RESEARCH MANUSCRIPT

Competition and Surplus in Long-Term Iterated Games

A categorical semantics of horizon-dependent payoffs

Anonymous

Abstract

A stage game can have a fixed total payoff at each state even when the players' actions affect which totals will be available later. We formalize this distinction in a time-graded category of stochastic reward processes. Sequential composition pulls continuation payoffs back along transition kernels, and summation is natural with respect to this operation. Finite-horizon constant-sum behavior is characterized by action contrasts on the orbit of the statewise surplus, while discounted infinite-horizon invariance is characterized by an exact reward potential. We derive finite crossover and robustness bounds, distinguish aggregate improvements from equilibrium incentives, and use observation quotients and sheaves to separate local improvement from improvement at a uniform horizon. Three models develop Luca Dellanna's cases of authorship, consulting, and repeat negotiation.

Keywords. Stochastic games, constant-sum games, compositional semantics, continuation payoffs, reward potentials, sheaf semantics.

MSC 2020. 91A15, 91A20, 18F20.

Contents

1 The share and the surplus **4**

1.1 Comparisons and their quantifiers 5

1.2 Main results and position in the literature 5

1.3 Conventions and scope 6

2 A category of stochastic reward processes **7**

2.1 State interfaces and affine continuation 7

2.2 Aggregation and comparison 8

2.3 An order stable under substitution 9

2.4 Retained boundaries 9

2.5 An error estimate for a modular replacement 10

3 Policy evaluation and the strategic boundary **11**

3.1 Public histories and resolved arrows 11

3.2 Infinite evaluation and truncation 12

3.3 Restriction, substitution, and equilibrium 13

3.4 Expected utility and alternative objectives 14

3.5 A constant-sum repeated game 14

4 Finite-horizon constant sum **15**

4.1 Statewise constant sums 15

4.2 Observational reach and maximal delay 16

4.3 Terminal fields as additional tests 17

4.4 One starting law can conceal a contrast 17

5 Discounted potentials and long-run invariance **18**

5.1 A characterization by a bounded potential 18

5.2 A single resolvent and a family of forecasts 19

5.3 Boundary examples 20

5.4 Average payoffs and patience 20

6 Crossover and sensitivity to the horizon **22**

6.1 Persistent per-period differences 23

6.2 A single reversal in marginal gains 24

6.3 Observable margins and incomplete tails 24

6.4 Sensitivity to the unit of patience 25

7 From feasible cooperation to equilibrium **26**
7.1 A one-step deviation theorem . . . 26
7.2 Stationary incentives and public contingencies . . . 27
7.3 Making a transfer feasible . . . 28
7.4 An impossibility from one-sided capture . . . 28
7.5 Participation at intermediate histories . . . 29

8 Welfare boundaries, coalitions, and discount conventions **30**
8.1 A family of linear aggregations . . . 30
8.2 Heterogeneous discounting . . . 30
8.3 Coalition-specific surplus . . . 31
8.4 A decomposition with excluded participants . . . 31
8.5 External effects and incentives can move separately . . . 32

9 Observation quotients and value preservation **33**
9.1 Exact descent of a resolved process . . . 33
9.2 Controlled descent and strategic qualifications . . . 34
9.3 How weak an observable can suffice? . . . 35
9.4 Approximate descent under bounded model error . . . 35

10 Sheaves of horizon comparisons **37**
10.1 Local payoff fields . . . 37
10.2 The existential horizon and its witnesses . . . 37
10.3 Uniform horizons under persistence . . . 38
10.4 Refinement and what it certifies . . . 39
10.5 Comparing gains in all components . . . 40

11 Authorship and demand creation **41**
11.1 A fixed sale and an expandable market . . . 41
11.2 Joint provision and free riding . . . 42
11.3 Persistence, maturation, and saturation . . . 43
11.4 Where the analogy has empirical content . . . 43
11.5 Demand growth with capacity constraints . . . 44

12 Consulting and shared reputation **45**
12.1 Replacing a competitor and shrinking demand . . . 45
12.2 Recovery rather than permanent loss . . . 46
12.3 Private quality incentives . . . 46
12.4 Accounting for replacement value without double counting . . . 47

13 Negotiation and the survival of a relationship **48**
13.1 An active deal and an outside state . . . 48

13.2 Profile comparisons at every horizon . . . 48
13.3 A one-step choice and an equilibrium reading . . . 49
13.4 Opportunity cost and state-dependent disagreement . . . 50

14 Robust certificates and empirical interpretation 51
14.1 A policy deviation identity . . . 51
14.2 Bounds for transition and reward perturbations . . . 52
14.3 Separating a causal claim from a payoff identity . . . 53
14.4 A protocol for interpreting model statements . . . 53
14.5 Two observationally identical initial periods . . . 54

15 Feasible surplus and discounted occupation 55
15.1 The discounted flow equations . . . 55
15.2 Upper and lower certificates . . . 56
15.3 Allocations in the feasible value set . . . 57

16 Approximate invariance and coarse policy comparisons 58
16.1 A residual in the potential equation . . . 58
16.2 An approximate converse . . . 59
16.3 Finite forecast errors . . . 59
16.4 Combining a margin with imperfect observation . . . 60

17 Parallel markets and a common clock 61
17.1 Synchronous composition . . . 61
17.2 Correlated transitions and interactions . . . 62
17.3 Asynchronous opportunities . . . 63

18 Prior work and conclusions 64
18.1 Dynamic programming and stochastic games . . . 64
18.2 Open games and categorical probability . . . 64
18.3 Repeated games, reputation, and value creation . . . 65
18.4 Conclusion . . . 65
18.5 Questions left open . . . 66

References 67

1 The share and the surplus

Luca Dellanna's essay on long-term iterated games [1] contrasts rivalry for the next transaction with a shared interest in creating more transactions. Two authors may want the same reader to buy their respective books on a particular day. The same authors may each benefit when readers learn why the subject matters and buy more books over the following years. A consultant might gain a client after a competitor's failure, even though that failure reduces demand for consulting. A negotiator may receive more from a deal while reducing the probability of dealing with the same counterparty again. The examples distinguish decisions that change a division from decisions that change the object being divided.

Here is a minimal mathematical statement of the distinction. At state s, every joint action a distributes an amount $c(s)$ among the players. Its effect on the next state has law $P(s, a, -)$. With discount $\delta \in (0, 1)$, the total payoff over the current and next dates is

$$c(s) + \delta \sum_{y} P(s, a, y)c(y). \tag{1.1}$$

The first term is fixed at the current state. The second varies with a whenever available actions shift probability between states of different expected surplus. A common current total therefore imposes no requirement that the two-stage game have a common total. This observation becomes a theorem about an arbitrary number of stages when the functions $c(y)$ in (1.1) are replaced by an appropriate family of future forecasts.

An absolute positive payoff sum depends on the choice of origin. Adding a fixed amount to every profile changes the sign of an absolute sum while preserving incentives and all rankings. We use *strict aggregate improvement* for a positive difference from a specified baseline, and *constant sum* when all profiles in the chosen comparison class give the same aggregate. A strict Pareto improvement raises every modeled player's value. Equilibrium concerns each player's opportunities to deviate with the others held fixed. A longer horizon may affect all of these predicates, and they can change at different dates. The fixed-profile comparison also differs from the equilibrium of a game that publicly announces a different stopping date.

1.1 Comparisons and their quantifiers

A claim that the short-run game is constant-sum can refer to the entire stage game at a specified state or to a particular intervention that only changes a split. For the stage game the quantifier ranges over all feasible joint actions at that state. For an intervention we may compare only two joint actions, or two profiles on a larger game with many other possibilities. A strict pairwise improvement at a later horizon establishes variation at that horizon, but the behavior of other profiles still requires examination.

We always start a comparison from the same initial state or public history. If a population distribution of starting states is used, it is shared between the profiles. Otherwise a difference can result from initial conditions rather than a changed action or continuation. Theorems about invariance from every state quantify over all initial states, including ones that a particular application assigns probability zero initially.

For the modeled players, assume payoffs are cardinal and expressed in compatible units, with a common discount unless stated otherwise. The sum may represent producer surplus for a specified group or a weighted welfare convention fixed in advance. An industry revenue increase does not imply social welfare growth when customer surplus, competitors outside the modeled set, or resource costs are omitted. Treating a payment as a pure transfer is valid when the payer and recipient both lie within the boundary and units match. Any payment to an excluded party changes the measured aggregate. All subsequent uses of total refer to this declared boundary.

1.2 Main results and position in the literature

Stochastic games provide the underlying notion of an action-controlled transition [3]. The usual repeated game with a fixed stage game arises as a special case. Open games organize strategic components through sequential and parallel composition [5]. The iteration of open games has also been treated using a final coalgebra and a subgame-perfection operator [6]. Here we retain stochastic transitions and rewards after a strategy profile has been chosen. This makes the evaluation functor explicit and isolates when a scalar observable can detect a change in the transition.

Our first result identifies the precise subspace of state observables relevant to finite-horizon constant-sum behavior. It is spanned by successive forecasts of the one-stage total under any one deterministic baseline. Each controlled action must agree with the baseline on this span, even if the two kernels differ on other state observables. The first forecast on which a contrast occurs identifies the first horizon with a possible strict aggregate ordering. A second result identifies the corresponding

infinite-horizon condition through a single bounded potential. It is an expected-reward form of discounted telescoping, related to potential-based reward shaping [7].

Subsequent chapters give quantitative crossover tests, distinguish transfers from incentives, and study what can be seen through coarser observations. A sheaf interpretation distinguishes horizons chosen locally across contexts from a single horizon valid across a whole domain. Later examples use the same semantics for three distinct mechanisms: market creation, a shared reputation loss, and a change in the survival of a relationship. The examples translate mechanisms in the essay into explicit assumptions. They use no estimates of the parameters for the people or markets in the essay.

1.3 Conventions and scope

Time is discrete, starting at date zero. An N-stage payoff contains dates $0, \dots, N-1$. Terminal value, when supplied, is evaluated at date N and discounted by δ^N. A common $\delta \in (0,1)$ makes all comparisons between expected sums well defined for bounded rewards. A geometric survival probability can sometimes be absorbed into the discount for a fixed policy. Keeping survival in the state transition is important when a present action changes it.

We permit finite or countable state interfaces in the categorical construction. The strategic games have finite state and action sets except where public histories are used as states or an example explicitly uses an infinite family of separate contexts. Independent behavioral randomization at a public history suffices for the proofs. Correlation can be represented by an explicit public signal included in the state before decisions. We assume that the reward argument available to a player is the expected one-stage utility. Expectation equivalence in this model can leave full payoff distributions different.

2 A category of stochastic reward processes

2.1 State interfaces and affine continuation

Fix a finite nonempty player set I and $0 < \delta < 1$. For a finite or countable set X, let $\mathcal{V}_I(X) = \ell^\infty(X, \mathbb{R}^I)$, with coordinatewise order and supremum norm over states and coordinates. A stochastic kernel P from X to Y sends x to a probability distribution $P(- \mid x)$ on Y. It pulls back bounded fields by

$$(P^* v)(x) = \sum_{y \in Y} P(y \mid x) v(y).$$

The map is positive, fixes constant fields, and has operator norm at most one. If Q goes from Y to Z, write QP for the kernel obtained by using P and then Q. Absolute convergence gives $(QP)^* = P^* Q^*$.

Definition 2.1. A morphism $f : X \to Y$ in $\mathcal{R}_{\delta,I}$ is (n, P, r), with grade $n \in \mathbb{N}$, kernel P from X to Y, and expected reward $r \in \mathcal{V}_I(X)$ in the units of the starting date. For $g = (m, Q, u) : Y \to Z$, define

$$g \circ f = (n + m, QP, r + \delta^n P^* u). \tag{2.1}$$

The identity of X is $(0, \mathrm{id}_X, 0)$.

Grade zero permits a change of interface at a single instant. The applications to observation use grade-zero arrows with zero reward. The category allows other grade-zero rewards algebraically, while stage games below associate one date of elapsed time to each one-stage arrow.

Proposition 2.2. *Formula (2.1) defines a category. Grade is a functor to the one-object category $(\mathbb{N}, +, 0)$. Forgetting grades and rewards gives a functor to the category of stochastic kernels. The operation*

$$f^\sharp : \mathcal{V}_I(Y) \longrightarrow \mathcal{V}_I(X), \qquad f^\sharp(v) = r + \delta^n P^* v \tag{2.2}$$

is a contravariant functor to ordered real vector spaces with monotone affine maps.

Proof. For a third arrow (k, L, w) either bracketing yields the kernel LQP, grade $n + m + k$, and reward

$$r + \delta^n P^* u + \delta^{n+m} P^* Q^* w.$$

The two identities follow using grade zero and the identity kernel. Evaluating (2.1) on v gives $(g \circ f)^\sharp(v) = f^\sharp(g^\sharp(v))$. Positivity of P^* and $\delta^n > 0$ make the affine map monotone. □

Composition keeps the terminal distribution and the expected accumulated payoff. A stage reward may in reality be random and correlated with the next state. Its conditional expectation given the source state and current action suffices for expected utility with additive intertemporal rewards. Correlations with future observations may need to be kept in the transition state when they influence later decisions. This restriction explains why the semantics applies to the evaluation of policies on adequate state interfaces.

One can recover the three components of a morphism from $f^\sharp$ for nonempty X and Y. Evaluating at zero gives r. Evaluating on a constant coordinate field and subtracting r recovers δ^n, hence n. Evaluating on singleton indicators then recovers the rows of P. This separation uses the full space of terminal fields, which need not be available to an actual game.

2.2 Aggregation and comparison

Write $\mathsf{sum}_X(v)(x) = \sum_{i \in I} v_i(x)$. The scalar reward category is $\mathcal{R}_{\delta,\{*\}}$. Define

$$\mathsf{A}f = (n, P, \mathsf{sum}_X r), \qquad \mathsf{sum}_X f^\sharp(v) = (\mathsf{A}f)^\sharp \mathsf{sum}_Y(v). \tag{2.3}$$

The equality follows from commutation of sum with kernel pullback. Therefore A is a functor and sum is a natural transformation between the continuation functors on the common domain $\mathcal{R}_{\delta,I}$. The zero-sum directions form the canonical subspace $\ker \mathsf{sum}_X$. Choosing $u \in \mathbb{R}^I$ with $\sum_i u_i = 1$ gives a splitting $r = (\mathsf{sum}_X r)u + z$, where $z \in \ker \mathsf{sum}_X$. A different choice of u changes the reference allocation. The kernel and scalar projection are intrinsic.

Proposition 2.3 (Continuation-sensitive comparison)**.** *Let f and f' have the same source, target, grade n, and aggregate reward. For their respective kernels P and P', the aggregate comparison against a common terminal field v is*

$$\mathsf{sum}_X(f^\sharp(v) - f'^\sharp(v)) = \delta^n (P^* - P'^*) \mathsf{sum}_Y(v). \tag{2.4}$$

The two arrows agree in aggregate against every bounded terminal field if and only if $P = P'$.

Proof. Subtract (2.2) and use (2.3). For each $y \in Y$, put $\mathbf{1}_{\{y\}}$ in one payoff coordinate and zero in the others. Equality for these fields forces equality of every kernel entry. The converse follows from the displayed equation. □

For a restricted family of terminals, (2.4) only tests the span of their aggregate fields. An action can therefore alter a distribution and leave aggregate value fixed in all environments under consideration. The distinction between changes of kernel and changes of observable is essential to the finite-horizon theorem.

