Structure Learning with Hierarchical Models for Computational Motor Control

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Tim Genewein
aus Bludenz, Österreich

Juli 2016

Tag der mündlichen Prüfung: Mittwoch, 29. März 2017

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Prof. Dr. Dr. Daniel A. Braun
2. Berichterstatter: Prof. Dr. Martin A. Giese

Prüfungskommission: Prof. Dr. Felix Wichmann
Prof. Dr. Martin A. Giese
PD Dr. Axel Lindner
Prof. Dr. Dr. Daniel A. Braun

Erklärung

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:
"Structure Learning with Hierarchical Models for Computational Motor Control"
selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den _____ _____

(Datum) (Unterschrift)

Declaration

I hereby declare that I have produced the work entitled:
"Structure Learning with Hierarchical Models for Computational Motor Control",
submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, on _____ _____

(date) (signature)

Acknowledgements

This thesis concludes my work in Daniel Braun's group at the MPI in Tübingen. While I am excited to move on to new challenges, I am very thankful for the wonderful environment that I've spent the last four years in—it will be hard to beat (both, our group and Tübingen). I want to thank Daniel for bringing our group together, showing me how science works and giving me the freedom to pursue my own ideas. A big thanks goes to Pedro, who inspires me to ask fundamental questions, think deep and who taught me much when I arrived to Tübingen. Another big thanks to Jordi, Felix and Zhen who have been great colleagues but have also become good friends that I will certainly miss.

Large parts of my education would have not been possible without the help of a number of people, which I shall not forget. First and foremost my mother, who always encouraged and supported me, despite tight financial constraints. I have also had help from Hugo, Bruno, Dietmar, Albert, Evelyn and the Bruderschaft St.Christoph.

Along the way, I've had the fortune to get to know and spend time with Tom, Matthäus, Chris, my brother Moritz and Mandy—thank you!

Tim Genewein

Extra kudos go to the Austrian federal aid system, which made it possible to almost fully finance a Bachelor's and Master's (including cost of living) and graduate without any debt. This is by no means to be taken for granted—hopefully education will remain important enough for our society to keep this system up.

Abstract

Recent advances in machine learning have enabled the use of data-sets of unprecedented scale to tune models with millions of parameters. This has lead to breakthroughs in areas such as computer vision and natural language processing. With this wave of big-data methods the term "artificial intelligence" (AI) has been revived and heavily marketed, sometimes along with bold claims. However, it seems that one important feat of intelligence is almost orthogonal to the current trends in machine learning: the capability to rapidly learn from few, sparse, noisy and ambiguous observations. In contrast to current state-of-the-art artificial systems, animals and humans excel at learning from very few observations, especially in sensorimotor tasks. It is thought that this key capability of biological intelligence heavily relies on the ability to extract abstract knowledge which can then be transferred to novel situations, leading to generalizations that go far beyond simple inter- or extrapolation. The process of extracting abstract, invariant knowledge from concrete observations is called *structure learning*.

This thesis aims to shed some light on the computational models and principles underlying structure learning, with a focus on human sensorimotor processing. Structure learning, or the formation of abstractions, depends on the ability to separate relevant information from irrelevant variation or noise. In the theoretical part of this thesis, this idea is fleshed out by viewing structure learning from an information-theoretic angle. In particular, rate distortion-theory, the framework for lossy compression, provides mathematically well grounded ideas on structure learning that can lead to the emergence of natural levels of abstraction in a decision-making scenario. Interestingly, there is also a deep, formal connection to a thermodynamic treatment of bounded rational decision-making, where gains in utility are traded off against information processing demands. Under this viewpoint, abstractions facilitate decision-making under information processing limitations by focusing resources on processing relevant information.

In the experimental part of this thesis, hierarchical Bayesian models and ideas from Bayesian model selection are applied to describe human behavior in sensorimotor structure learning tasks, performed in virtual reality. The hierarchical Bayesian framework turns out to be very suitable for capturing hierarchies of abstractions and for modeling hierarchical inference processes. We found that human behavior was consistent with a Bayesian model in two sensorimotor model selection tasks and that human sensorimotor processing was well described by a hierarchical Bayesian model in a structured sensorimotor integration task.

