The Geometry of Competent Contact: What a Scalar Return Cannot See in an Agent's Grip on the World

PIATRA . INSTITUTE

July 2026

Abstract

Reinforcement learning scores a policy by a single number, the expected return $J(\pi) = \mathbb{E}[R]$, and treats everything an agent does as a means to that scalar. Competent action resists the scalar. When a hand turns a cup to read its shape, or a robot taps a surface to tell metal from plastic, what improves is not a running total but the agent's grip on the world, the way its uncertainty about hidden states collapses into task-relevant structure. This paper gives that grip a formal object. The agent's belief over world-states is a sufficient statistic for control, it lives on the probability simplex, and Chentsov's theorem forces the Fisher-Rao metric on that simplex, so the movement of belief has a geometry that is not chosen but compelled. Competent contact is read off that geometry: the contraction of the uncertainty set, the Fisher-Rao length of the belief trajectory, and the separability the terminal belief keeps between world-states. The central result is a quotient. If reward depends on the world only through a partition g, the expected return depends on the belief only through its pushforward g_*b, so two beliefs at the maximal Fisher-Rao distance π can be reward-identical, and a scalar objective is blind to everything the geometry resolves within a reward-class. A small active-perception POMDP makes this exact. A reward-only policy and a competent-contact policy attain the identical single-task return of 0.8, yet leave beliefs of very different shape: terminal entropy 1.09 nats against 0.68, and, in a representational dissimilarity matrix, world-states that differ only in the reward-irrelevant attribute collapse to Fisher-Rao distance 0.0 under the reward-only policy while the competent policy holds all six apart at least 2.44. The discarded geometry is what transfers: on a second task the reward never mentioned, the reward-only agent scores at chance, 0.5, and the competent agent scores 0.85, a gap of 0.35 that no single-task return could have priced. A from-scratch curvature routine returns the simplex curvature of 0.25 to machine precision, certifying the geometry the argument runs on.

The Return and the Grip

A reinforcement learner is defined by what it maximizes. The standard formulation makes the target a scalar, the expected sum of rewards $J(\pi) = \mathbb{E}[R]$, and the agent is good to the exact degree that this number is large (Sutton and Barto, 2018). Everything else, perception included, is instrumental to the number. Sensing is worthwhile when it raises expected reward and worthless otherwise, and the quality of an agent's contact with its world has no standing in the theory except through its eventual cash value in return.

Perception was described very differently by the tradition that took it seriously as an achievement. Gibson held that an animal does not receive a retinal image and infer a world from it but actively picks up invariant structure in the ambient array, and that what it picks up are affordances, the possibilities the environment offers to that body (Gibson, 1979). Merleau-Ponty described perception as tending toward an optimal hold on its object, the body settling into the world until the scene comes into the clearest grip it admits, and Dreyfus named that tendency a "maximal grip," the drive to reduce the felt disequilibrium between the perceiver and the perceptual field (Merleau-Ponty, 2012; Dreyfus, 2002). O'Regan and Noë made the point operational: to see is to exercise mastery of the sensorimotor contingencies, the lawful ways sensory input changes as the agent

moves, so that vision is a skill of probing rather than a stream of pictures (O'Regan and Noë, 2001). Across these accounts perception is a competence, a graded quality of coupling between an agent and a world, and it is measured by how well the coupling resolves the world into the structure the agent's activity requires.

This paper takes that idea and gives it a mathematical object. The claim is that an agent's competence is the geometry of how its actions transform its belief over the world, and that this geometry is strictly richer than any scalar return computed from it. Competent contact is a property of a trajectory on a curved manifold, and the return is a coarse projection of that trajectory onto a line. The projection discards structure. The paper identifies which structure it discards, proves that the discard is generic rather than accidental, and shows on an exact model that what falls through the projection is what lets an agent carry its competence from one task to another.

The Belief Simplex Is the Arena

The right state variable for an agent that cannot see the world directly is its belief about the world. Åström proved for a partially observed Markov process that the posterior distribution over hidden states is a sufficient statistic for optimal control, so an optimal policy can be written as a function of the belief alone and needs nothing else from the history (Åström, 1965). Kaelbling, Littman, and Cassandra built the modern theory of planning on this fact, recasting a partially observable problem as a fully observable Markov decision process whose states are beliefs, so that the object the agent actually navigates is the space of distributions over world-states (Kaelbling, Littman, and Cassandra, 1998). That space is the probability simplex. A belief over N world-states is a point b in Δ^{N-1}, and an action, through the observation it yields and Bayes' rule, is a map that sends one point of the simplex to another.