2.3 An order stable under substitution

For parallel arrows of equal grade, define $f \preceq g$ when $f^\sharp(v) \leq g^\sharp(v)$ for every bounded v. Taking $v = 0$ implies $r_f \leq r_g$. If the kernels differed, a singleton indicator multiplied by a sufficiently large signed scalar in one coordinate would contradict the inequality at some state. Thus $f \preceq g$ is equivalent to equality of the kernels and coordinatewise comparison of the rewards. This gives a partial order on each hom-set if arrows of different grades are incomparable. Composition is monotone in both variables, since postcomposition supplies a terminal field to the compared arrows, and precomposition applies a common monotone continuation operator. Hence the category is locally posetal as a 2-category. The functor A preserves the order. A horizon-specific comparison can order arrows with different kernels because it tests only the actual continuation.

For probability measures μ, ν on a countable set, use $\mathrm{TV}(\mu, \nu) = \sup_B |\mu(B) - \nu(B)|$. Then for bounded scalar v,

$$\left|\sum_y (\mu(y) - \nu(y))v(y)\right| \leq \mathrm{TV}(\mu, \nu)\ \mathrm{osc}(v), \qquad \mathrm{osc}(v) = \sup v - \inf v. \tag{2.5}$$

Subtract $\inf v$ and bound the positive and negative parts of the signed measure separately. Applied to (2.4), this bounds the magnitude of an effect through its transition shift and the range of future surplus.

2.4 Retained boundaries

A terminal boundary is part of a valuation. Suppose two arrows of the same grade agree at terminal zero. Every later evaluation compares them by a kernel contrast on the remainder's value. The common prefix case is different. Applying the same earlier arrow to two continuations pulls the difference backward and multiplies it by the

prefix discount. A strict downstream inequality may turn weak at a particular initial state if its downstream strict region has zero probability. If a coordinate difference is nonnegative everywhere and positive on a set of positive probability, its expectation is positive. These are consequences of positivity of the kernel and require no equilibrium or optimality criterion.

There is a corresponding statement for processes with different grades. A deterministic wait of k dates with zero reward at an unchanged state maps v to $\delta^k v$. The state distribution can be identical before and after a wait while its present value is strictly lower for a positive terminal reward. The grade records an economically relevant observable that the underlying category of kernels omits. This also explains why time spent expanding a market and time spent distributing its receipts should be counted on the same clock.

2.5 An error estimate for a modular replacement

Suppose a grade-n prefix is shared by two evaluated plans, and the two tails have values that differ by at most ε in the supremum norm. Their complete values then differ by at most $\delta^n \varepsilon$. For multiple replacements at different boundaries one can form a chain of intermediate plans, replacing one component at a time. If the jth replacement changes the continuation by at most ε_j at a boundary after n_j dates, the total error is bounded by $\sum_j \delta^{n_j} \varepsilon_j$. This argument requires each local comparison against the actual suffix of the intermediate plan. A bound obtained by evaluating a local component at zero may fail when a nonzero continuation is attached.

For a scalar illustration, two initial components may both pay zero and feed a high-value or low-value state with different probabilities. Against terminal zero they tie. A suffix that pays a high-state dividend makes their values differ. Supplying a class of possible suffix values at the boundary lets an analyst control the comparison before choosing the final plan. Equation (2.5) gives a bound using only the largest allowed oscillation and the component's kernel discrepancy.

3 Policy evaluation and the strategic boundary

3.1 Public histories and resolved arrows

Let S be a finite state set. At state s, the finite nonempty set $A_i(s)$ gives the actions of player i and $A(s) = \prod_i A_i(s)$ gives the joint actions. The transition $P(s, a, -)$ is a probability distribution on S, and $r(s, a) \in \mathbb{R}^I$ is the expected stage payoff. Write $M = \sup_{s,a} \|r(s, a)\|_\infty$. A public history is a sequence $h = (s_0, a_0, s_1, \ldots, a_{t-1}, s_t)$ with all the recorded transitions possible and all actions feasible. A history records actual actions and states, so it can encode the effects of a player's past choice on opponents' future behavior.

A behavioral strategy gives a distribution on $A_i(s_t)$ for each feasible h. A profile π induces a joint distribution π_h at each h by independent randomization. The values of a strategy at histories that have zero probability under π are still specified, because unilateral deviations can reach them. We can include a public signal in the history before a simultaneous decision and then take products of conditional distributions. This is the convention for public correlation. Private signals and privately observed states require other notions of admissible strategies and are outside the equilibrium theorem below.

Let X_t be the set of feasible histories of length t. A fixed profile resolves a stage to $f_t^\pi = (1, P_t^\pi, r_t^\pi) : X_t \to X_{t+1}$. Explicitly,

$$r_t^\pi(h) = \sum_{a \in A(s_t)} \pi_h(a) r(s_t, a),$$
$$P_t^\pi((h, a, y) \mid h) = \pi_h(a) P(s_t, a, y).$$

The stage field is defined for every feasible h even when the profile cannot reach it. The probability of appending any impossible history is zero. This representation separates the original action-controlled kernel from the induced law after strategies have been selected.

For a segment starting at a public history h, reset the index of discounting to zero. Compose the next N resolved arrows and evaluate at zero. Induction using (2.1) and

the law of iterated expectation gives

$$V_N^\pi(h) = \mathbb{E}_h^\pi \left[\sum_{t=0}^{N-1} \delta^t r(s_t, a_t) \right]. \tag{3.1}$$

At a nonzero terminal field u on date-N histories, the result adds $\delta^N \mathbb{E}_h^\pi u(h_N)$. If future behavior depends only on the current state and the date, the state s_t suffices as an interface. Using histories proves the formula without a Markov assumption on the policy. One may use distinct interfaces at different dates. This lets the formalism handle a finite tree and a stationary stochastic game in the same category.

3.2 Infinite evaluation and truncation

The bounded expected value

$$V_\infty^\pi(h) = \mathbb{E}_h^\pi \left[\sum_{t \geq 0} \delta^t r(s_t, a_t) \right]$$

exists, and

$$\sup_{\pi,h} \| V_\infty^\pi(h) - V_N^\pi(h) \|_\infty \leq \frac{M\delta^N}{1-\delta}. \tag{3.2}$$

The geometric bound proves convergence uniformly, and justifies interchanging the sum and expectation. The same proof works when some histories correspond to the game having stopped. We assign the stopped state reward zero and keep it absorbing, which turns variable-length play into a fixed sequence of dates.

For a stationary Markov profile, let P_π and r_π be its resolved state kernel and field. The stationary evaluation is the unique bounded solution of

$$V_\infty^\pi = r_\pi + \delta P_\pi^* V_\infty^\pi. \tag{3.3}$$

The map on the right is a contraction of ratio at most δ. A history-dependent policy is stationary on the disjoint union of its public histories in the sense that there is one kernel that appends the next stage to any given history. Thus (3.3) remains valid as an equation on bounded history fields. This fixed point evaluates a specified policy. It does not solve the set of strategic best-response equations. Bellman's functional equation provides the dynamic-programming lineage of the evaluation [2], and Shapley's model solves a zero-sum strategic problem with a related contraction [3].

A useful consequence of the affine structure concerns boundary changes. If two terminal values differ uniformly by at most ε, they affect an N-stage evaluation by at most $\delta^N \varepsilon$. The discount suppresses errors far from the beginning, while errors made every period may accumulate with a factor as large as $(1-\delta)^{-1}$. Later chapters distinguish these two sources of uncertainty. Terminal payoffs can also encode resale, liquidation, or obligations left by the original game. Setting them to zero should be understood as choosing a particular experiment in which post-horizon payments are excluded.

3.3 Restriction, substitution, and equilibrium

Sequential composition identifies an interface at which a continuation can be substituted. Suppose two tails have equal value at each interface state. Replacing one with the other leaves the value of every common prefix unchanged by (2.2). This remains true for values depending on full histories. At a state-only interface, the equality needs to hold conditional on the state and any other information that a replacement is allowed to read. Forgetting history can give an apparent equality that fails after substitution.

For a stronger equivalence of strategic games, a tail replacement would also have to preserve the effects of unilateral deviations and which continuation choices are available. An arbitrary equality between two on-path value fields does not preserve that information. Consider a common path with mutual cooperation. One strategy can prescribe punishment following a unilateral deviation, while another prescribes continued cooperation. Their values on the common path agree. In the first case the opponent's continuation can deter the deviation. The second continuation leaves its immediate reward unopposed. The difference is encoded in values at off-path histories, before any attempt to aggregate the strategic data.

The open game literature addresses this distinction by making an open game's strategic behavior sensitive to a supplied context and continuation [5, 6]. Our use of policy-resolved arrows avoids imposing an equilibrium in each factor of a long composite. An allocation stage that would admit a deviation when evaluated in isolation can be sustained by the continuation attached to it. Conversely, behavior that is optimal in a closed one-step game can become suboptimal when the same physical transition feeds a different continuation.

3.4 Expected utility and alternative objectives

The evaluation is linear in distributions. If a player instead cares about a nonlinear functional of the discounted stream, the same expected reward can correspond to different values. For instance, the deterministic payoff 1 and an equiprobable payoff of 0 or 2 have the same expectation. With payoff assessed by $\sqrt{x}$ on nonnegative terminal amounts, their values differ. Applying a nonlinear objective after forming the discounted sum can also introduce path dependence even if its expected one-stage rewards form a Markov process.

When cumulative wealth is a sufficient statistic, the state can include wealth and the terminal continuation can evaluate a utility of that wealth. That construction changes which scalar fields can be compared and generally changes whether a flow is a pure transfer in the chosen units. The results below use additive expected payoff. The restriction is substantive for an application involving credit constraints, bankruptcy risk, or a short-run loss with sharply nonlinear effects on the players.

3.5 A constant-sum repeated game

Suppose the same stage game repeats with no action-controlled market state and with $\sum_i r_i(a) = c$ for every joint action. Profiles may condition on all realized actions. For every realized play the aggregate flow is c at every date. Hence $\mathsf{sum} V_N^\pi = c(1 - \delta^N)/(1 - \delta)$ for all profiles, and $\mathsf{sum} V_\infty^\pi = c/(1 - \delta)$. Future responses can change individual payoffs, so equilibrium behavior can be history dependent even though no profile improves on another in the aggregate. The mechanism in Dellanna's essay requires a future opportunity or a future aggregate payment that changes with behavior, even if that opportunity is represented only as an effect of history.

4 Finite-horizon constant sum

4.1 Statewise constant sums

Let the game data be stationary with finite states and actions. The hypothesis of one-stage constant sum at each state is

$$\sum_i r_i(s,a) = c(s) \qquad \text{for all } s,\ a \in A(s). \tag{4.1}$$

The function c may vary by state. We write $P_a f(s) = \sum_y P(s,a,y) f(y)$ for an action available at s. Fix any deterministic Markov baseline b, with $b(s) \in A(s)$ and $P_b f(s) = P_{b(s)} f(s)$. Let

$$\mathcal{K}_b(c) = \operatorname{span}\{c, P_b c, P_b^2 c, \ldots\}, \qquad L_{s,a} f = P_a f(s) - P_b f(s).$$

The sequence $P_b^j c$ consists of forecasts of the surplus payable j dates after the current state when the baseline is used. Its span is invariant under P_b. On a finite state set this is a finite-dimensional subspace even though it is generated by an infinite sequence.

Theorem 4.1 (Finite-horizon invariance)**.** *Under* (4.1)*, the following are equivalent.*

(i) From every initial state and at every finite horizon, expected aggregate value with terminal reward zero is independent of the behavioral profile.

(ii) Every $L_{s,a}$ vanishes on $\mathcal{K}_b(c)$.

If (ii) fails, let k be the least nonnegative integer for which $L_{s,a} P_b^k c \neq 0$ for some s, a. Then $k+2$ is the first horizon at which aggregate invariance fails somewhere. A single changed initial joint action followed by the baseline witnesses failure.

Proof. Let $g_n = \sum_{j=0}^{n-1} \delta^j P_b^j c$, with $g_0 = 0$. This is the baseline aggregate value of n stages. Starting at s, force action a and then follow the baseline. The difference from playing the baseline throughout the $(n+1)$-stage game is

$$\delta L_{s,a} g_n. \tag{4.2}$$

Under (i) this vanishes for every n, s, a. Since $g_{j+1} - g_j = \delta^j P_b^j c$, subtracting the equalities for consecutive n gives (ii).

Conversely, assume (ii). If all n-stage continuation values equal g_n conditional on any possible history at the given current state, then every action at that history gives

aggregate value $c(s) + \delta P_a g_n(s) = g_{n+1}(s)$. Mixing among the actions yields the same value. Induction starting at $n = 0$ proves (i), including history-dependent profiles.

If k is the first nonvanishing contrast index, the same induction applies whenever $n \leq k$, so all horizons through $k+1$ remain invariant. At a witnessing state and action, (4.2) for $n = k+1$ equals $\delta^{k+1} L_{s,a} P_b^k c \neq 0$. □

The criterion is independent of which baseline was chosen. The theorem makes a stronger statement about the span itself under invariance. If it holds, then for every deterministic baseline b' and $f \in \mathcal{K}_b(c)$ we have $P_{b'} f = P_b f$ pointwise. Inductively $P_{b'}^j c = P_b^j c$ for every j, and $\mathcal{K}_{b'}(c) = \mathcal{K}_b(c)$. Before the first failure, this calculation applies to those generators on which all contrasts are known to vanish. Agreement between baseline orbits is a consequence of the observable restriction, even if the underlying transition matrices differ considerably.

4.2 Observational reach and maximal delay

If c is globally constant, all kernels fix c and (4.1) remains constant-sum at every horizon. If $P(s, a, -)$ is independent of a at every state, every action contrast vanishes regardless of c. The theorem also covers cases in which both c and the kernel are nontrivial. For example, two actions can give very different distributions inside a collection of states which all have the same current and future aggregate forecasts. A later shift between collections with different forecasts would be visible.

A bound on the first witness follows by linear algebra. Suppose $\dim \mathcal{K}_b(c) = m > 0$. The first m vectors $c, P_b c, \ldots, P_b^{m-1} c$ span it. Indeed, once a generator belongs to the span of its predecessors, applying P_b puts every later generator in that span. If a contrast is nonzero, its first occurrence is therefore at $k \leq m - 1$, giving a first failure by horizon $m + 1$. When $c = 0$ the span is zero and there is no witness. This is a dimension statement about the model. No enumeration of finite games enters the proof.

The delay can be made arbitrarily long by increasing the number of states. Take a branch at an initial state, with one action leading to a zero absorbing state. The other leads into a deterministic string of states whose first nonzero surplus is received at date L from the branch. Use no further choices and place a surplus of 1 at that date. Both profiles tie at all horizons $N \leq L$. At horizon $L + 1$ their difference is δ^L. Thus in a family of games with longer maturation times, common one-stage data can coexist with witnesses at arbitrarily remote dates. The delay in this example is a property of the transition and reward structure. Increasing patience increases the value of the eventual difference but leaves the first date at which it is nonzero unchanged.

4.3 Terminal fields as additional tests

The invariance theorem sets the terminal field to zero. For a scalar terminal u, the baseline n-stage value becomes $g_n + \delta^n P_b^n u$. In the forced-action comparison, the extra contrast equals $\delta^{n+1} L_{s,a} P_b^n u$. A family of admissible terminal fields therefore changes the observables against which arrows are tested.

Corollary 4.2. *Let U be any collection of bounded scalar terminal fields. Assume that $0 \in U$ and (4.1) holds. Independence of the scalar value from policy, from every state and at every finite horizon for every $u \in U$, is equivalent to the vanishing of every $L_{s,a}$ on*

$$\operatorname{span}\big(\{P_b^j c : j \geq 0\} \cup \{P_b^j u : j \geq 0,\ u \in U\}\big).$$

Proof. The zero terminal and Theorem 4.1 give the tests on c. For a fixed terminal u and each n, comparing initial actions against the baseline continuation yields the additional test on $P_b^n u$. Conversely, induction with initial field u gives the baseline value from every history, since each $L_{s,a}$ vanishes on its n-stage continuation. □

If U includes every state indicator, independence already at horizon one forces equality of all transition rows for each fixed source state, as in Proposition 2.3. An actual market setting might instead specify a narrow set of continuation claims, such as terminal inventory value or residual contract value. The corresponding notion of indistinguishability can be much coarser.