The thesis concludes by applying the insights gained from the rate-distortion viewpoint on structure learning to perception-action systems and hierarchies of abstractions. In particular, we found that under information processing limitations, perception and action should become tightly coupled in a particular way and cannot be treated as separate problems. Based on the same theoretical framework, a second interesting finding is an optimality principle that leads to the emergence of a two-level hierarchy of abstractions, where more abstract and thus potentially transferable knowledge is captured by the upper level of the hierarchy.

Contents

1 Introduction

> *"Arguably the most elusive aspect of intelligence is the ability to make robust inferences that go far beyond one's experience. [...] Such inductive leaps are thought to result from the brain's ability to infer latent structure that governs the environment."*
>
> Tervo, Tenenbaum, Gershman [Tervo et al., 2016]

1.1 What is structure learning?

Animals and humans excel at flexibly adapting their behavior in order to efficiently cope with variation and uncertainty in their environment. Robust inference and decision-making in the light of noisy, ambiguous or sparse data is paramount for survival in a natural environment that is subject to permanent change. At the same time natural environments are highly structured in the sense that their perpetual variation exhibits rich regularities on different scales, thus allowing for organisms to thrive if they can efficiently exploit these regularities. It has been argued before that: "One reason that animals and humans can rapidly learn new problems is perhaps because they take advantage of the high degree of structure of natural tasks." [Gershman and Niv, 2010].

By exploiting the structure of natural tasks and stimuli, animals and humans are able to solve intricate problems, such as perceiving complex objects, learning abstract concepts and rules and developing sophisticated motor skills. It is thought that these feats of biological intelligence heavily rely on two key-features of intelligence that are enabled by structure learning:

1. **Extraction of transferable knowledge**: learning and inferring the (latent) structure that governs the environment. The structure is invariant under variations of the environment and thus enables the extraction and transfer of abstract knowledge to novel but similar situations.
2. **Robust inference from few observations**: generalizing well from few, potentially noisy and ambiguous observations. This is possible if the search-space over hypotheses can be sufficiently narrowed down by some abstract, general prior-knowledge (the learned structural regularities).

The question of how to extract abstract, transferable knowledge while learning specific tasks is central to this thesis and the corresponding computational process is referred to as *structure learning*. More formally, structure learning refers to the extraction of higher-level statistical invariants that are shared across a number of tasks (or stimuli). Structure learning enables the formation of abstractions and allows for hierarchical information processing, which both are hallmarks of human and animal intelligence that lead to the ability to rapidly adapt to novel but

similar problems by transferring previously acquired knowledge. Transferring knowledge from one task to another is one of the central open problems in contemporary AI research, where specific tasks can be learned very well but even mild task variations often require complete re-learning. A theoretical understanding of structure learning is thus highly desirable from a robotics or machine-learning point-of-view, but could also be crucial for understanding complex behavior and learning processes in animals and humans.

This thesis aims to contribute towards understanding the *computational* principles underlying structure learning and facilitated adaptation to novel tasks due to exploitation of structural knowledge. In the experimental studies that are part of this thesis (Chapters 4, 5, 6), the focus is on structure learning in sensorimotor tasks, that is movement-tasks with sensory feedback. The theoretical part of the thesis (Chapter 7) has a broader scope and applies equally to inference- and decision-making-scenarios.

1.1.1 Structure learning - an intuitive example

To intuitively illustrate the concept of structure learning and how structural knowledge can facilitate adaptation to novel situations consider the example illustrated in Figure 1.1, which is a variation of an example originally published by Goodman in 1955 [Goodman, 1955]. In the example, a box (Box 1) is presented - the box contains bags that are filled with red and blue marbles. The first N bags are opened and all the marbles inside are revealed (see Figure 1.1 A). All marbles within an observed bag turn out to be uniform in color, that is either all marbles inside a single bag are blue or all marbles inside a single bag are red. Finally, a new bag is pulled out from the box, a single marble is drawn from the new bag and turns out to be blue. The goal is now to predict the colors of the remaining marbles in the new bag. Arguably, a reasonable prediction would be that all marbles in the new bag are blue. This prediction is based on the observation that all bags observed so far were uniform in color and the assumption is that the new bag follows the same regularity. The *structural invariant* of the bags in Box 1 is the color-uniformity within a bag.