The simplex is not a featureless triangle. It carries a metric, and the metric is forced. Rao observed that the Fisher information of a family of distributions can be read as a Riemannian metric, giving a geodesic distance between distributions (Rao, 1945), and Chentsov proved the uniqueness that removes all arbitrariness from the choice: on a finite sample space the Fisher-Rao metric is, up to an overall scale, the only Riemannian metric invariant under the maps that carry one statistical model into another by a sufficient statistic (Chentsov, 1982; the extension beyond finite sample spaces is Ay, Jost, Lê, and Schwachhöfer, 2017). Any other way of measuring how far one belief sits from another would import structure the statistics does not contain. If belief is the arena, its geometry is Fisher-Rao and nothing else.

On the simplex that geometry has a shape one can compute with. The map $b \mapsto 2\sqrt{b}$ carries Δ^{N-1} isometrically onto a piece of a sphere of radius 2, on which the Fisher-Rao distance between two beliefs is the great-circle distance

$$d_{FR}(b, b') \;=\; 2\arccos\Big(\textstyle\sum_i \sqrt{b_i\, b'_i}\Big),$$

running from 0 for identical beliefs to π for beliefs on disjoint supports (Amari and Nagaoka, 2000). A sphere of radius 2 has constant Gaussian curvature a quarter, and the simulation computes that

quarter rather than assuming it. A curvature routine built only from the metric components and the Brioschi formula, fed the Fisher metric on a grid across a three-state marginal simplex, returns 0.25 with a maximum deviation of 0.0 across the grid. It is computed rather than assumed. The instrument recovers the known value before it is trusted on the unknown ones. With the arena fixed, competence becomes a question about paths: how an agent's actions move its belief across this curved surface, how far the belief travels, and where it comes to rest.

Reward Is a Quotient

The scalar return sees the belief through a narrow window. A task assigns a reward $R(d, w)$ to each decision d the agent might commit to and each world-state w that might obtain, and the Bayes value of holding a belief b is the best expected reward it supports, $V(b) = \max_d \sum_w b(w)R(d, w)$. The question is how much of the belief this value can register. The answer is that it registers only a quotient.

Proposition 1. *Suppose the reward depends on the world-state w only through a partition $g\colon W \to \{1, \dots, T\}$, so that $R(d, w) = \tilde{R}(d, g(w))$ for every decision d. Then the Bayes value $V(b) = \max_d \sum_w b(w)R(d, w)$ depends on the belief b only through its pushforward g_*b, the induced belief on classes. Consequently any two beliefs b, b' with $g_*b = g_*b'$ are reward-equivalent for the task, even when their Fisher-Rao distance $d_{FR}(b, b')$ is as large as the diameter π of the simplex.*

Proof. Write $c = g(w)$ for the class of w and group the sum by class: $\sum_w b(w)R(d, w) = \sum_w b(w)\,\tilde{R}(d, g(w)) = \sum_c \big(\sum_{w:\, g(w)=c} b(w)\big)\,\tilde{R}(d, c) = \sum_c (g_*b)(c)\,\tilde{R}(d, c)$. The inner expression depends on b only through the class-sums $(g_*b)(c)$, so the maximum over d does too. If $g_*b = g_*b'$ then $V(b) = V(b')$. Beliefs supported on distinct states within a single class have disjoint support yet identical pushforward, and disjoint support gives $d_{FR} = \pi$. □

The proposition is elementary. Its content is that the reward-relevant part of a belief is its projection onto the coarsest partition the reward respects, and that everything finer is invisible to the return. Blackwell's comparison of experiments is the same fact seen from the side of information: one experiment is sufficient for another when it is at least as useful in every decision problem, and an experiment can be sufficient for a particular task while discarding distinctions that a different task would need (Blackwell, 1953). A scalar objective asks the belief for its answer to one question. The geometry of the belief holds the answers to all of them.