4.4 One starting law can conceal a contrast

The universal initial state quantifier is necessary to recover every statewise condition. Suppose there are two disjoint components and the initial distribution is supported in one. A contrast in the other component never contributes to an expectation from that distribution. Even within the support, a specific pair of policy interventions can create positive and negative differences in different states which cancel when averaged. Conversely, if every state is initially assigned positive probability and one can choose any state-contingent profile, a nonzero contrast can be isolated at its witnessing state for a one-step comparison against the baseline. Changing only the action there makes the average difference the witness times the initial state's positive mass. A report of invariance under a single initial distribution should specify which interventions were available when assessing it.

5 Discounted potentials and long-run invariance

5.1 A characterization by a bounded potential

In this chapter write $c(s,a) = \sum_i r_i(s,a)$ without assuming (4.1). Let the game data be stationary, with bounded rewards. Consider the proposition that the infinite discounted aggregate value from a state is the same for every profile.

Theorem 5.1 (Discounted potential criterion)**.** *The discounted aggregate value is independent of the behavioral profile from every state if and only if there is a bounded $\phi : S \to \mathbb{R}$ such that*

$$c(s,a) = \phi(s) - \delta P_a \phi(s) \quad \text{for every } s, a. \tag{5.1}$$

The function ϕ is unique and equals the common infinite value. When (5.1) holds, every finite-horizon aggregate is

$$\mathsf{sum} V_N^\pi(s) = \phi(s) - \delta^N \mathbb{E}_s^\pi[\phi(s_N)]. \tag{5.2}$$

Proof. If all profiles have common value $\phi(s)$ at s, forcing action a first and then using a deterministic baseline gives $\phi(s) = c(s,a) + \delta P_a \phi(s)$. The tail bound makes ϕ bounded. Conversely, assume (5.1). At each history and joint action the conditional expectation of $\delta^{t+1}\phi(s_{t+1})$ is the term $\delta^{t+1} P_a \phi(s_t)$ in the discounted reward. After taking expectations, consecutive potential terms cancel. This gives (5.2), whose second term tends uniformly to zero. Every solution is therefore the common infinite value. □

The formula is related to potential-based reward shaping in Markov decision processes [7]. The shaping term used there is a discounted difference of successive potentials. Here the equality determines whether the entire aggregate payoff is policy invariant for a fixed controlled system. The equation is imposed on expected stage rewards, so it permits realized reward variation around the specified conditional expectation.

For a scalar arrow $f = (n, P, r) : X \to Y$, choose a bounded field ϕ_X on each object and put

$$r_f^\phi = r - \phi_X + \delta^n P^* \phi_Y.$$

This defines an automorphism of the scalar category fixing grades and kernels. Indeed, when composing, the Y potential appears once positively and once negatively and cancels, leaving the endpoint correction for the composite. The inverse uses $-\phi$. At the level of continuation maps, write $T_X(v) = v + \phi_X$. Then

$$(f^{\phi})^{\sharp} = T_X^{-1} \circ f^{\sharp} \circ T_Y.$$

Thus the automorphism is realized by a coherent change of payoff origins across interfaces. When (5.1) holds, this change makes every one-step aggregate reward zero. Evaluating the original arrows at zero and evaluating the transformed arrows at zero correspond to different terminal choices. The correction in (5.2) accounts for that difference exactly.

5.2 A single resolvent and a family of forecasts

Under (4.1), the baseline value is

$$\phi_b = (\mathrm{id} - \delta P_b)^{-1} c = \sum_{j \geq 0} \delta^j P_b^j c.$$

This follows by summing the geometric series of bounded operators. Since $\phi_b = c + \delta P_b \phi_b$, equation (5.1) for all actions is equivalent to $L_{s,a}\phi_b = 0$ for every s, a. At a fixed discount, a single observable is enough. Finite-horizon invariance instead tests the whole span $\mathcal{K}_b(c)$. When $c(s, a)$ also depends on the action, the baseline still gives a unique candidate $\phi_b = (\mathrm{id} - \delta P_b)^{-1} c_b$ with $c_b(s) = c(s, b(s))$. The exact test is then

$$c(s, a) - c_b(s) + \delta L_{s,a}\phi_b = 0$$

for every action and state.

Corollary 5.2 (Invariance across discount factors)**.** *Under* (4.1)*, if infinite aggregate values are policy invariant from every state for discount factors in a subset of* $(0, 1)$ *with an accumulation point in* $(0, 1)$*, then aggregate values are policy invariant at every finite horizon for every discount in* $(0, 1)$*.*

Proof. For each s, a the analytic function

$$F_{s,a}(z) = \sum_{j \geq 0} z^j L_{s,a} P_b^j c$$

converges for $|z| < 1$ since the coefficients are bounded. The hypothesis and the potential criterion give zeros at the specified discount factors. The identity theorem

makes each $F_{s,a}$ identically zero, so all its coefficients vanish. Theorem 4.1 gives the conclusion. □

On a finite state set the contrast is a rational function in δ, by the matrix inverse formula. If it is nonzero as a function, its zeros in $(0,1)$ are finite. Consequently, infinite-horizon policy invariance at exceptional isolated discounts can coexist with finite-horizon variation. The algebraic condition uses an entire interval of possible choices of δ, whereas Theorem 5.1 fixes one discount.

5.3 Boundary examples

Fix a discount $\delta_0 \in (0,1)$. The starting state pays zero in the aggregate and has two actions. The first leads to a state paying 1 next period, then to a state paying $-1/\delta_0$ at the following period, and then to a zero absorbing state. The second leads directly to that absorbing state. All noninitial states have one available joint action. From the initial state the aggregate difference at horizon two is δ_0 and at horizon three it is zero when the discount is δ_0. The infinite difference at a general discount is

$$\delta - \frac{\delta^2}{\delta_0}.$$

It is positive below δ_0 and negative above it. Here increasing patience can reverse the ordering because a gain precedes a later loss. The positive-tail condition used later excludes this reversal.

There is a second source of terminal effects when the closure assigns nonzero value to unfinished work. Suppose a program builds an asset and pays no dividend before date N. A zero terminal field can rank it below an immediate transfer even when a date-N liquidation value gives the opposite ranking. The asset and liquidation value should be represented separately from the flow, particularly when the asset is imperfectly marketable. Equation (5.2) gives an exact mathematical case in which a terminal correction restores invariance at all horizons. For real applications the appropriate correction depends on which payments and rights are included in the model.

5.4 Average payoffs and patience

The discounted value diverges as δ tends to one when the long-run mean flow is positive. Its normalized version is $(1-\delta)V_\infty^\pi$. Under a fixed stationary Markov policy

on a finite state set, the mean of expected per-period rewards has a limit

$$\bar{r}^{\pi}(s) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} (P_{\pi}^{*})^{t} r_{\pi}(s).$$

The limit follows from the decomposition into transient and recurrent classes, or the finite matrix mean ergodic theorem. Recurrent classes can give different limiting payoffs. The value at a transient state is a mixture using its probabilities of entering those classes. Periodic state distributions can fail to converge, while their averages still converge.

For any bounded sequence u_t with convergent Cesaro mean $\bar{u}$, Abel summation gives

$$(1-\delta) \sum_{t\geq 0} \delta^{t} u_t = (1-\delta)^2 \sum_{n\geq 1} \delta^{n-1} \sum_{t<n} u_t \;\longrightarrow\; \bar{u}.$$

To prove the limit, write the right side as a weighted mean of the prefix averages with weights $(1-\delta)^2 n\delta^{n-1}$. Their sum is one, and the total weight of any fixed finite prefix tends to zero. This applies componentwise to $\bar{r}^{\pi}$. A positive average difference implies a positive discounted comparison for discounts sufficiently near one. The converse can fail when the profiles differ by a finite windfall but have the same average. In the essay's examples a persistent increase in the size of the market and a one-time retention opportunity accordingly have different asymptotic signatures.

This limit concerns fixed profiles. It does not exchange the limit with optimization, nor assert convergence of equilibrium strategies. Stochastic-game folk theorems concern strategic payoff sets under additional hypotheses on the game [9, 11]. The ability to change the recurrent class is economically central and can also invalidate irreducibility assumptions used in those results.

6 Crossover and sensitivity to the horizon

Fix two full profiles π and ρ from the same initial history or law, and let $d_t = \mathbb{E}^{\pi}[r(s_t, a_t)] - \mathbb{E}^{\rho}[r(s_t, a_t)]$. Their payoff difference and aggregate difference are

$$D_N = V_N^{\pi} - V_N^{\rho} = \sum_{t<N} \delta^t d_t, \qquad C_N = \mathsf{sum} D_N.$$

Both are measured in units of date zero. We use ρ as the baseline. An aggregate improvement is $C_N > 0$. A strict Pareto improvement is $D_{N,i} > 0$ for each i. Fixing the profiles allows the observation horizon to vary without changing behavior before the horizon. When an announced deadline changes behavior, each deadline instead induces a new family of profiles.

Proposition 6.1 (Finite crossover and persistence)**.** *Let b_t be a coordinate of d_t or its aggregate, and put $H_N = \sum_{t<N} \delta^t b_t$.*

(a) If $H_1 \leq 0$ and $b_t \geq 0$ for $t \geq 1$, then a strict positive crossing occurs at some finite horizon if and only if $\sum_{t\geq 1} \delta^t b_t > -b_0$. Every horizon after a strict crossing is positive.

(b) If $|b_t| \leq B$, then $|H_\infty - H_N| \leq B\delta^N/(1-\delta)$. A strictly positive discounted limit gives positivity at all sufficiently large finite horizons. A certificate $H_N > B\delta^N/(1-\delta)$ gives positivity at every horizon $L \geq N$, including the infinite one.

Proof. In (a) the partial sums are nondecreasing from $N = 1$ and converge. A positive partial sum forces a positive limit, and a positive limit forces a positive partial sum. In (b) the difference between any two sums starting at horizon N is bounded by the indicated geometric tail. Taking the limit gives the statement about H_∞. □

With $\|r\|_\infty \leq M$, one can take $B = 2M$ for a coordinate or $B = 2|I|M$ for the sum. If every coordinate of D_∞ is positive, choose the smallest positive limiting margin over the finite player set and make the common tail bound smaller than that margin. This produces one horizon valid for every player. A strict aggregate crossing can occur when some coordinates never cross.

The positive-tail hypothesis is sufficient for a permanent crossover from the first strict gain. Without it, early and late horizons can disagree repeatedly. A nonzero limit nevertheless controls the eventual sign. At a zero limit, the value of a finite prefix is minus the discounted value of its remaining tail, and its sign can alternate. For instance,

prescribe $H_N = (-\delta)^{N-1}$ for $N \geq 1$. Then $b_0 = 1$, and $b_t = (H_{t+1} - H_t)/\delta^t = (-1)^t(1 + 1/\delta)$ for $t \geq 1$, a bounded sequence. Its discounted limit is zero and it changes sign at every finite extension.

6.1 Persistent per-period differences

For an immediate loss k_i and constant later gain $w_i \geq 0$, write $d_0 = -k$ and $d_t = w$ for $t \geq 1$. Then

$$G_N(\delta) = \sum_{t=1}^{N-1} \delta^t = \frac{\delta(1 - \delta^{N-1})}{1 - \delta}, \qquad D_N = -k + G_N(\delta)w. \tag{6.1}$$

When $k_i, w_i > 0$, a finite strict gain exists precisely when $\delta > k_i/(k_i + w_i)$. Above that threshold, define $\eta_i = 1 - k_i(1 - \delta)/(\delta w_i) \in (0, 1)$. The first gain occurs at the smallest integer N with

$$N - 1 > \frac{\log \eta_i}{\log \delta}.$$

The strict inequality matters when the ratio is an integer. At the critical discount, the infinite values tie and every finite horizon remains unfavorable to player i. If $k_i = 0$ and $w_i > 0$, the first gain is at horizon two. If $w_i = 0$ and $k_i > 0$, there is no crossing. If $k_i < 0$, the immediate benefit already gives positivity under the nonnegative tail assumption. These edge cases distinguish a cash-flow problem from a case in which the modeled mechanism delivers no value to a participant.

The aggregate expression uses $\sum_i k_i$ and $\sum_i w_i$. When the first is zero and the second positive, horizon two gives the first aggregate gain, even if a player needs many additional dates to recover their initial concession. With a common discount δ, the threshold depends on the ratio of an immediate loss to a per-date benefit, not on the absolute magnitude of payoffs.

If the difference matures only after date $L \geq 1$, replace G_N by $\delta^L(1 - \delta^{N-L})/(1 - \delta)$ for $N > L$ and by zero for $N \leq L$. A longer maturation delay decreases the infinite discounted value by a factor δ per added date. In a model in which dates have equal calendar duration $\tau > 0$ and a discount rate $\lambda > 0$ is specified per unit calendar time, $\delta = e^{-\lambda\tau}$. An assertion that benefits arrive within a decade therefore requires a choice of stage duration and a model of the delay. The discount threshold alone determines neither.

6.2 A single reversal in marginal gains

Some models have a transition from losing increments to winning increments, with no subsequent reversal. Suppose there is $T \geq 0$ with $b_t \leq 0$ for $t < T$ and $b_t \geq 0$ for $t \geq T$. A strict finite crossing then exists if and only if the infinite value is strictly positive, provided $H_0 = 0$. Before T the sums are nonpositive, and after T they are nondecreasing. If $H_\infty = 0$, every finite value is nonpositive. This generalizes the first part of Proposition 6.1; it will be useful when a better relationship needs to survive several periods before its chance of continuing offsets a worse initial split.

A similar comparison holds as patience varies while the increments are fixed. For $\delta' > \delta > 0$ and the same turning point T, the signed differences in $\sum_t (\delta'/\delta)^{t-T}\delta^t b_t$ weakly improve term by term: at an early negative increment its multiplier is less than or equal to one, and at a later positive increment the multiplier is at least one. Multiplying the resulting sum by $(\delta'/\delta)^T$ yields the value at δ'. If the sum was strictly positive at δ, it remains strictly positive at every larger discount below one. Profiles which themselves vary with discount need not satisfy this comparison, even if their values at one discount do.

6.3 Observable margins and incomplete tails

Suppose a finite comparison is based on known payoff differences through date $N-1$, with only the bound $|b_t| \leq B$ thereafter. The interval of possible infinite values is contained in

$$\left[H_N - \frac{B\delta^N}{1-\delta},\ H_N + \frac{B\delta^N}{1-\delta}\right].$$

If no restriction besides the bound is imposed on future differences, the interval is sharp: choose the remaining increments to be any fixed value in $[-B, B]$. If the model implies $b_t \geq 0$ thereafter, replace the lower endpoint by H_N. A more informative model may restrict the expected tail through transitions, contracts, or accumulation of capacity. That information enters the valuation as a continuation field rather than as a change of definition of surplus.

Suppose instead the observed prefixes give estimates $\widehat{b}_t$ with $|\widehat{b}_t - b_t| \leq \varepsilon_t$. Then the bound for an estimated prefix $\widehat{H}_N$ becomes

$$|H_\infty - \widehat{H}_N| \leq \sum_{t<N} \delta^t \varepsilon_t + \frac{B\delta^N}{1-\delta}.$$

Adding dates tightens the second term but can increase the first term. When both are significant, a longer reporting window need not strengthen a sign certificate.

The inequality is deterministic conditional on the error bounds. It is distinct from a statistical confidence statement about the data that produced those bounds.