Now assume that another box (Box 2) is presented. Again, the box contains bags filled with marbles. Similar to Box 1 the contents of N bags are revealed. This time, all bags contain marbles of both colors (see Figure 1.1 B) but there are always more red marbles than blue marbles inside each bag. Like before, a new bag is pulled from Box 2 and a single marble is drawn and its color is revealed - the color of the single marble turns out to be blue. What would now be a good prediction for the colors of the remaining marbles in the new bag from Box 2? Perhaps that the bag contains marbles of both colors, probably more red than blue marbles? The crucial point is that the prediction over the color-distribution for the new bag from Box 2 is *different* from the prediction over the marbles' colors in the new bag from Box 1, even though the specific situation was *identical*: a single observed marble drawn from a novel bag that turned out to be blue. The difference between the two situations is subtle but important - by observing the contents of many bags from each of the boxes, different structural regularities could be learned for each box. That is, for Box 1 all bags are uniform in color and for Box 2 all bags have mixed colors and always contain more red than blue marbles. Applying this abstract, structural knowledge to the corresponding novel bags then leads to different predictions under the same observation.

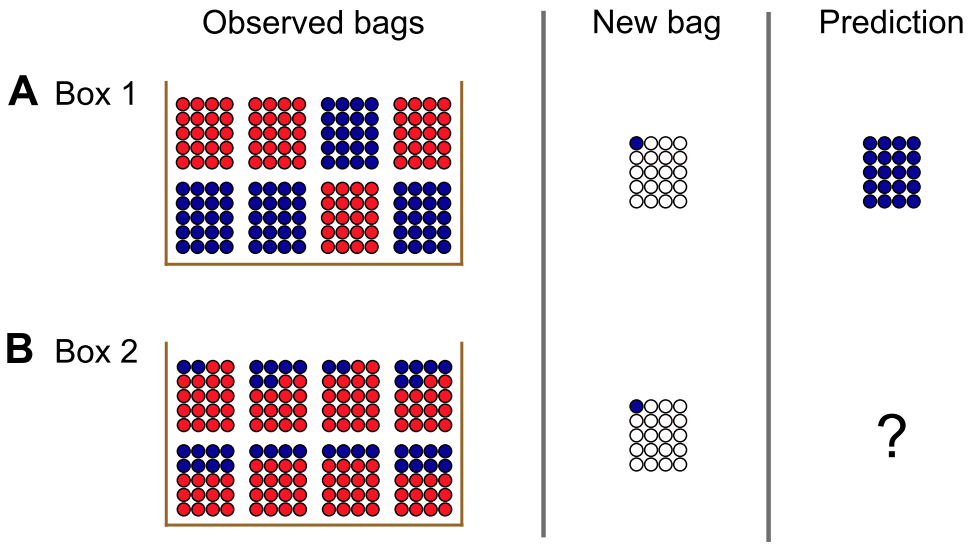

Figure 1.1: Bags & Marbles example. **A** A box (Box 1) contains bags filled with marbles. 8 of the bags are opened and their content is presented. Each bag is exclusively filled with either red or blue marbles. Subsequently, a new bag is taken from Box 1 and a single marble is shown from the new bag - the marble turns out to be blue. The question is: what is the prediction for the colors of all marbles in the new bag? Since all bags observed from Box 1 share the regularity that the bags contain only uniformly colored marbles, a reasonable assumption would be that all marbles in the new bag are blue. Note how knowledge about (structural) invariants can lead to strong predictions based on very sparse data (a single observation). **B** Another box (Box 2) is presented and the contents of 8 bags drawn from the box are revealed. In contrast to Box 1, the bags from Box 2 contain mixed-colored bags and each bag contains more red than blue marbles. A new bag is drawn from Box 2 and a single marble is revealed. Just like previously, the marble turns out to be blue. What is now the prediction for the colors of the remaining marbles in the novel bag? Interestingly, the exact same observation (a blue marble drawn from a novel bag) now leads to different predictions. The reason is that the predictions are based on different assumptions about the structural invariants that are shared across the bags in each box. This structural knowledge was extracted from the data provided by the fully observed bags and is transferred to the novel bag.

