynamic processes of adjustment.

Conclusion

This article has developed a dynamic theoretical framework to analyse judicial behaviour when moral convictions interact with institutional incentives over time. By modelling moral dissonance as a source of disutility rather than as a stabilising input by construction, the analysis departs from standard economic approaches that assume an inherent alignment between morality and institutional rewards. The framework shows that alignment between law and morality need not arise monotonically. Instead, depending on the relative strength of moral responsiveness, reputational discipline, and institutional inertia, legal systems may converge towards alignment, exhibit dampened adjustment, or display persistent bounded oscillations.

The contribution is situated specifically within the literature on judicial decision-making under institutional incentives, rather than advancing a general theory of institutional change. Moral motivation, when misaligned with institutional incentives, introduces friction into judicial decision-making. These frictions do not necessarily signal institutional failure. Under plausible conditions, they generate

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press

bounded dynamic patterns that allow legal systems to accommodate moral disagreement while preserving overall stability.

An important implication of the framework is that institutional stability should not be equated exclusively with convergence towards a fixed moral–legal equilibrium. Stability may instead consist in the containment of moral–legal tensions within bounded dynamic trajectories. Legal systems that combine judicial discretion with reputational discipline may remain stable precisely because moral deviation is neither fully suppressed nor allowed to escalate uncontrollably. From this perspective, oscillatory dynamics reflect a structural feature of institutions that rely on judicial interpretation in normatively salient contexts, rather than an anomaly or pathology.

The analysis also highlights the role of institutional design in shaping the temporal structure of moral–legal interaction. Judicial independence, reputational enforcement, and the speed of institutional updating interact to determine whether moral disagreement is rapidly absorbed, persistently expressed, or transformed into recurrent cycles of contestation and consolidation. The framework does not imply that any particular configuration dominates across contexts. Instead, it suggests that different institutional arrangements correspond to different dynamic regimes, each characterised by distinct trade-offs between predictability, adaptability, and the scope for moral reasoning in judicial decisions.

The model is deliberately parsimonious and subject to clear limitations. Social morality and institutional baselines are treated as slowly moving reference points rather than fully endogenous variables, and the analysis abstracts from strategic interaction among judges, political shocks, and external enforcement mechanisms. These simplifications are intentional and serve to isolate the core feedback mechanism between moral motivation and institutional incentives. Future research could extend the framework by endogenising social norms, introducing heterogeneity in institutional responses, or exploring strategic interaction within judicial panels.

Finally, the framework opens several avenues for empirical investigation. Its implications could be explored using textual analysis of judicial opinions, citation networks, or longitudinal measures of doctrinal change in high-salience areas of law. By providing a tractable language for studying dynamic moral–legal interaction, the article aims to contribute to a more nuanced understanding of how judicial institutions evolve over time when moral conviction and institutional incentives are persistently misaligned.

References

Aoki M. (2001). *Toward a Comparative Institutional Analysis*. Cambridge, MA: MIT Press.

Bicchieri C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Brown v. Board of Education, 347 U.S. 483 (1954). *Supreme Court of the United States*.

Ellickson R.C. (1991). *Order without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.

Epstein L. and Knight J. (1998). *The Choices Justices Make*. Washington, DC: CQ Press.

Gennaioli N. and Shleifer A. (2007). The evolution of common law. *Journal of Political Economy* **115**(1), 43–68.

Guthrie C., Rachlinski J.J. and Wistrich A.J. (2001). Inside the judicial mind. *Cornell Law Review* **86**(4), 777–830.

Hayek F.A. (1973). *Law, Legislation and Liberty, Volume I: Rules and Order*. Chicago: University of Chicago Press.

Jolls C., Sunstein C.R. and Thaler R. (1998). A behavioral approach to law and economics. *Stanford Law Review* **50**(5), 1471–1550.

Landes W.M. and Posner R.A. (1975). The independent judiciary in an interest-group perspective. *Journal of Law and Economics* **18**(3), 875–901.

North D.C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Pietrzyk M. (2025). In search of a happy ending in legal interpretation: Cognitive dissonance in judicial decision-making. *International Journal for the Semiotics of Law* **38**(5), 1619–1638.

Plessy v. Ferguson, 163 U.S. 537 (1896). *Supreme Court of the United States*.

Warren C.J.E. (1954). Brown v. board of education. *United States Reports* **347**(1954), 483.

Cite this article: Marletta M (2026). Moral dissonance and judicial dynamics: a differential model of normative misalignment. *Journal of Institutional Economics* **22**, e33, 1–13. https://doi.org/10.1017/S1744137426100721

https://doi.org/10.1017/S1744137426100721 Published online by Cambridge University Press