6.4 Sensitivity to the unit of patience

For fixed increments with $|b_t| \leq B$, the infinite comparison is differentiable as a function of the discount on $(0, 1)$, by locally uniform convergence of the derivative series. For $0 < \delta, \delta' \leq d < 1$,

$$|H_\infty(\delta') - H_\infty(\delta)| \leq |\delta' - \delta| \sum_{t \geq 1} t d^{t-1} B = \frac{B|\delta' - \delta|}{(1-d)^2}.$$

The proof bounds each difference of powers using the mean-value theorem before summing. A sign conclusion with a margin exceeding this bound survives the specified uncertainty in discount. This is uniform only on intervals bounded away from one. With known positive later increments, H_∞ is increasing in the discount, so one can instead evaluate at the lower end of the interval to certify a positive sign. The appropriate bound depends on whether the mechanism produces signed or persistent tail effects.

7 From feasible cooperation to equilibrium

7.1 A one-step deviation theorem

Let π be a behavioral profile with bounded rewards and public histories. For each player i, history h, and pure alternative a_i' available at h, let $Q_i(h, a_i'; \pi)$ be the expected discounted value obtained by changing only that player's present action and following π from the next date onward. Let W^π be the continuation field of π, defined at every feasible history.

Theorem 7.1 (One-step incentive criterion)**.** *The profile π is subgame perfect if and only if*

$$W_i^\pi(h) \geq Q_i(h, a_i'; \pi) \quad \text{for every } i,\ h,\ a_i'. \tag{7.1}$$

This applies both to a finite game and to the discounted infinite game, using the relevant continuation field in each case.

Proof. Necessity follows by considering a single deviation in the indicated subgame. For sufficiency, fix opponents' strategies and a subgame. The inequality for pure alternatives extends to randomized choices by linearity. A deviation restricted to the first L dates of the subgame can be removed backward. At any history at its last date, switching to π_i cannot lower the deviator's conditional value, since after that date the continuation is π. The conclusion remains true after taking expectations at the previous dates. Repeating the argument eliminates every changed date and proves the finite assertion. An arbitrary infinite deviation agrees with its truncated version through date $L-1$, and the difference of their later values is bounded by $2M\delta^L/(1-\delta)$. Letting L tend to infinity rules out a strict gain. □

Let $r_i^\pi(h)$ be the expected one-step reward under the candidate and $r_i^{\text{dev}}(h)$ the corresponding reward under a one-step deviation. If $g_i = r_i^{\text{dev}}(h) - r_i^\pi(h)$ is the current temptation, (7.1) reads

$$g_i \leq \delta\big(P_\pi^* - P_{\text{dev}}^*\big)W_i^\pi(h). \tag{7.2}$$

The state for the kernels can be a public history. Then the contrast records changed physical transitions and changed reactions to the revealed action. An aggregate

continuation difference can have the opposite sign from this individual continuation difference. An equilibrium proof requires the latter at all histories, including histories created by a deviation.

There is a useful approximate form. If the right side of (7.1) exceeds the left side by at most ε at each history and for each deviation, any L-date deviation gains at most $\sum_{t<L} \delta^t \varepsilon$. The backward removal argument now charges at most ε at the current date before discounting. Thus an infinite deviation gains at most $\varepsilon/(1-\delta)$. This statement controls a subgame-perfect approximation uniformly. Checking small deviation gaps only along a realized path does not yield that conclusion.

7.2 Stationary incentives and public contingencies

For a stationary Markov candidate π, the continuation $W^{\pi}(s)$ solves (3.3). A one-step deviation at any history with state s has the same comparison as at every other such history. The criterion can then be checked as a statewise inequality for the stationary candidate, and it still rules out history-dependent deviations. The reduction is valid because the opponents' prescribed responses and all physical transitions depend on the current state. A candidate with a punishment phase must either use the history or put the phase in the state.

In a repeated prisoner's dilemma, let mutual cooperation pay $(2,2)$, mutual defection pay $(1,1)$, and unilateral defection pay $(3,0)$ or $(0,3)$. The always-cooperate and always-defect paths have a strict Pareto ordering at every positive horizon. Always cooperating fails (7.1) at any discount, because a unilateral defection yields a current gain of 1 and leaves the opponents' future behavior unchanged. In the infinite game, cooperation until the first defection followed by permanent mutual defection is subgame perfect precisely when

$$\frac{2}{1-\delta} \;\geq\; 3 + \frac{\delta}{1-\delta}, \quad \text{or equivalently} \quad \delta \geq \frac{1}{2}.$$

In the punishment phase defection is a best response to defection. With a publicly known finite terminal date, backward induction instead gives mutual defection at every date. The favorable ordering of the two stationary paths persists as a counterfactual comparison. The finite strategic conclusion changes because the continuation incentives disappear at the end.

Repeated-game folk theorems, including Fudenberg and Maskin [8], concern which payoff vectors can be supported by some equilibrium at sufficiently patient discounts under their hypotheses. Stochastic versions allow physical consequences of actions and can require additional control, recurrence, or monitoring conditions

[9, 11]. The one-step theorem here checks a supplied strategy. It neither presumes that an equilibrium implementing a given gain exists nor characterizes the set of all equilibrium values.

7.3 Making a transfer feasible

Assume the payoff coordinates share a transferable unit and side payments can be committed at date zero. Let $C_N = \sum_i D_{N,i}$. For $n = |I|$ and $C_N > 0$, the zero-sum transfer

$$t_i = C_N/n - D_{N,i}$$

leaves each player a gain C_N/n. Conversely, summing strict positive gains after a zero-sum transfer proves $C_N > 0$. For two players, a transfer x to player one from player two gives strict improvement exactly for

$$-D_{N,1} < x < D_{N,2}. \tag{7.3}$$

A feasible set of transfers restricts this open interval. A firm that can pay no more than $\bar{x}$ before collecting its revenue may be unable to implement the transfer even when the net value would cover it.

Transfers may also be contingent on a future event. Under commitment and a common discount, a transfer of x at date t conditional on an event of probability p is worth $\delta^t p x$ in present value. If the only admissible event has $p = 0$ under the intended profile, no transfer on that event changes participation values. If one party can refuse the transfer when the event arrives, its one-step deviation constraint includes the temptation to withhold it. There must then be a continuation or a mechanism making compliance privately valuable. The analysis of relational contracts by Baker, Gibbons, and Murphy [19] studies this dependence of sustainable arrangements on the gains from reneging and the value of future relationships.

7.4 An impossibility from one-sided capture

Take a two-player state transition which generates a permanent benefit $w > 0$ to player one and no later benefit to player two. Entering it requires player two to spend $k > 0$ in the common payoff unit. For $\delta w/(1-\delta) > k$ the move has a positive aggregate discounted return. Player two has a loss at every horizon. If player two controls the irreversible decision, transfer promises cannot be committed, and all later decisions by player one affect only that player’s own current money, the promise of future compensation is not credible in the modeled game. In every continuation after entry,

refusing a positive promised payment is strictly profitable for player one. Anticipating this, player two declines entry. An enforceable transfer or a continuation in which player two can withdraw valuable future cooperation changes the game.

The example separates three levels of analysis. A positive aggregate shows that value exists in the modeled surplus boundary. A feasible commitment technology may allocate it so that participation is beneficial. A strategic protocol additionally has to survive each player's subsequent opportunities to deviate. The implications run through those assumptions, and a larger discount factor cannot compensate for an absent incentive channel.

7.5 Participation at intermediate histories

A positive gain at date zero can coexist with a later incentive to terminate. Suppose both players gain initially from joining a project and one learns at date one that its continuation has negative value relative to an immediate outside option. If withdrawal is available there, the outside option is a one-step deviation with its own successor state. The inequalities at date zero do not exclude it. An agreement feasible under commitment can therefore fail when participation is voluntary at each date. The model should put termination and its payoff into the action set at every history where it is available.

If the withdrawal action gives current payoff $o_i(h)$ and transition law P_{exit}, the relevant participation inequality is

$$W_i^{\pi}(h) \geq o_i(h) + \delta P_{\text{exit}}^{*} W_i^{\pi}(h).$$

This compares the candidate's remaining value with the actual outside path at that history. When a relationship changes the participant's future outside opportunities, using a fixed zero outside payoff throughout can change which agreements survive.

8 Welfare boundaries, coalitions, and discount conventions

8.1 A family of linear aggregations

For fixed real weights ω_i, define $\mathsf{A}_\omega(v) = \sum_i \omega_i v_i$. The same proof as (2.3) makes this aggregation natural with respect to affine continuation when the scalar arrow is assigned reward $\mathsf{A}_\omega(r)$. Positivity of the weights is required if coordinatewise improvement is to imply weak aggregate improvement. If every $\omega_i > 0$, a weak Pareto improvement with a strict gain for at least one player yields strict weighted aggregate gain. A weighted aggregate gain can still coexist with losses to individual players.

This formal freedom does not select a normative rule. An unweighted sum of market revenues uses a different welfare boundary than a sum of consumer and producer surpluses. If an allocation stage merely transfers wealth between two participants, its sum is unchanged in common monetary units. If the same transfer is weighted by ω, its contribution is $(\omega_1 - \omega_2)x$. Weighting can intentionally value one recipient's dollar differently, but the constant-sum premise must then be tested again with the declared weights.

Positive affine transformations of each player's *total* utility preserve all coordinatewise rankings. A player-specific scaling generally changes an unweighted scalar sum. A date-independent shift of stage reward preserves fixed-horizon comparisons if the game includes the shift at every date for every path, including any absorbing states. If a shift is paid only while a relationship survives, it changes the value of survival. In that case it is an economic modification, because different policies can earn the shift a different number of times.

8.2 Heterogeneous discounting

Suppose player i has discount $\delta_i \in (0, 1)$. The category of Chapter 2 can be defined with the continuation formula

$$(f^\sharp(v))_i = r_i + \delta_i^n P^* v_i.$$

Its composition rule uses this coordinatewise discount and is associative by the equality $\delta_i^{n+m} = \delta_i^n \delta_i^m$. A scalar aggregation A_ω with one scalar discount $\bar{\delta}$ commutes with continuation for all fields and grades exactly when $\delta_i = \bar{\delta}$ for each i with $\omega_i \neq 0$. To prove necessity, use a grade-one identity kernel and a terminal field with only its ith coordinate nonzero. The equality tests $\omega_i \delta_i = \omega_i \bar{\delta}$. Sufficiency follows by factoring the common discount out of the sum.

Thus if two people value the same date differently, a zero-sum transfer at that date can have a nonzero sum of subjective present values. For instance, the stream which pays one dollar to player one and charges it to player two at date one has date-one total zero and time-zero subjective sum $\delta_1 - \delta_2$. An analyst can still compute a constant-sum social accounting measure with a chosen common discount, while treating individual incentives with each person's own discount. The two calculations answer different questions. In the heterogeneous model the one-step proof applies to player i using δ_i, and a separate social aggregate can still use the scalar results in Chapters 4 and 5.

8.3 Coalition-specific surplus

A coalition $J \subseteq I$ observes the field $\sum_{i \in J} r_i$. This gives a scalar reward category via weights $\omega_i = \mathbf{1}_{\{i \in J\}}$. A transfer between members of J disappears from that total. A transfer from outside J increases it by the amount received, even when the total for I is unchanged. Consequently two distinct claims can both hold: an initiative raises the modeled industry total, and one subgroup can improve its own payoff by diverting a transaction from another subgroup without raising the total.

If players outside J are held to a fixed policy, the remaining coalition can choose joint actions as one controller for the purpose of computing its feasible aggregate reward. The finite-horizon theorem applies to this induced controlled system whenever the resulting one-stage coalition total is independent of coalition actions at each adequate state. The potential criterion gives the corresponding infinite-horizon characterization without that stage hypothesis. Such a calculation does not imply that members of J have a mechanism to coordinate a joint deviation, that they can share the gains, or that outside players would actually hold their policies fixed following a change. A strategic coalition concept requires specifying these rights and responses.

8.4 A decomposition with excluded participants

Let I denote two sellers and let B denote buyers. A market education action might shift the expected number of later sales by $\Delta q > 0$. If m_i is the seller's incremental net

margin per additional sale and b the buyer's net surplus, the increase in per-period surplus depends on the boundary:

$$w_I = \sum_i m_i \Delta q_i, \qquad w_{I \cup B} = w_I + b \sum_i \Delta q_i$$

in a simple homogeneous buyer model. A teaching effort with cost borne by the sellers lowers w_I if the cost is ongoing, or adds an initial loss if it is one-time. If buyer surplus is negative because the action induces poorly matched purchases, the total for sellers can rise while the enlarged total falls. The sign comparison cannot resolve the empirical or normative question of which marginal buyer values are appropriate.

A contract reallocation gives another boundary example. A failed project may transfer fees from one consultant to another. The fees sum to zero in their pairwise comparison only after controlling for amounts paid by the client and costs incurred by both firms. Rework paid by the client, opportunity costs of lost time, and damage to other participants belong to separate coordinates if they are included. The consulting model in Chapter 12 will isolate two effects: a fixed current net transfer and a future difference in demand. Adding genuine current rework costs gives an aggregate loss that the same crossover inequalities can accommodate.

8.5 External effects and incentives can move separately

A policy may raise the total for I while raising no player's continuation value by enough to offset their own immediate temptation. Conversely an individual may prefer a reliable agreement without any positive aggregate effect, when reliability redistributes future transactions in that individual's favor. The first case is a gap between social and private value. The second is a gap between private value and market expansion. Equation (7.2) and Theorem 4.1 can both be applied to the same game to distinguish them.

The value-creation framework of Brandenburger and Stuart [16] makes the distinction between value created by participants and value appropriated by a particular participant explicit. Our temporal formulation adds a controlled law for which opportunities exist at later dates. Its aggregate observable should be read as conditional on the set of participants and the payoffs chosen to define creation.

9 Observation quotients and value preservation

9.1 Exact descent of a resolved process

Let $q : X \to \bar{X}$ and $p : Y \to \bar{Y}$ be surjections between state interfaces. A field on X descends along q precisely when it is constant on each fiber. Its descended value is unique. A kernel $P : X \to Y$ descends relative to the two quotient maps when its pushforward to $\bar{Y}$ depends only on $q(x)$:

$$\sum_{y:\, p(y)=\bar{y}} P(y \mid x) = \sum_{y:\, p(y)=\bar{y}} P(y \mid x') \quad \text{if } q(x) = q(x'), \quad \bar{y} \in \bar{Y}. \tag{9.1}$$

This is the block-sum condition for strong lumpability in the one-kernel case [13]. It can also be expressed by saying that P^* sends fields constant on fibers of p to fields constant on fibers of q. The implication from (9.1) follows by grouping the sums defining P^*. The converse follows by using the indicator fields of the fibers of p.

Theorem 9.1 (Descent of stochastic reward arrows)**.** *Let $f = (n, P, r) : X \to Y$. There is a unique arrow $\bar{f} = (n, \bar{P}, \bar{r}) : \bar{X} \to \bar{Y}$ for which*

$$f^\sharp p^* = q^* \bar{f}^\sharp \tag{9.2}$$

if and only if r is constant on the fibers of q and (9.1) holds. Equivalently, the square with the quotient maps as zero-reward grade-zero deterministic arrows commutes in $\mathcal{R}_{\delta,1}$.

Proof. If both conditions hold, set $\bar{r}(q(x)) = r(x)$ and set $\bar{P}(\bar{y} \mid q(x))$ to either side of (9.1). These formulas define an arrow and verify (9.2). Conversely, evaluate the identity at zero to recover $\bar{r} \circ q = r$. Subtract the zero-field identities from the identities for singleton indicator fields in a payoff coordinate. Divide by δ^n to recover the block sums. Surjectivity makes both parts of the quotient arrow unique. By (2.1) the two paths in the claimed square have respective kernels pP and $\bar{P}q$, and rewards r and $q^*\bar{r}$, giving the equivalent formulation. □

If successive arrows descend through compatible quotients, their composite descends by composing the commuting squares. The zero terminal is a descended field. Pullback therefore identifies each finite evaluation with the one made on the quotient.