The simulation exhibits the extreme case directly. Take a world whose state is a pair $w = (A, B)$ of two independent attributes, an A in $\{0, 1, 2\}$ and a B in $\{0, 1\}$, so there are 6 world-states on the simplex Δ^5. Let task 1 reward depend on the world only through A. The two beliefs concentrated on $(A, B) = (0, 0)$ and on $(0, 1)$ have the same A-marginal, a point mass at $A = 0$, so both value certainty about task 1 at 1.0 and support the same optimal decision. They sit at Fisher-Rao distance 3.14, the diameter of the simplex, because their supports are disjoint. And they answer a task about B in opposite ways: the decision that is optimal under the first scores 0.0 under the second, the entire stake of that task. Two beliefs as far apart as the manifold allows, indistinguishable to one

task and opposite on another. The return cannot be the measure of competence, because it cannot see the difference.

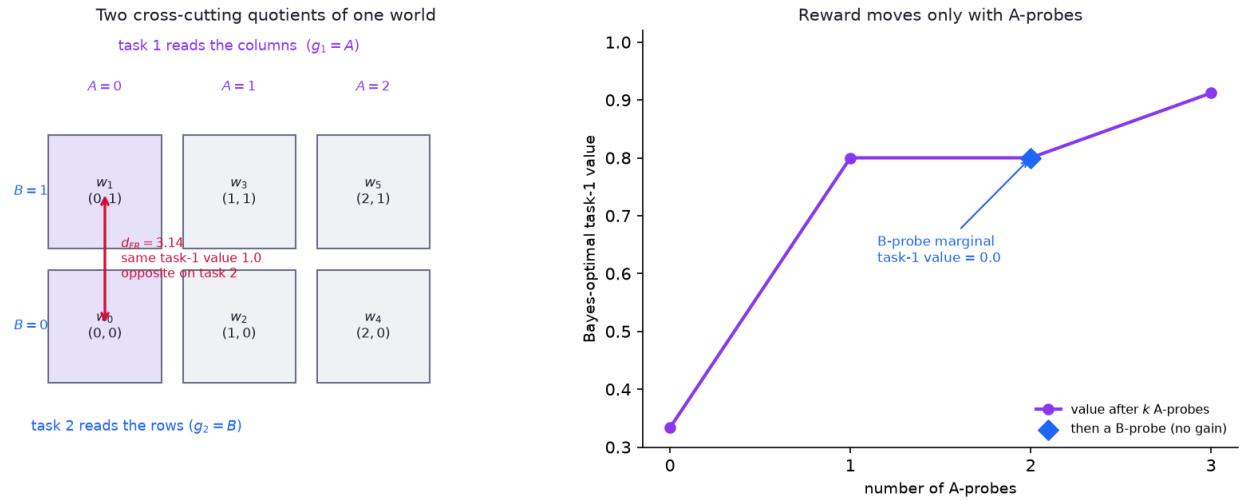

Figure 1: The world and its quotients. Left: the 6 world-states as a grid, A across the columns and B down the rows. Task 1 reads only the columns (the partition $g_1 = A$); task 2 reads only the rows ($g_2 = B$). The beliefs at $w_0 = (0, 0)$ and $w_1 = (0, 1)$ sit at Fisher-Rao distance 3.14, the diameter, yet share the task-1 value 1.0 and take opposite task-2 decisions. Right: Bayes-optimal task-1 value against the number of A-probes, rising from a prior value with a diminishing return; a B-probe added afterward leaves the task-1 value unchanged, a marginal contribution of 0.0, because B lies in the quotient the reward discards.

Two Policies the Return Cannot Tell Apart

To act on the world the agent probes it. A probe is a sensing action with a fixed observation channel, and in the model there are two: an A-probe that reports the attribute A through a confusion channel, correct with probability 0.8, and a B-probe that reports B, correct with probability 0.85. Each probe outcome updates the belief by Bayes' rule, so a plan of probes traces a path across the simplex. Because the world is static and the channels are memoryless, every expectation below is computed exactly by enumerating observation outcomes weighted by their probabilities, with no sampling and no seed.