Under the uniform reward bound, taking limits yields the same statement for the discounted infinite value. Aggregation also commutes with pullback, so the scalar value can be evaluated on the quotient whenever the vector arrows descend. It can sometimes descend with weaker assumptions: for scalar arrows it suffices that the aggregate stage reward descend and the same kernel condition hold, even when the allocation varies inside an observation class.

9.2 Controlled descent and strategic qualifications

In a stationary controlled model with quotient $q : S \to \bar{S}$, suppose $A_i(s)$ depends only on $q(s)$ and every fixed joint action satisfies both conditions of Theorem 9.1 on the blocks where it is available. Then there is a controlled quotient game with the same action labels. A profile defined on observable histories of states and actions lifts to the hidden state game by composition with the observation map. Given a hidden history and its public image, the lifted action distribution is the same and the conditional distribution of the next observation depends only on that image. By induction, the lifted profile and the quotient profile have the same expected vector value, including against any lifted deviation. Thus evaluation and the deviation test agree for the information structure in which every strategy uses only the quotient histories.

The implication for a game with more information has a separate argument under these strong hypotheses. Suppose all players except one use the lifted profile and the lifted candidate passes the one-step test on the quotient. A fully informed deviator can select a different action in distinct hidden states. At each hidden history, however, the continuation after a one-step change has exactly the value of the corresponding quotient test, because the next observation law, the reward, and later candidate actions descend. Thus the inequality holds for every pure action at each fully observed history, and Theorem 7.1 gives robustness of that candidate against fully informed unilateral deviations in the full public-state game. This argument uses descent of vector rewards for every feasible action and assumes no extra feasible actions at a hidden state. Value preservation for a single policy alone would be insufficient.

If (9.1) fails, a coarser observation may still be adequate when a posterior belief over the hidden state is used as an interface. The resulting update depends on the observation process and actions. In games with private information, each player's posterior and feasible deviation set can differ. We use the quotient only under the explicit exactness or approximation hypotheses and make no equilibrium claim for those more general information structures. The recursive treatment of imperfect public monitoring in Abreu, Pearce, and Stacchetti [10], and the imperfect monitoring stochastic results in [11], solve distinct incentive problems.

9.3 How weak an observable can suffice?

For the finite-horizon theorem, one only needs the space $\mathcal{K}_b(c)$, rather than the entire algebra of all state fields. Define $s \sim s'$ when $f(s) = f(s')$ for every $f \in \mathcal{K}_b(c)$. Every f in the span descends through the quotient $S/\!\sim$. Because P_b preserves the span, its pullback on those fields descends as well. This does not imply lumpability on all functions on $S/\!\sim$. The linear span may fail to contain the indicator functions of its equivalence classes.

For example, let $S = \{0, 1, 2\}$, $c(s) = s$, and let P_b be the identity. Then $\mathcal{K}_b(c)$ is spanned by c. At state 1 an alternative action can send the process to 0 and 2 with equal probabilities. This preserves $P_a c(1) = 1$ and hence all expected total values in the finite-horizon theorem. The alternative nevertheless changes the distribution of the next surplus. Against the terminal payoff $\mathbf{1}_{\{1\}}$, the two actions differ. Against a nonlinear objective applied to the next period's surplus, they can also differ. The equivalence depends on the class of observable fields.

A one-step forecast matching condition weaker than lumpability can therefore preserve every continuation value of a specific game. Exact model reduction that must work for any terminal field needs the stronger commuting-square condition. These correspond to different questions about data collection. Estimating full transition probabilities is unnecessary if only one aggregate forecast needs to be transported. Estimating only that forecast can conceal distributional risks and the effects of changing the terminal objective.

9.4 Approximate descent under bounded model error

Suppose a proposed quotient arrow $\widehat{f} = (1, \widehat{P}, \widehat{r})$ is compared to a grade-one arrow on $X \to Y$ with observation maps as above. Assume

$$\sup_x \|r(x) - \widehat{r}(q(x))\|_\infty \le \varepsilon_r, \qquad \sup_x \mathrm{TV}\Big(p_* P(- \mid x), \widehat{P}(- \mid q(x))\Big) \le \varepsilon_p.$$

For any bounded terminal v on $\bar{Y}$, the uniform deviation from (9.2) is at most

$$\varepsilon_r + \delta \varepsilon_p \max_i \operatorname{osc}(v_i). \tag{9.3}$$

The first term is the reward error and the second follows from (2.5). For a sequence of approximate quotients with rewards bounded in norm by M and zero terminal, every quotient continuation has norm at most $M/(1-\delta)$. The valuation error over N stages

is bounded by

$$\frac{1-\delta^N}{1-\delta}\left(\varepsilon_r + \frac{2\delta M \varepsilon_p}{1-\delta}\right)$$

when the local bounds hold uniformly. To prove this, compare the true continuation first with the lifted approximate continuation at the next date, incurring at most δ times the next-step error. Then apply (9.3) to the quotient's next-date field. Iteration from zero terminal gives the result. Coordinatewise sign claims with margins larger than the sum of the error bounds for the two compared profiles survive this reduction.

The error can also be used to compare two candidate policies on the same quotient. If a finite-horizon gain in a coordinate exceeds twice a common valuation error bound, that coordinate remains favorable before reduction. For the aggregate the corresponding conservative factor is twice the number of modeled players. In a reduced model that preserves the aggregate reward more accurately than the individual rewards, apply the scalar version of the estimate directly. It can give a sharper aggregate certificate without making a claim about allocation within an observation class.

10 Sheaves of horizon comparisons

10.1 Local payoff fields

Let X now be a topological space of public contexts, for example a parameter space of jointly observable market conditions. This is separate from the state set of a single game. Work in the sheaf topos $\mathrm{Sh}(X)$ with covers given by unions of opens. Let $\mathcal{C}$ be the sheaf $U \mapsto C(U, \mathbb{R})$ of continuous real fields. Its order relation $f > 0$ on U means pointwise strict positivity. A fixed-profile family gives compatible fields $d_{t,i} \in \mathcal{C}(X)$ whenever the expected date-t differences vary continuously with context. For example, finite-state reward and kernel data and fixed resolved policies that vary continuously in the context give continuous finite-horizon comparisons by finite products and sums. Define

$$D_{N,i} = \sum_{t<N} \delta^t d_{t,i}, \qquad C_N = \sum_i D_{N,i}.$$

The evaluation of any finite horizon and the summation operation commute with restriction to an open subset. If the differences are uniformly bounded on each neighborhood by a bound independent of date, the discounted limits are continuous by locally uniform convergence. This is a local boundedness assumption on the family, rather than a requirement that the bounded functions on arbitrary opens form a sheaf.

For the aggregate comparison put $U_N = \{x : C_N(x) > 0\}$. For strict Pareto comparison use $U_N^{\mathrm{P}} = \bigcap_{i \in I} \{x : D_{N,i}(x) > 0\}$. Both are open, with the latter a finite intersection. In $\mathrm{Sh}(X)$ they are subterminal objects, meaning they have at most one section over any open context. The subterminal for an open V has a section on U exactly when $U \subseteq V$. Joins and finite meets of these objects correspond to unions and finite intersections of opens.

10.2 The existential horizon and its witnesses

Let $\underline{\mathbb{N}}$ be the natural-number object, whose sections on U are locally constant functions from U to $\mathbb{N}$. We use positive N for the horizon. The relation $C_N > 0$ is the subobject $R \subseteq X \times \underline{\mathbb{N}}$ specified on a constant section N by the open U_N. Here X denotes the terminal sheaf of the topos in the product notation. For a variable locally constant section n on U, membership means that each point $x \in U$ lies in $U_{n(x)}$. This definition

is compatible with local restriction and gluing. The image of the projection $R \to X$ is the union $\bigcup_N U_N$, expressing the internal predicate $\exists N\ (C_N > 0)$.

In Kripke–Joyal semantics [15], this predicate holds over U precisely when there is an open cover $\{U_\alpha\}$ of U and a horizon witness locally on each U_α. Equivalently,

$$U \Vdash \exists N\ (C_N > 0) \quad \Longleftrightarrow \quad U \subseteq \bigcup_N U_N. \tag{10.1}$$

For the forward direction, refine the witness cover so the locally constant horizon is constant on each piece. For the reverse direction take the cover $\{U \cap U_N\}$ and use the constant horizon N there. The existence of this cover does not assert a single constant witness. It may also fail to give a global locally constant witness when the local witnesses cannot be made compatible on overlaps. This distinction is the difference between the local existential image and a global section of the relation.

For an explicit illustration, consider $X = (-1, 1)$ and strict positivity sets $U_1 = (-1, 1/2)$, $U_2 = (-1/2, 1)$ with all later sets empty. Their union is X, but neither covers X. Any locally constant function $X \to \mathbb{N}$ is constant because X is connected, so a global witness does not exist. These opens can be realized by continuous comparisons $C_1 = 1/2 - x$, $C_2 = x + 1/2$, $C_N = 0$ for $N \geq 3$. Choose discounted increments by differences of successive C_N. The chosen fields have bounded increments on X for any fixed $\delta > 0$. This example concerns a crossover at some horizon, without persistence.

10.3 Uniform horizons under persistence

Suppose $C_N(x) > 0$ implies $C_L(x) > 0$ for every $L \geq N$ and every x in the comparison domain. Then the opens U_N are increasing. This hypothesis holds, for instance, under the pointwise nonnegative tail condition of Proposition 6.1. The same implication holds for strict Pareto opens if the coordinate comparisons satisfy persistence.

Theorem 10.1 (Uniformization on compact contexts)**.** *Let $K \subseteq X$ be compact and suppose the positivity opens are increasing. If each $x \in K$ has a positive comparison at some finite horizon, there is a single finite N for which $C_L(x) > 0$ at every $x \in K$ and every $L \geq N$. The same assertion holds for strict Pareto improvement with increasing strict Pareto opens.*

Proof. The positivity opens cover K. Choose a finite subcover and take the largest index among its members. By monotonicity the open at that index already covers K, and every later open contains it. □

Compactness alone does not suffice for a uniform horizon when the opens are not increasing. A finite subcover then provides a finite menu of horizons, with

each context possibly requiring a different choice. The sign margin bound gives a separate route to uniform eventual improvement without a monotone sequence. If $\inf_{x\in K} C_\infty(x) > 0$ and the tail bound is uniform on K, choose N making the tail smaller than that infimum. If C_∞ is continuous and strictly positive on compact K, its minimum on K is positive and supplies the first condition. The argument applies coordinatewise to strict Pareto comparisons with finitely many coordinates.

To see why noncompact domains are different even with bounded rewards, take $X = \mathbb{N}_{\geq 1}$ with the discrete topology. At context m let the two profiles differ by an aggregate payment 1 only at date m. This can be implemented by the delayed-payment chain of Chapter 4. The positivity opens are $U_N = \{m : m < N\}$. Every context admits a strict and persistent crossover and the internal existential is valid. A single finite horizon fails at all sufficiently large m. There is a global locally constant witness $n(m) = m + 1$, which is unbounded. The infinite gain is $\delta^m > 0$ pointwise and its infimum is zero, consistent with the quantitative criterion.

10.4 Refinement and what it certifies

If a continuous map $g : Y \to X$ refines or restricts contexts, pullback of the actual fields gives $C_N \circ g$. Their positivity opens are $g^{-1}(U_N)$, and g^{-1} commutes with unions and finite intersections. Thus a local certificate pulls back to the refined domain. For a surjective g, if a fixed N is positive on all of Y then it is positive on all of X, since every context has a preimage. This statement uses pulled-back fields. It does not recover value fields on a hidden state space from arbitrary data upstairs. That is the separate issue of descent in Chapter 9.

The same construction can represent incentive margins. For a fixed finite comparison of deviations, each strict margin defines an open. For an unbounded collection of histories the intersection of all strict-margin regions need not be open. Its interior is the largest open on which every listed strict inequality holds pointwise. Consequently, verifying the one-step inequalities separately in neighborhoods that shrink with the history need not give one neighborhood certifying all of them. A uniform positive margin together with a uniform continuity or error bound can produce such a neighborhood. The distinction concerns the topology of the evidence and is independent of the strategic proof of Theorem 7.1.

The sheaf argument separates three ways to state a long-horizon claim. A pointwise statement names a date for each context. A local statement says a fixed date works throughout a sufficiently small neighborhood of each context, which follows here from strictness and continuity. A uniform statement names one date valid on the entire

domain, and needs additional assumptions such as those of Theorem 10.1. None of these assertions alone identifies the actual transition parameters of an industry.

10.5 Comparing gains in all components

The logical order of quantifiers matters when players need different horizons. Without persistence, every player may gain at some date while there is no date at which they all gain. In the sheaf lattice, the first assertion uses

$$\bigcap_{i \in I} \bigcup_{N} \{D_{N,i} > 0\},$$

while the second uses $\bigcup_N \bigcap_{i \in I} \{D_{N,i} > 0\}$. The second open is contained in the first. For two players with $D_1 = (1, -1)$, $D_2 = (-1, 1)$ and all later differences zero, the inclusion is strict even over a one-point space. If positivity persists for every coordinate, the two opens coincide: at a point in the first open choose a horizon for each coordinate and take the maximum. The compactness result then concerns whether that pointwise common horizon can be made uniform across contexts. Finiteness of the player set resolves one quantifier exchange. Compactness and persistence resolve a different one.

11 Authorship and demand creation

11.1 A fixed sale and an expandable market

In the authorship example of Dellanna [1], a reader deciding which book to buy next creates a local allocation problem. At that moment, suppose the purchase generates a fixed net amount B that is divided between two authors, for example $(B, 0)$ or $(0, B)$. This is a constant-sum comparison among those allocations conditional on a sale and within the modeled author boundary. When a reader can buy neither book, that option can have a different total even at the same date. The formal premise therefore conditions on a class of comparable current transactions. An education initiative can subsequently change the number of future purchases even when the initial purchase is held fixed.

To formalize this, use an initial state s_0 with current baseline split (b_1, b_2). The baseline leads to an absorbing low-awareness state L, paying (ℓ_1, ℓ_2) each later period. A market education program leads to an absorbing high-awareness state H with probability $p \in [0, 1]$, and to L otherwise. At H, later rewards are $(\ell_1 + w_1, \ell_2 + w_2)$, where $w_i \geq 0$. In the baseline both authors keep the initial split. In the program author one transfers $d \geq 0$ of the current reward to author two, for instance as a share of current receipts required by the agreement. This is a pure transfer when both authors are in the accounting boundary. The initial stage aggregate is $b_1 + b_2$ in both cases.

The difference of expected payoffs between the program and the baseline at horizon N is

$$D_N = (-d, d) + pG_N(\delta)(w_1, w_2). \tag{11.1}$$

The common continuation after the initial choice uses the same actions in each realized state. The difference in aggregate value is the initial kernel contrast applied to the later flow, as in (2.4). If $p > 0$ and $w_1 + w_2 > 0$, the total first differs at horizon two. When $d > 0$ and $pw_1 > 0$, author one strictly prefers the program at some finite horizon precisely when $\delta > d/(d + pw_1)$. Author two gains immediately when $d > 0$ and gains later whenever $pw_2 > 0$. If $d = 0$, strict Pareto improvement at horizon two requires both $pw_i > 0$.

The interpretation of the initial concession matters. If author one pays an external provider rather than author two, the initial difference is $(-d, 0)$. Then even the aggregate initially loses d. A later aggregate gain exists exactly when $p(w_1 + w_2)\delta/(1 - \delta) > d$. When one author captures almost all incremental sales, the aggregate threshold and

each player's threshold can be far apart. This possibility is compatible with a joint expansion of the market. It is a question of how the increment is allocated.

11.2 Joint provision and free riding

Suppose education succeeds only if both authors contribute. At the initial date each chooses E or A (abstain). Choosing E costs author i a real amount $k_i > 0$ immediately. The current net sale split (b_1, b_2) is unchanged before costs. The action (E, E) leads to high awareness with probability p, and every other joint action leads to low awareness. After this initial decision there are no further actions and rewards are the flows given above. The improvement of joint contribution over mutual abstention is

$$D_{N,i} = -k_i + pG_N(\delta)w_i. \tag{11.2}$$

The comparison includes real costs, so its current aggregate is negative. For a fixed horizon, (A, A) is a strict Nash equilibrium at the initial decision because a unilateral contributor pays a cost with no increase in success probability. The profile (E, E) is also an equilibrium exactly when $pG_N(\delta)w_i \geq k_i$ for both authors. Against a contributing opponent, unilateral abstention saves the current cost and removes all of the continuation gain, giving that inequality. In this simple model it also gives strict individual participation relative to mutual abstention when the inequality is strict.

This distinction gives two different limitations of an appeal to patience. A higher discount can make contribution individually worthwhile conditional on the other author's contribution. It does not remove the strict equilibrium of mutual abstention. Selecting an outcome still depends on how expectations and commitments about joint action are formed. The two payoff profiles may look similar in an aggregate report, but the test for a unilateral move needs the interaction between both actions and success.

More generally, write p_{ab} for the probability of high awareness after actions $a, b \in \{\mathrm{E}, \mathrm{A}\}$. If both are intended to contribute, the incentive condition for author one is

$$k_1 \leq (p_{EE} - p_{AE})\, G_N(\delta)w_1,$$

and the corresponding condition for author two uses $p_{EE} - p_{EA}$. The joint comparison against (A, A) instead uses $p_{EE} - p_{AA}$ in both coordinates. The total increase in success probability can be large while each author's marginal contribution at the intended profile is small. For example, if a single contribution already reaches the entire audience, then $p_{EE} = p_{EA} = p_{AE} > p_{AA}$. Each author would rather leave the cost to the other at a joint-contribution profile, regardless of the horizon. The game

then concerns the provision and allocation of effort in addition to the value of demand creation.

11.3 Persistence, maturation, and saturation

An absorbing high-awareness state assumes the demand effect persists. A model with decay replaces this by a survival probability $q \in [0,1]$ for high awareness at each subsequent date, with low awareness absorbing. Starting at H at date one, the chance of remaining there at date $t \geq 1$ is q^{t-1}. In (11.1) replace $G_N(\delta)$ by

$$B_N(\delta, q) = \sum_{t=1}^{N-1} \delta^t q^{t-1} = \frac{\delta(1-(\delta q)^{N-1})}{1-\delta q}. \tag{11.3}$$

The discounted infinite gain to author one is $-d + pw_1\delta/(1-\delta q)$. If $q < 1$, this converges as δ tends to one to $-d + pw_1/(1-q)$. Thus even nearly undiscounted evaluation may leave the author worse off when demand decays too quickly. A larger horizon cannot replace persistence of the effect. If the initiative takes several periods to mature, additional powers of δ reduce its value as in Chapter 6.

Market saturation can also change the sign of the later increments. If an initiative mainly moves purchases from a later date to an earlier date, the number of buyers over both dates may be unchanged. At common units and without discount, this is a pure timing shift. Under discount, accelerating a positive sale increases present value, but the normalized long-run average still ignores a finite timing shift. If educational effort expands the repeat-purchase rate, the later increase can be persistent instead. These are different transition mechanisms. The total at one cutoff cannot distinguish them without more information about the state or expected tail.

11.4 Where the analogy has empirical content

The algebra identifies the data that would bear on the comparison: the baseline purchase path, the success probability of the education action, the incremental net receipts by author, and how long the new demand lasts. The claim that other authors sometimes help create a market is represented by a positive probability shift toward states with a higher aggregate continuation. An observed increase in sales following an initiative does not by itself establish that contrast. The comparison uses a common initial population and asks what would have happened there under the baseline.

The model also allows both mechanisms to act simultaneously. In the high-awareness state, the authors can still choose prices or titles that affect their split.

If those actions change current shares while preserving the measured current total and the future forecast span, the finite-horizon invariance theorem applies within that state for those choices. The same authors can therefore face a genuine allocation contest inside a larger process whose total depends on earlier choices.

11.5 Demand growth with capacity constraints

A high-awareness state can generate requests that an author cannot fulfill at the same net margin. To incorporate this, let extra demand for author i be $q_i \geq 0$, let capacity for additional sales be $m_i \geq 0$, and let the net margin on a fulfilled additional sale be γ_i. In the simplest static capacity model the per-period increment is $w_i = \gamma_i \min(q_i, m_i)$. When an author is already at capacity, additional interest can raise the other author's payoff and leave its own unchanged. Formula (11.1) then gives a positive total with no offset for the capacity-constrained author's initial concession.

If capacity can be increased by a later investment, high awareness can also change the feasible choices at that later state. Its value then includes the continuation of the investment decision, rather than a fixed per-period reward. The stochastic arrow still pulls this continuation back to the initial period. When investment is chosen strategically, its individual value and funding must be evaluated in the expanded game. This distinguishes increased demand from the ability to convert that demand into compatible gains for both authors.

12 Consulting and shared reputation

12.1 Replacing a competitor and shrinking demand

Dellanna's consulting anecdote [1] distinguishes a firm's prospect of taking over a particular engagement from the perception of consulting among later clients. Let the ordinary period in a healthy market H pay (b_1, b_2) to the modeled consultants. In a low-trust state L, let the period pay $(b_1 - \ell_1, b_2 - \ell_2)$ with $0 \leq \ell_i \leq b_i$. Assume these flows are in net receipts in a common unit. In a focal initial comparison, a reliable project keeps the market healthy and yields (b_1, b_2). A failed project reallocates $g > 0$ of the current receipts toward consultant one and away from consultant two, giving $(b_1 + g, b_2 - g)$. It also sends the market to L with probability p. Holding the current modeled total fixed in the two scenarios isolates the effect in the essay.

This is a comparison of scenarios, conditional on specified laws for market perception. It does not give one consultant control over a competitor's failure. Such control, or a joint intervention that changes the failure probability, would need to be supplied as an action in a strategic version. For the base calculation, both market states are absorbing after the focal period. The difference between reliability and failure is

$$D_N = (-g, g) + pG_N(\delta)(\ell_1, \ell_2). \tag{12.1}$$

The first consultant gains the current opportunity in the failure scenario but prefers reliability at a finite horizon precisely when $p\ell_1 G_N(\delta) > g$. The second prefers reliability at horizon one and at every later horizon. The aggregate favors reliability at every horizon from two onward if $p(\ell_1 + \ell_2) > 0$. If the first consultant is completely insulated from the loss of trust, $\ell_1 = 0$, it keeps a private preference for the failed scenario at every horizon.

The effect $p\ell_i$ combines the probability of a shift with the size of the resulting loss. One can have a large reputational loss conditional on a rare public failure, or a high probability of a smaller loss. This simplest model identifies the product from a known expected payoff shift. It cannot distinguish the factors without observing market states, measuring exposure, or changing the conditions under which failures are public.

12.2 Recovery rather than permanent loss

Suppose that starting at date one a low-trust market remains low at each subsequent date with probability q, and otherwise returns permanently to H. The probability of low trust at date $t \geq 1$ after the failed project is pq^{t-1}. Replace G_N in (12.1) by $B_N(\delta, q)$ from (11.3). Consultant one's infinite value of reliability relative to failure is

$$-g + \frac{p\ell_1\delta}{1-\delta q}. \tag{12.2}$$

There is a finite horizon at which this consultant strictly prefers reliability if and only if the displayed value is strictly positive. When $q < 1$ and $p\ell_1/(1-q) \leq g$, no discount in $(0,1)$ gives a strict reversal. Recovery removes the private reason to forgo the transfer if the effects of damage are too short-lived.

The parameter q can reflect public information, the pace at which new clients arrive, or recovery from an industry-wide incident. If market reputation consists of separate client communities, the future loss can be a mixture of several decay rates. With community losses $\ell_{ij} \geq 0$, exposure probabilities p_j, and persistence probabilities q_j, replace the future term for consultant i by

$$\sum_j p_j \ell_{ij} B_N(\delta, q_j).$$

The positivity test still uses the sum, and estimating one average recovery probability need not give the same value. For instance, for fixed $\delta > 0$ the factor $\delta/(1-\delta q)$ is strictly convex in q. Replacing a nontrivial mixture by its appropriately weighted average persistence generally understates this infinite continuation loss. This is one reason to distinguish observable submarkets from a single aggregate reputation score.

12.3 Private quality incentives

A strategic version lets a consultant choose costly quality assurance, at cost k to itself, to reduce the probability of reaching L from p_0 to $p_1 < p_0$. Suppose its own current revenue is otherwise fixed and the induced continuation rewards are those above. Its private criterion for assurance at horizon N is

$$k \leq (p_0 - p_1)\ell_i B_N(\delta, q).$$

The joint value criterion for the two modeled firms instead compares k with $(p_0 - p_1)(\ell_1 + \ell_2)B_N(\delta, q)$. The interval between these two amounts describes an externality

from quality provision. If clients also suffer from failures, the wider welfare boundary adds their losses. The scenario in (12.1) is different: it compares two initial allocations and a probability shift while suppressing a real assurance cost. The quality-choice extension puts that cost back into the model and uses the same continuation law.

The repeat-purchase account of contract performance in Klein and Leffler [17] ties performance to future rents placed at risk by nonperformance. The present calculation instead distinguishes a firm's own exposure from the exposure it creates for its competitors. Reputation models such as Kreps and Wilson [18] make public beliefs about a type change following strategic behavior. Our two-state trust parameter can summarize belief effects when it has a sufficient transition law. It is specified as a reduced-form state here, with no claim that the binary law follows from Bayesian inference.

12.4 Accounting for replacement value without double counting

If the first consultant acquires the client permanently, it can have a positive future reallocation term as well as the reputation loss. Let $u_t \geq 0$ be its additional retained net fees at date t, offset by the same loss to the second consultant. In the reliability-minus-failure comparison, the tail gains then include $(-u_t, u_t)$ in addition to $pq^{t-1}(\ell_1, \ell_2)$. The aggregate formula is unchanged under the common units and discount. The first consultant's private threshold changes and the positive-tail assumption may fail for its coordinate. The geometric error bound still applies. If retaining the client also changes the project's quality or the number of future contracts, those effects require additional transition or reward terms; they should not be counted twice as both transfers and total growth.

The model is compatible with repeated competition for consulting work in both states. Some allocation choices preserve future demand, and others can alter it. The analytic question is which of those choices have nonzero contrasts on the subsequent surplus observable and which players capture or bear the effects. The answer depends on a specified counterfactual law for the market, rather than on the sign of an individual current gain.

13 Negotiation and the survival of a relationship

13.1 An active deal and an outside state

In Dellanna's negotiation example [1], a more balanced current deal can make a counterparty more likely to return. Let A be an active relationship and O the stopped state with zero future joint gains above the chosen outside baseline. Every active deal divides an amount $B > 0$. Put $a = B/2$ and choose $0 < g < a$. A balanced deal F pays (a, a), and a squeeze X pays $(a + g, a - g)$. These are both feasible and acceptable in the base model. Under F the next relationship is active with probability p_F, and under X with probability p_X, where $0 \leq p_X < p_F \leq 1$. The reward and transition repeat in every active period.

This specification gives a direct action-controlled hazard. It is a reduced-form law that may reflect willingness to transact again, outside offers, or a decision to end the relationship. The comparison initially holds that law fixed and treats the proposer as choosing between available deals. An equilibrium interpretation with a responder who chooses whether to participate requires specifying the responder's options and continuation, as done in a special version below.

13.2 Profile comparisons at every horizon

Compare the policy of always choosing F while active with always choosing X. Under the two policies the probabilities of still being active at date t are p_F^t and p_X^t. Their values are

$$V_N^F = (a, a) \sum_{t=0}^{N-1} (\delta p_F)^t,$$

$$V_N^X = (a + g, a - g) \sum_{t=0}^{N-1} (\delta p_X)^t.$$

The aggregate difference equals $B \sum_{t<N} \delta^t (p_F^t - p_X^t)$. It is zero at horizon one and strictly positive at every larger horizon. The counterparty's own difference is strictly positive at every positive horizon because the current share and later survival both

improve. The proposer loses g at horizon one. Its infinite value comparison is strictly positive exactly when

$$\frac{a}{1-\delta p_F} > \frac{a+g}{1-\delta p_X} \quad \Longleftrightarrow \quad \delta\big((a+g)p_F - ap_X\big) > g. \tag{13.1}$$

This condition uses a per-period gain from future activity and a current loss of share.

For the proposer, the expected per-date difference before discounting is $ap_F^t - (a+g)p_X^t$. If $p_X > 0$, the ratio $(p_F/p_X)^t$ is strictly increasing and eventually exceeds $(a+g)/a$. Thus the per-date difference has at most one sign change, from negative to positive. If $p_X = 0$, it is positive from date one. The marginal reversal result of Chapter 6 now proves that (13.1) is also necessary and sufficient for a strict finite crossover. Once the proposer strictly gains at a horizon, it strictly gains at every larger one. At equality in (13.1), the infinite values tie and no finite horizon strictly favors F.

If $p_F = p_X$ were permitted, the survival mechanism would disappear. The aggregate would stay invariant at every horizon, even though the current allocation differs. If p_X is positive and $p_F - p_X$ is small, (13.1) can require a discount outside $(0,1)$ and hence no patience in the modeled range can make F strictly preferable to the proposer. An active relationship is valuable only to the extent that the current action actually changes the future opportunity.

13.3 A one-step choice and an equilibrium reading

Suppose the proposer is the sole controller in the base model. Under the stationary balanced policy its continuation while active is $W_1^F(A) = a/(1-\delta p_F)$. A single squeeze followed by reversion to F gains g immediately but changes the continuation probability by $p_X - p_F$. The condition for no profitable one-step squeeze is

$$g \le \delta(p_F - p_X)\,\frac{a}{1-\delta p_F}. \tag{13.2}$$

Algebra rearranges (13.2) to the weak version of (13.1). Thus in this one-state single-controller model the threshold for preferring the two stationary profiles also characterizes optimality of the balanced policy. The equality of thresholds uses the continuation at a unique active state, a common stopping state, and the specified two actions. In a general stochastic game the comparison between two full profiles does not imply a one-step incentive test.

One extension makes acceptance voluntary. At each active date the proposer chooses F or X. After observing the offer, the responder can accept, getting the stated current share and the stipulated transition, or reject, producing immediate stopping

with outside value zero for both. Give rejection its own decision node at the same date, so acceptance and offer choice are jointly counted as one stage for discounting. Both offers give the responder strictly positive current gains, and their later continuation under the proposed stationary strategy is nonnegative. Thus accepting either offer is strictly optimal at each acceptance node. If (13.2) holds, the proposer chooses F optimally at every active node, by backward elimination for finite deviations and the discounted tail bound for infinite deviations. Hence the strategy pair is subgame perfect for the stated extensive form. If the responder has a valuable outside offer, one or both acceptance inequalities may change. The base model's payoff comparison does not determine those outside values.

13.4 Opportunity cost and state-dependent disagreement

Suppose stopping leads to a stream paying o_i per period to player i, and active payoffs were stated in absolute units. Subtract o_i from every date and every state for that player's comparison. Then the stopped state has zero reward and the active differences above the outside path are the relevant gains. This translation preserves strategy comparisons if the outside stream is the same for all paths after entry into the stopped state, and the subtracted stream is the same constant at every date of each valuation. If outside opportunities improve after a particular kind of deal, stopping itself has an action-dependent value. That possibility requires multiple outside states or a terminal payoff that records the difference.