Consider two policies that differ only in what they try to resolve. The reward-only policy is built to maximize task-1 return, and it faces a sharp asymmetry: an A-probe raises the task-1 value, while a B-probe leaves it untouched, a marginal task-1 value of 0.0, because B falls in the quotient the reward cannot read. No policy that maximizes task-1 return ever spends a probe on B. The competent-contact policy is built instead to resolve the world, choosing probes by expected information gain about the full state w, which is the classical objective of Bayesian experimental design, the expected reduction in the posterior's Shannon entropy (Lindley, 1956), and equivalently the expected Bayesian surprise, the KL divergence from prior to posterior that measures how far the belief moves (Itti and Baldi, 2009; Gottlieb, Oudeyer, Lopes, and Baranes, 2013). To isolate the

one difference that matters, both policies are given the same 2-probe budget on A; the competent policy additionally spends 2 probes on B.

The return cannot tell them apart. Both policies leave the A-marginal identically distributed, since B-probes do not touch it, so both attain the same expected task-1 value of 0.8, a gap of 0.0. A theory that scores policies by single-task return would rank these two as equivalent. Read a per-probe cost of 0.05 into the return, the small price of taking an action in the world, and the ranking inverts against competence: the reward-only policy nets 0.7 on task 1 while the competent policy nets 0.6, because the competent policy paid for 2 probes that task 1 does not reward. By the scalar the field actually optimizes, the more competent agent is the worse agent.

The geometry of the two beliefs is not the same at all. The reward-only policy leaves a terminal belief of Shannon entropy 1.09 nats, most of it the untouched uncertainty about B; the competent policy leaves 0.68 nats, having driven the belief further toward a vertex. Measured as Fisher-Rao length from the prior, the competent belief travels 1.62 against the reward-only 1.28, a longer path across the manifold to a more resolved place. The sharpest reading is representational. For each policy, take the belief it ends with as a function of the true world-state and form the 6-by-6 matrix of Fisher-Rao distances between those terminal beliefs, the representational dissimilarity matrix of systems neuroscience (Kriegeskorte, Mur, and Bandettini, 2008; Kriegeskorte and Kievit, 2013). Under the competent policy every pair of world-states is held apart, the smallest separation being 2.44. Nothing collapses. Under the reward-only policy the world-states that share an A and differ only in B collapse onto each other, their separation falling to 0.0: the agent's representation cannot tell them apart, because it never asked. The reward-only agent has built a representation in which the reward-irrelevant distinctions of the world simply do not exist.

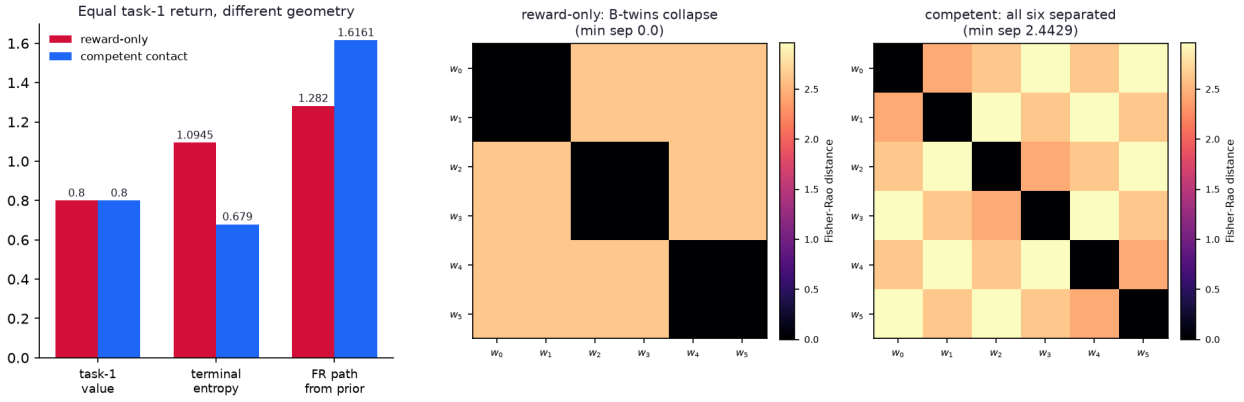

Figure 2: Two policies, one number. Left: the reward-only and competent-contact policies attain the identical task-1 value 0.8, yet the reward-only policy leaves higher terminal entropy (1.0945 against 0.679 nats) and a shorter Fisher-Rao path from the prior (1.282 against 1.6161). Middle and right: the representational dissimilarity matrices, Fisher-Rao distances between the terminal beliefs indexed by the true world-state. The reward-only matrix collapses the three pairs that differ only in B to distance 0.0 (dark off-diagonal blocks), a minimum separation of 0.0; the competent matrix separates all six states, minimum 2.4429.