The survival account relates to relational contracting models [19, 20]. Those models can make continuation value enforce behavior that formal contracts cannot fully specify, and they often model a richer set of actions, information, and enforceable payments. Here a balanced split affects the arrival probability of another deal directly. The mechanism is structurally similar to repeat-purchase discipline [17], but the increase in total is measured from a fixed outside baseline and comes from an additional probability of joint gains. An observed association between balanced deals and repeat transactions could also reflect partners who were likely to return independently of the deal. The transition law in this chapter is a modeling hypothesis that distinguishes those explanations.

14 Robust certificates and empirical interpretation

14.1 A policy deviation identity

An alternative to comparing two whole payoff streams is to evaluate their changes one step at a time against a fixed baseline continuation. For fixed profiles π and ρ, write W^ρ for the bounded infinite continuation from every public history and define the vector advantage

$$A^{\pi|\rho}(h) = r^\pi(h) + \delta(P_\pi^* W^\rho)(h) - W^\rho(h).$$

Here P_π is the history-append kernel. The field $A^{\pi|\rho}$ measures what happens if the current stage uses π and all future stages use ρ. Since W^ρ solves its own policy recursion, the advantage equals the current reward difference plus the kernel contrast against W^ρ. It vanishes at a history if the two profiles prescribe the same action law there.

Proposition 14.1 (Resolving a multi-stage contrast)**.** *For every initial history h,*

$$W^\pi(h) - W^\rho(h) = \mathbb{E}_h^\pi \sum_{t \geq 0} \delta^t A^{\pi|\rho}(h_t). \tag{14.1}$$

Proof. For the first N terms of the right side, use conditional expectation under π to replace $\delta(P_\pi^* W^\rho)(h_t)$ by the discounted next value in expectation. The intermediate terms telescope. The result is $V_N^\pi(h) - W^\rho(h) + \delta^N \mathbb{E}_h^\pi W^\rho(h_N)$. The last term tends to zero and the first tends to $W^\pi(h)$. All sums and limits are justified by boundedness. □

This identity differs from a unilateral incentive test. It can compare joint changes of profile and weights the one-step contrasts by the distribution under the new profile. Thus an advantage at a state rarely visited under ρ can matter when the changed policy visits that state frequently. The same equation holds after scalar aggregation. If every aggregate advantage is nonnegative and one is positive at a finite date with positive probability, the aggregate comparison is strict. If every coordinate advantage is nonnegative with a positive contribution for each coordinate somewhere along the

new law, the comparison is strictly Pareto. These conclusions do not require the future rewards themselves to increase at each date.

14.2 Bounds for transition and reward perturbations

Consider two games with common state and action sets and common initial state. Hold a profile on public histories fixed and assume a uniform reward discrepancy ε_r and a uniform total variation discrepancy ε_p in transition rows. To speak of the same profile at every history, extend it to the union of histories feasible in either model, choosing admissible actions there. Mixing over a common action law cannot increase the bounds. The history kernels differ by at most ε_p in total variation, because they first sample the same action law and then append the state from the two respective rows.

Assume rewards in both models have norm at most M. Comparing the two Bellman equations gives

$$\|W^\pi - \widehat{W}^\pi\|_\infty \le \varepsilon_r + \delta\|W^\pi - \widehat{W}^\pi\|_\infty + \frac{2\delta M}{1-\delta}\varepsilon_p.$$

Therefore their infinite evaluations differ by at most

$$E_\infty = \frac{\varepsilon_r}{1-\delta} + \frac{2\delta M \varepsilon_p}{(1-\delta)^2} \tag{14.2}$$

per coordinate. The bound follows also from the approximate descent calculation with identity observation maps. A coordinatewise improvement margin larger than $2E_\infty$ survives perturbation of both profile values in a comparison. An aggregate margin larger than $2|I|E_\infty$ survives as well. These are conservative bounds because different payoff and transition errors may cancel or act only in states of small occupancy.

A tighter one-step comparison uses the actual range of the continuation coordinate at a history and a row-specific estimate of the transition discrepancy, as in (2.5). In a state with a uniform continuation value, kernel error has no immediate effect on that value. A model may therefore be insensitive to some poorly estimated transition details. Metrics for approximate probabilistic bisimulation in decision processes provide a broader account of how state similarity controls value error [14]. The bound here only compares given profiles and uses prescribed discrepancy estimates.

14.3 Separating a causal claim from a payoff identity

In an application, a statement that a move grows the market includes a causal interpretation of $P(s, a, -)$. Two observed groups with different actions can differ in initial demand, firm quality, or outside opportunities. Their observed continuation distributions then need not estimate the effect of changing an action at a common state. The formal model assumes intervention laws for each compared action. The equations identify the implications of those laws; they do not identify the laws from observed association.

Three design choices are particularly important. The horizon must count the same economic dates for both policies. The reported aggregate must use the same participants and units at every date. Finally, the comparator must specify future actions and how they respond to observations. For example, a single generous deal followed by extraction, and a repeated balanced-offer policy, induce different transition sequences even if their initial action is the same. An unspecified counterfactual cannot be repaired simply by observing more years.

In some cases one can derive a sensitivity region without identifying each parameter. In the author model with decaying awareness, suppose $p \geq \underline{p}$, $w_i \geq \underline{w}_i$, $q \geq \underline{q}$ and the concession is at most $\bar{d}$. A sufficient bound for the conceding author's infinite gain is

$$-\bar{d} + \frac{\underline{p}\,\underline{w}_i\,\delta}{1 - \delta\underline{q}} > 0.$$

This follows from monotonicity of the value in all three lower-bounded parameters. If the lower bound fails, the sign may still be identified by additional information about dependence between the parameters. The simple rectangular bound treats their extreme values as simultaneously possible. Analogous inequalities for the consulting and negotiation cases separate the precision required for a claim about the aggregate from that required for an individual incentive claim.

14.4 A protocol for interpreting model statements

An aggregate calculation should name the comparator, the initial state or distribution, the payoff boundary, and the discount. A crossover claim should specify whether the first crossing persists and whether behavior is held fixed across horizons. A strategic statement should additionally identify the one-step alternatives and the off-path continuation. An observation statement should say which fields descend through the available interface and which comparisons require hidden-state assumptions.

These declarations make it possible to tell whether two claims about the same market actually disagree.

The mathematical certificates in this chapter are conditional on known deterministic error bounds or symbolic parameter bounds. They do not give sampling distributions or confidence levels. Empirical uncertainty can be represented by deriving those bounds from a separately specified study or by evaluating the payoff formulas over a declared parameter region. The strength of a resulting conclusion depends on which parameter values that study can exclude.

14.5 Two observationally identical initial periods

Even exact measurement of the initial allocation is insufficient to distinguish the hypothesis of future demand growth from the hypothesis of a pure transfer. Consider two otherwise identical models with an initial reward change $(-d, d)$ and the same subsequent reward at the baseline state. In the first, the action also shifts probability to an absorbing high-surplus state. In the second, both actions induce the baseline transition. All initial reward data coincide, and the initial aggregate contrast is exactly zero in both. Later contrasts differ by (2.4). A short observation window can fail to distinguish the models even when later outcomes are large, as in the delayed-payment chain. An empirical interpretation of the article's claim therefore requires evidence about continuation states or observations long enough to distinguish the proposed transition mechanisms.

15 Feasible surplus and discounted occupation

The previous criteria can show when feasible profiles have different totals without enumerating those profiles. A separate question is how large a gain can be achieved by controlling joint actions. In a finite stationary game, expected discounted occupancy gives a compact description of the values attainable by a coordinated controller. The controller is an accounting device. Its feasible values form an upper envelope for outcomes under decentralized play.

15.1 The discounted flow equations

Fix an initial distribution μ on S. For any profile π, define the normalized expected occupation of each feasible state-action pair by

$$x^{\pi}(s,a) = (1-\delta)\sum_{t\geq 0}\delta^{t}\Pr_{\mu}^{\pi}(s_t = s, a_t = a).$$

Summing over a gives the state occupation $m^{\pi}(s)$. Conditioning on the previous state and action gives

$$m^{\pi}(s) = \sum_{a} x^{\pi}(s,a) = (1-\delta)\mu(s) + \delta\sum_{y,b} x^{\pi}(y,b)P(y,b,s). \tag{15.1}$$

The initial mass enters at date zero. Summing (15.1) over s gives $\sum_{s,a} x^{\pi}(s,a) = 1$. This holds even with absorbing states, which retain probability mass after activity stops. The normalized vector payoff is the linear image

$$(1-\delta)\mathbb{E}_{\mu}V_{\infty}^{\pi} = \sum_{s,a} x^{\pi}(s,a)r(s,a). \tag{15.2}$$

The exchange of sums follows from boundedness and the finite player set.

Proposition 15.1 (Realizing coordinated occupancy)**.** *Every nonnegative array x satisfying* (15.1) *is the occupation array of a stationary randomized joint controller. The normalized payoff set of such controllers is therefore the image of a compact convex polytope under* (15.2)*.*

Proof. Set $m(s) = \sum_a x(s,a)$ and let the controller choose a at s with probability $x(s,a)/m(s)$ when $m(s) > 0$. Choose any action law where $m(s) = 0$. Call its state kernel P_x. Substituting the policy in (15.1) shows that $m = (1-\delta)\mu + \delta m P_x$. This affine equation has a unique solution because right multiplication by the stochastic matrix is nonexpansive in the ℓ^1 norm and $\delta < 1$. Its solution is the state occupancy of the constructed controller. Multiplying by the action probabilities gives x, including the zero-occupancy states. The flow equations and nonnegativity cut out a closed subset of the unit simplex, which is a compact convex polytope. □

A randomized joint law need not factor into the product of individual action laws. Public randomization can propose a joint action, yet each participant may prefer to deviate after seeing the proposal. Thus the polytope characterizes centralized or coordinated feasibility, while Theorem 7.1 determines whether a proposed protocol is stable. Discounted occupation measures in Markov decision processes and their use in constrained optimization are treated in [21]. The application here uses the aggregate and vector projections of a common occupation model.

15.2 Upper and lower certificates

Let $c(s,a) = \mathsf{sum} r(s,a)$ and suppose a bounded scalar function U satisfies

$$U(s) \geq c(s,a) + \delta P_a U(s) \qquad \text{for all } s, a. \tag{15.3}$$

For any profile, sum the discounted inequalities after conditioning on each action. Telescoping gives $\mathsf{sum} V_N^\pi(s) \leq U(s) - \delta^N \mathbb{E}_s^\pi U(s_N)$ and hence $\mathsf{sum} V_\infty^\pi(s) \leq U(s)$. If there is a deterministic policy choosing equality at every state, its value is U. Reversing the inequalities gives an analogous lower certificate. Integrating (15.3) against x and using (15.1) also yields

$$\sum_{s,a} x(s,a) c(s,a) \leq (1-\delta) \sum_s \mu(s) U(s).$$

This bound can be evaluated either through the time recursion or through the normalized flow equations.

Since the sets $A(s)$ are finite and nonempty, the map

$$(\mathcal{T}_+ U)(s) = \max_{a \in A(s)} \{c(s,a) + \delta P_a U(s)\}$$

is a contraction of ratio at most δ in the supremum norm. Its fixed point admits a deterministic maximizer at every state and is the maximal aggregate value for

coordinated control, even relative to history-dependent joint actions. Replacing the maximum by the minimum gives the minimal value. Their equality from every state is equivalent to the potential criterion of Theorem 5.1: a shared value is both a maximum and a minimum and therefore satisfies equality for every action. If they differ at a state, the maximizer and minimizer witness a strict aggregate ranking there. Neither choice is guaranteed to be a Nash equilibrium.

15.3 Allocations in the feasible value set

For fixed μ the expected vector value set under a centralized controller is convex by Proposition 15.1. A strictly positive aggregate difference from a baseline ρ identifies a direction whose inner product with $\mathbf{1}$ is positive. Strict Pareto improvement additionally asks whether the feasible set intersects $V^{\rho} + \mathbb{R}^{I}_{>0}$. A point with a higher sum can lie outside that orthant because one coordinate is negative. The discrepancy need not be small, even with a large scalar gain.

Suppose all feasible changes from a baseline have the form $(u, -\theta u)$ with $u \geq 0$ and $0 < \theta < 1$. Then every nonzero change raises total value while the second player loses. In the transfer model of Chapter 7 an enforceable side payment can move such a point parallel to the hyperplane of constant aggregate. Absent that instrument, the convexity of attainable vectors provides no such movement. Mixing plans along the ray leaves the second coordinate weakly worse. This geometric distinction explains why showing that a coalition creates value can fall short of showing that its internal allocation is acceptable to all members.

At a finite horizon, introduce one occupancy variable for each date and use unnormalized flows from the initial law. The date-t variables carry weight δ^{t} in the payoff. The existence of a feasible gain then depends on the stage count and on the available joint actions. The crossover theorem in Chapter 6 instead follows two fixed profiles as the truncation date changes. One compares envelopes of feasible values. The other compares two designated paths through those envelopes.

16 Approximate invariance and coarse policy comparisons

Exact invariance says that the aggregate reward is unchanged across a family of profiles. In an application, a more useful claim may be that every change of profile has small aggregate effect at the horizons relevant to a decision. Exact potentials and forecast contrasts give two quantitative versions of this claim.

16.1 A residual in the potential equation

Given a bounded scalar field ϕ on the state space, set

$$e_\phi(s,a) = c(s,a) - \phi(s) + \delta P_a \phi(s).$$

By the telescoping argument of Theorem 5.1,

$$\mathsf{sum} V_N^\pi(s) = \phi(s) - \delta^N \mathbb{E}_s^\pi \phi(s_N) + \mathbb{E}_s^\pi \sum_{t<N} \delta^t e_\phi(s_t, a_t). \tag{16.1}$$

Thus if $|e_\phi| \le \varepsilon$, every infinite discounted aggregate lies within $\varepsilon/(1-\delta)$ of $\phi(s)$, from the same initial state. The difference between any two lies within $2\varepsilon/(1-\delta)$ of zero. A bound for a finite comparison adds the difference of terminal expectations, which by (2.5) is bounded by $\delta^N \operatorname{osc}(\phi)$ and can be smaller when terminal distributions are close.

A deterministic baseline b gives a canonical candidate ϕ_b for which $e_{\phi_b}(s, b(s)) = 0$ at every state. From any fixed start the aggregate gain against this baseline is at most $\sup e_{\phi_b}/(1-\delta)$ and at least $\inf e_{\phi_b}/(1-\delta)$. If residuals vanish at all states except a set with small discounted occupancy under the compared policies, the gain can be bounded using just that occupancy. To see this, restrict the sum in (16.1) to the residual's support, and bound each term by ε times the probability of visiting the support. The improvement identity (14.1) is the vector version when the potential is a baseline continuation on histories.

The factor $(1-\delta)^{-1}$ cannot be omitted without further assumptions. In a single-state system with one action paying aggregate ε and another paying $-\varepsilon$ at each date, the potential $\phi = 0$ has residual bound ε. Always choosing the first and always

choosing the second gives a difference $2\varepsilon/(1-\delta)$, exactly the bound. A small flow discrepancy can accumulate into a material present value when the horizon and patience are large.

16.2 An approximate converse

Suppose that for every state the range of infinite aggregate values over profiles has diameter at most η. Fix a deterministic baseline b and let ϕ_b be its value. Forcing a single action and then reverting to b gives an infinite value whose difference from $\phi_b(s)$ is exactly $e_{\phi_b}(s,a)$. It follows that $|e_{\phi_b}(s,a)| \le \eta$ for every state and action. Combined with (16.1), this gives an equivalence of exact zero residual and exact invariance, as well as bounds in both directions for approximate invariance. The constants differ because a one-step certificate bounds a quantity that may occur again at every date.