What the Discarded Geometry Was Worth

The distinctions the reward-only policy discarded were not noise. They were the structure of the world along a dimension that one task happened not to reward, and a different task rewards it fully. Let task 2 depend on the world only through B, and let it arrive after the probing is done, to be answered from the belief the agent already holds, with no further contact. This is the ordinary situation of transfer: an agent that understood its world for one purpose is asked to act for another, and cannot go back to sense again.

The competent agent transfers and the reward-only agent does not. The reward-only policy, having never probed B, holds a belief whose B-marginal is still the uniform prior, so its best zero-shot task-2 value is 0.5, chance. The competent policy resolved B to the channel's fidelity and scores 0.85, a transfer gap of 0.35 that is the within-class structure the scalar quotient threw away. Average over a task family that draws its reward from either attribute with equal probability, and the competent agent's expected value is 0.825 against the reward-only agent's 0.65. The gap is not a matter of the competent agent being generally stronger. On task 1 the two are equal, and net of cost the competent agent is behind. The gap is entirely the value of the geometry that task 1 did not price and task 2 redeemed.

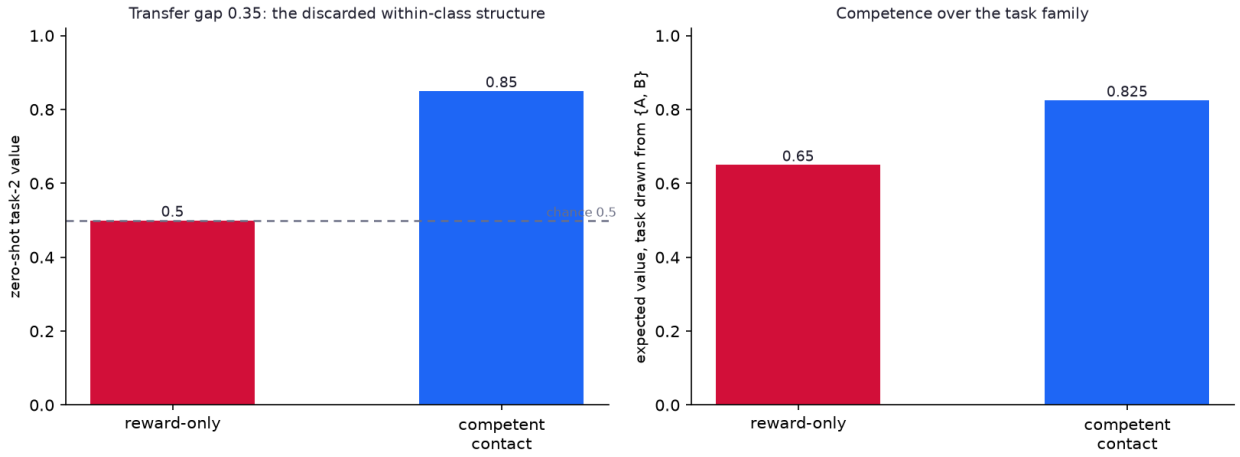

Figure 3: What the geometry buys. Left: zero-shot task-2 value with no further probing. The reward-only policy sits at chance, 0.5, having never resolved B; the competent policy scores 0.85; the transfer gap of 0.35 is the within-class structure the reward-only quotient discarded. Right: expected value over a task family drawing its reward from either attribute, 0.65 for the reward-only policy against 0.825 for the competent one.

This is the sense in which competent contact is a real quantity and the return is a shadow of it. The competent-contact policy is seeking Blackwell-sufficiency for the world rather than for a task: an experiment sufficient for the full state is preferred in every decision problem the world could pose, and that is what makes it transfer (Blackwell, 1953). The information-theoretic drives studied as intrinsic motivation are neighbors of this quantity, and it is worth being precise about how they differ. Empowerment measures the channel capacity from an agent's actions to its later sensor states, the most an agent can inject into its own sensory future, and it rewards an agent for being

in control (Klyubin, Polani, and Nehaniv, 2005; Shannon, 1948; Cover and Thomas, 2006). Competent contact measures something adjacent but distinct, how much the agent's probing resolves the world's hidden structure, and an agent can have high empowerment over a world it barely understands. The epistemic term in active inference is closer still: expected free energy scores a policy partly by its expected information gain about hidden states, which is the competent-contact objective wearing the clothes of variational inference (Friston, Rigoli, Ognibene, Mathys, Fitzgerald, and Pezzulo, 2015; Tishby and Polani, 2011). The contribution here is not to add another intrinsic reward to the pile. It is to identify the object all of these are groping toward, the belief manifold and its Fisher-Rao geometry, and to prove that this object is what a scalar return generically fails to see.

Contact Is Always with a Carved World

The belief manifold is defined over the agent's own hypothesis space, and that is a boundary the geometry cannot cross. The world-states w are the possibilities the agent represents, the attributes A and B are the dimensions along which it can tell things apart at all, and the Fisher-Rao distances are distances in the agent's own belief, not in the world. A different agent, with a hypothesis space that carved the same substrate along different attributes, would inhabit a different manifold, and the competence measured on one is not commensurable with the competence measured on the other. Uexküll made this the center of his biology: every animal lives in an Umwelt, a world articulated by its own functional cycles, so that the tick's world and the dog's world and the physicist's world are genuinely different worlds built on one physical ground (Uexküll, 2010). Competent contact is contact with the world as the agent has carved it, and there is no view from nowhere in which an agent's grip could be scored against the world in itself.

This bounds the claim in a useful way. The geometry does not certify that an agent's carving is the right one, and an agent that resolves its own hypothesis space perfectly may still be resolving the wrong distinctions, competent within an Umwelt that a harder world will break. Two of the model's stipulations mark the same edge. The reward-relevant partition and the reward-irrelevant one were fixed in advance, and the whole demonstration lives in the gap between them; in a world where every distinction is eventually rewarded there is no discarded geometry to recover, and competent contact and reward maximization coincide. The interesting worlds are the ones where they do not, where an agent must resolve more than any single task rewards because it does not yet know which task it will face. And the policies were hand-specified to isolate the phenomenon rather than derived as optima, so the numbers are facts about the construction and not measurements of any agent in any world. What survives the stipulations is the proposition, which needs none of them: wherever reward factors through a coarsening of the world, the return is blind to the belief's finer geometry, and that geometry is what an agent carries from the task it was trained on to the task it was not.

A learner built to maximize a scalar will, if the scalar is faithful to its training task, throw away the structure that would have let it do the next thing. It will look most efficient at the moment it is least prepared to be asked something new. The geometry of competent contact is the account

of what it threw away, and the reason that the throwing-away has a cost is that a world is always larger than the reward an agent is currently being paid to collect from it.

References

Amari, S., and Nagaoka, H. (2000). *Methods of Information Geometry* (D. Harada, Trans.). Translations of Mathematical Monographs 191. American Mathematical Society and Oxford University Press.

Ay, N., Jost, J., Lê, H. V., and Schwachhöfer, L. (2017). *Information Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, 64. Springer.

Åström, K. J. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1), 174–205.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2), 265–272.

Chentsov, N. N. (1982). *Statistical Decision Rules and Optimal Inference* (Translations of Mathematical Monographs 53). American Mathematical Society. (Original work published 1972.)

Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.

Dreyfus, H. L. (2002). Intelligence without representation: Merleau-Ponty's critique of mental representation. *Phenomenology and the Cognitive Sciences*, 1(4), 367–383.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11), 585–593.

Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2), 99–134.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation* (Vol. 1, pp. 128–135). IEEE.

Kriegeskorte, N., and Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412.

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis: Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4), 986–1005.

Merleau-Ponty, M. (2012). *Phenomenology of Perception* (D. A. Landes, Trans.). Routledge. (Original work published 1945.)

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–973.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37, 81–91.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.

Tishby, N., and Polani, D. (2011). Information theory of decisions and actions. In V. Cutsuridis, A. Hussain, and J. G. Taylor (Eds.), *Perception-Action Cycle: Models, Architectures, and Hardware* (pp. 601–636). Springer.

Uexküll, J. von (2010). *A Foray into the Worlds of Animals and Humans* (J. D. O'Neil, Trans.). University of Minnesota Press. (Original work published 1934.)