A baseline need not be optimized to serve as a potential witness. A tighter symmetric certificate may be obtained by a different ϕ. For instance, take the midpoint $\phi = (U+L)/2$ of the extremal values from Chapter 15. If $0 \le U - L \le \eta$ pointwise, the upper inequality for U and lower inequality for L give

$$-\tfrac{1}{2}(U-L)(s) + \tfrac{\delta}{2}P_a(U-L)(s) \;\le e_\phi(s,a) \;\le \tfrac{1}{2}(U-L)(s) - \tfrac{\delta}{2}P_a(U-L)(s).$$

These bounds follow by writing $c + \delta P_a U \le U$ and $c + \delta P_a L \ge L$, then replacing U and L by $\phi \pm (U-L)/2$. In particular $|e_\phi(s,a)| \le \eta/2$. For this midpoint witness the next-state spread can reduce the residual further. The approximation expresses closeness of all aggregate returns, independent of whether the policies are close in action distribution.

16.3 Finite forecast errors

Assume (4.1) and let g_n be the finite baseline field from Chapter 4. Put $\xi_n = \sup_{s,a} |L_{s,a} g_n|$. Define e_n to be the supremum, over histories and policies, of the absolute difference between the true n-stage aggregate continuation and g_n at the current state. Conditioning on the first step shows

$$e_{n+1} \le \delta e_n + \delta \xi_n, \qquad e_0 = 0.$$

The term e_n comes from replacing the actual continuation at the next state by the baseline value. The remaining discrepancy is the kernel contrast on that value. Expanding

the recursion gives

$$e_N \leq \sum_{n=0}^{N-1} \delta^{N-n} \xi_n.$$

This bound retains the first date at which a discrepancy becomes visible. If the first k forecast contrasts vanish exactly, early terms contribute zero rather than a generic error allowance. If known bounds are on individual generators, say $|L_{s,a} P_b^j c| \leq \zeta_j$, then $\xi_n \leq \sum_{j<n} \delta^j \zeta_j$ by linearity. The finite theorem is recovered when all ζ_j vanish.

This comparison needs a baseline forecast and a bound on how each action changes its expectation. It can be used without building an exact lumped model of the entire state process. A failure of the bound to certify approximate invariance can reflect large uncertainties or a genuine change in aggregate prospects. The claim certified by a small bound is specific to the horizon and payoff definition used to form g_n.

16.4 Combining a margin with imperfect observation

Suppose a reduced model supplies an estimated finite aggregate improvement $\widehat{C}_N$. If the two reduced values differ from the true values by at most E_N per coordinate each, the aggregate comparison error is at most $2|I|E_N$. If the true payoff increments are bounded in magnitude by B in aggregate, then

$$\widehat{C}_N > 2|I|E_N + \frac{B\delta^N}{1-\delta}$$

certifies strict aggregate improvement at N and all later horizons, including the discounted infinite horizon. This is the triangle inequality combined with Proposition 6.1. The same reasoning on each coordinate uses its own margins and bounds. If one coordinate is nearly indifferent, an aggregate certificate can remain valid while no coordinatewise certificate is possible.

The sufficient margin separates two problems. Improving the model of current and subsequent transitions can reduce E_N. Extending the reporting horizon reduces the second term. With approximate descent errors incurred at each new date, the first effect may need separate evidence, because a longer horizon does not itself make a quotient more accurate.

17 Parallel markets and a common clock

The category in Chapter 2 composes stages sequentially. Some applications also combine separate markets over the same dates, or couple them by a shared state. A synchronous product distinguishes adding independent revenue streams from changing how those streams develop.

17.1 Synchronous composition

Fix a common player set and discount. For parallel independent processes $f = (n, P, r) : X \to Y$ and $g = (n, Q, u) : X' \to Y'$ of the same grade, define

$$f \otimes g = (n, P \otimes Q, r \boxplus u) : X \times X' \to Y \times Y', \qquad (r \boxplus u)(x, x') = r(x) + u(x').$$

The product kernel has independent conditional components, and the rewards add in each player's coordinate. The grade remains n, since the two processes run over the same dates. To state the categorical domain, take the category whose objects are pairs of state interfaces and whose arrows are pairs of reward arrows with equal grade. It is a fiber product over the grade category. The operation $\otimes$ is a functor from this category to $\mathcal{R}_{\delta,I}$.

For functoriality, consider pairs of grade n followed by pairs of grade m. Kernel composition commutes with independent products, and for future fields v, w on the two respective factor states,

$$(P \otimes Q)^*(v \boxplus w) = P^* v \boxplus Q^* w.$$

This equality follows by summing out each other factor using its probability mass one. Applying (2.1) gives

$$(f_2 \otimes g_2) \circ (f_1 \otimes g_1) = (f_2 \circ f_1) \otimes (g_2 \circ g_1)$$

for matching grades in both pairs. Identities are preserved by products. Product reassociation and swap, viewed as zero-reward grade-zero deterministic arrows, give the usual coherent associativity and symmetry isomorphisms. For each grade,

a singleton process with zero reward and trivial transition serves as the inactive companion for that grade. These are compatible with sequential composition.

With a terminal field of the form $v \boxplus w$, the composite continuation factors additively:

$$(f \otimes g)^\sharp(v \boxplus w) = f^\sharp(v) \boxplus g^\sharp(w).$$

Consequently, when each factor's policy and initial marginal are held fixed, total expected additive rewards are their sum. An initial correlation does not change that equality for separated policies, since the expectation of a sum depends on the marginal laws. A later policy that chooses the first market's action based on the second market's state can change the first marginal law. The product comparison then needs to be justified for the resolved process, rather than inferred from the initial state description.

17.2 Correlated transitions and interactions

Independence is sufficient for the product law but unnecessary for the last displayed equality with additive fields. Suppose a joint kernel has first marginal $P(- \mid x)$ and second marginal $Q(- \mid x')$ at every source pair (x, x'). It then pulls back an additive terminal field to the same additive result, by linearity and the marginal identities. Two joint kernels can agree on all such fields while differing on interaction observables. This is another instance of Proposition 2.3 with a restricted family of terminals.

For example, suppose each of two markets has a binary next-date demand shock with marginal probability $1/2$ of being high. One joint law makes the shocks coincide, and another makes them opposite. The expectation of the number of high shocks is one under either law. The expected indicator of both being high is $1/2$ under the first law and zero under the second. If a later joint venture is profitable only when both markets are high, its terminal field tests that interaction. Under the common discount the two laws then differ by $\delta/2$ in the expected contribution of a unit joint gain. Marginal invariance does not imply invariance of aggregate value when the relevant aggregate field itself has an interaction term.

A shared reputation state in consulting can couple markets in two ways. It can shift the marginal probability of demand in each community, as in Chapter 12, or change the correlation of demand without changing those marginal probabilities. An industry-wide capacity constraint or risk-sensitive cash need can make the latter economically relevant. An additive expected flow computed before the constraint would not capture that effect. A controlled state with the relevant joint observable makes the mechanism explicit.

17.3 Asynchronous opportunities

When two activities take different lengths of time, the equal-grade product above does not apply directly. One can put both on a common clock and include an explicit wait in the shorter process, with its own intervening rewards. A wait of k dates followed by the same continuation multiplies that continuation by δ^k. Waiting can also carry holding costs or intermediate opportunities, which enter through the reward of the waiting process. Modeling the shorter activity as immediately realized at the later shared boundary without a correction can change present values.

A different case is a random end date, as in negotiations. An absorbing state gives a common clock by making stopped activity pay zero, while active activity continues to affect future rewards. When an action changes the probability of remaining active, its effects belong in the transition law. Replacing the law by a single discount factor would give different effective discounts to different policies and conceal the difference between the common intertemporal preference and policy-specific survival. The survival formulas in Chapter 13 use the product δp_F or δp_X only after the policy is resolved.

18 Prior work and conclusions

18.1 Dynamic programming and stochastic games

Bellman's dynamic programming formulation [2] organizes sequential evaluation by the value remaining after a decision. Blackwell's treatment of discounted dynamic programming [4] addresses optimality of policies in a controlled stochastic system. Shapley's stochastic games [3] bring strategic control of the transition into the picture. The fixed-policy contraction and the history-state construction used here belong to this established framework. They also let us formulate the same problem at interfaces where different reward processes are composed.

The question here is a property of all policies: which controlled changes disappear when a vector process is read as a scalar total? The finite-horizon theorem expresses that question as simultaneous annihilation on the baseline's observable orbit. The potential theorem expresses the infinite version as cancellation of an endpoint correction. The two criteria show precisely how a one-stage fixed total can survive or fail under iteration. The difference between testing an orbit and testing one discounted resolvent explains why individual finite horizons can disagree with the infinite comparison.

18.2 Open games and categorical probability

Ghani, Hedges, Winschel, and Zahn [5] make games compositional by retaining strategies, play, and best responses to continuations. Ghani, Kupke, Lambert, and Nordvall Forsberg [6] develop infinite iterated open games using final coalgebra semantics and formalize subgame perfection in that setting. Our reward category is built after policy resolution. Composition transports bounded expected rewards and their terminal context, giving naturality of aggregation and a simple order for substitutable arrows. The one-step deviation theorem restores an incentive test at the full game level while keeping the evaluation construction separate.

Stochastic kernels, as used here, form the standard categorical model of discrete probability. The framework of Markov categories extends this viewpoint to abstract probability and statistical structures [12]. Our category adds a time grade and accumulated affine reward, using ordinary sums and expectation on countable sets. It does not require a disintegration operation. The exact quotient theorem is the block-sum condition for strong lumpability together with a reward condition, expressed as a

commuting square. The sheaf construction uses restriction in a space of contexts, a different operation from probabilistic lumping. It isolates the quantifiers of a claim about how long improvement takes.

18.3 Repeated games, reputation, and value creation

The folk-theorem results of Fudenberg and Maskin [8] and of Dutta [9] address the support of feasible, individually rational outcomes as equilibria under their respective repeated or stochastic game hypotheses. The recursions of Abreu, Pearce, and Stacchetti [10] and of Hörner, Sugaya, Takahashi, and Vieille [11] treat imperfect monitoring or related stochastic equilibrium questions. The aggregate criterion is logically prior to those support problems. It may identify a gain that the game cannot sustain, or it may show that an entire equilibrium payoff set lies on a common aggregate level.

Brandenburger and Stuart [16] analyze value creation and appropriation within a set of participants. The distinction is complementary to a controlled transition that changes which future opportunities exist. Klein and Leffler [17] use future market rents to explain reliable performance. Kreps and Wilson [18] give an informational mechanism for reputation. Baker, Gibbons, and Murphy [19] and Levin [20] treat self-enforcing relational contracts and incentives. Our examples isolate three reduced-form pathways for the changes described in the motivating essay: success in building demand, exposure to shared reputational damage, and survival of a mutually valuable relationship.

18.4 Conclusion

When actions preserve the current aggregate reward, their total effect at a later horizon comes from the continuation through the states those actions make more or less likely. Agreement on all relevant continuation forecasts characterizes preservation of the total across finite horizons. At a fixed infinite discount, even visibly different prefixes can have the same total through a discounted potential. Consequently, a longer horizon reveals an aggregate difference only under conditions on what present behavior changes and how those changes affect the chosen payoff boundary.

In the authorship case, the next book sale can remain competitive while awareness increases the future number of sales. In consulting, replacement receipts can be offset by lower demand if failure harms shared trust. In negotiation, retaining a counterpart can be worth a current concession when the probability shift and future joint gain are large enough. The aggregate crossover, the first gain for each person, and support by equilibrium can occur at different horizons. Their distinction is the formal content

of the claim that rivals can simultaneously compete for a share and affect the future surplus.

18.5 Questions left open

The present method evaluates specified histories and profiles. Under private information the observation available to one player can differ from the observation available to another. An analogue of the quotient theorem would need to identify which posterior statistics preserve each player's continuation and deviation values. The aggregate forecast span alone would generally be too small for that task.

A second direction concerns endogenous market formation. The examples treat market awareness, trust, and relationship survival as parts of a specified stochastic law. A more detailed model could let entry, demand, or new counterparty matching change endogenously. The invariance question would then be which of those changes a scalar continuation can observe. That extension requires a common boundary for the changing population, or an explicit rule for how arriving and departing participants enter the payoff aggregate.

Evidence and authorship

The manuscript uses mathematical proofs and symbolic examples. The three applications are illustrative models of the mechanisms in Dellanna's essay and contain no estimates about his businesses or the markets described. The byline is anonymous for editorial review. Generative AI assisted with drafting and literature discovery. The responsible author or research group is to be identified to the journal in its submission workflow as required.

References

[1] Luca Dellanna. Long-Term Iterated Games. Essay, 29 October 2026. Author's publication.

[2] Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences* 38(8), 716–719, 1952. doi:10.1073/pnas.38.8.716.

[3] Lloyd S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences* 39(10), 1095–1100, 1953. doi:10.1073/pnas.39.10.1095.

[4] David Blackwell. Discounted dynamic programming. *Annals of Mathematical Statistics* 36(1), 226–235, 1965. doi:10.1214/aoms/1177700285.

[5] Neil Ghani, Jules Hedges, Viktor Winschel, and Philipp Zahn. Compositional game theory. *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, 472–481, 2018. doi:10.1145/3209108.3209165; open preprint.

[6] Neil Ghani, Clemens Kupke, Alasdair Lambert, and Fredrik Nordvall Forsberg. A compositional treatment of iterated open games. *Theoretical Computer Science* 741, 48–57, 2018. doi:10.1016/j.tcs.2018.05.026.

[7] Andrew Y. Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. *Proceedings of the 16th International Conference on Machine Learning*, 278–287, 1999. Paper.

[8] Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54(3), 533–554, 1986. doi:10.2307/1911307.

[9] Prajit K. Dutta. A folk theorem for stochastic games. *Journal of Economic Theory* 66(1), 1–32, 1995. doi:10.1006/jeth.1995.1030.

[10] Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58(5), 1041–1063, 1990. doi:10.2307/2938299; working paper.

[11] Johannes Hörner, Takuo Sugaya, Satoru Takahashi, and Nicolas Vieille. Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem. *Econometrica* 79(4), 1277–1318, 2011. doi:10.3982/ECTA9004.

[12] Tobias Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics* 370, 107239, 2020. doi:10.1016/j.aim.2020.107239.

[13] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer, 1976. Publisher's record.

[14] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov decision processes. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 162–169, 2004. arXiv:1207.4114.

[15] Saunders Mac Lane and Ieke Moerdijk. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer, 1992. doi:10.1007/978-1-4612-0927-0.

[16] Adam M. Brandenburger and Harborne W. Stuart Jr. Value-based business strategy. *Journal of Economics & Management Strategy* 5(1), 5–24, 1996. doi:10.1111/j.1430-9134.1996.00005.x.

[17] Benjamin Klein and Keith B. Leffler. The role of market forces in assuring contractual performance. *Journal of Political Economy* 89(4), 615–641, 1981. doi:10.1086/260996.

[18] David M. Kreps and Robert Wilson. Reputation and imperfect information. *Journal of Economic Theory* 27(2), 253–279, 1982. doi:10.1016/0022-0531(82)90030-8.

[19] George Baker, Robert Gibbons, and Kevin J. Murphy. Relational contracts and the theory of the firm. *Quarterly Journal of Economics* 117(1), 39–84, 2002. doi:10.1162/003355302753399445.

[20] Jonathan Levin. Relational incentive contracts. *American Economic Review* 93(3), 835–857, 2003. doi:10.1257/000282803322157115.

[21] Eitan Altman. *Constrained Markov Decision Processes*. Chapman & Hall/CRC, 1999. Publisher record.