The example illustrates the two key-features of structure learning: one, the extraction of transferable knowledge in the form of invariants and two, robust inference from few observations by exploiting the learned structure. These features allow for good generalizations and rapid adaptation to novel tasks that share the same structure. Remarkably, structure learning in this example seems very intuitive and effortless to humans—some researchers even argue that humans might have an innate, general ability to extract and exploit structural invariants [Tenenbaum et al., 2011]. Despite this intuitive simplicity, the general computational principles underlying such structure learning processes are far from being fully understood.

1.1.2 Hierarchical probabilistic models as a key component

The aim of this thesis is to gain a quantitative understanding of the computational processes underlying structure learning. This section illustrates why hierarchical probabilistic models

might play an important role for structure learning and how the corresponding computational processes can be described by (hierarchical) Bayesian inference. The Bags & Marbles example from the previous section can be modeled with a hierarchical probabilistic model (see Figure 1.2), which has been previously illustrated by Kemp, Perfors and Tenenbaum [Kemp et al., 2007] who also suggested that the human brain, when faced with this example, must have a (flexible) mechanism that allows to implement the same computational principles. On the lowest level of the hierarchical model, the color-distribution of marbles within a single bag b can be described by a Binomial distribution $p(y_b = R|\theta_b)$ which quantifies the probability that exactly R out of all the N_m marbles y_b in bag b are red:

$$y_b \sim \text{Binomial}(\theta_b, N_m).$$

The Binomial distribution is parameterized by θ_b, which is the probability of drawing a red marble (from an infinite bag).

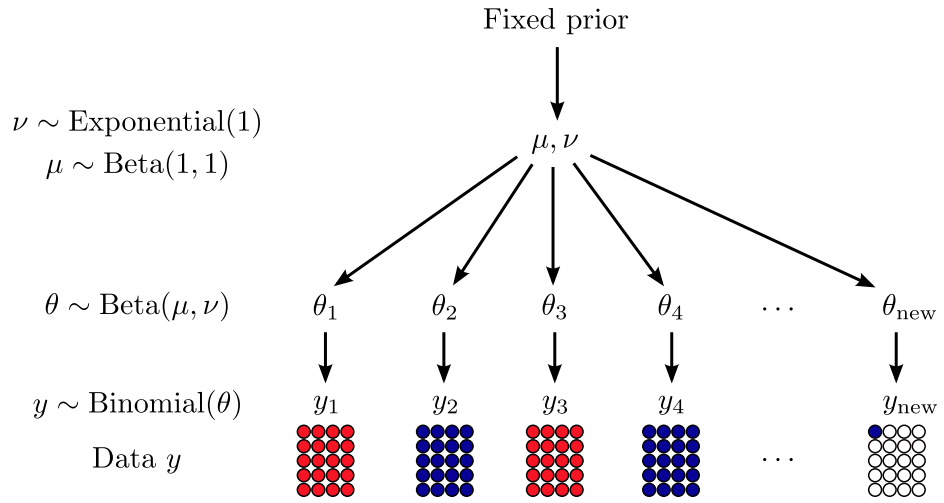

Figure 1.2: Hierarchical probabilistic model for the Bags & Marbles example. On the lowest level of the hierarchy is the data D, consisting of observations y_i, that is the colors of the marbles inside a single bag. The distribution over colors within a single bag can be modeled with a Binomial distribution. Importantly, the binomial parameters θ of the observed bags are not distributed randomly but follow some statistical regularities. These regularities are captured by a Beta-distribution that models the distribution of θ-values across bags. The Beta-distribution has two parameters, μ and ν, which are unknown but can be inferred from the data (by observing many bags). In a Bayesian setting, this inference is performed by placing prior distributions over μ and ν and computing the corresponding posteriors in light of the observed data.

In Box 1 all the observed bags $b \in \{1 \dots N\}$ are uniform in color (Figure 1.1 A) which means that the Binomial parameter for each bag is (very close to) either $\theta_b = 0$ in case of an all-blue bag or $\theta_b = 1$ in case of an all-red bag. This is interesting because the parameters θ are not completely random, instead they show clear regularities. These regularities are captured by the statistics over θ. In particular, the distribution over θ for all bags from Box 1 can be modeled by a Beta-distribution $p(\theta|\mu,\nu)$:

$$\theta_b \sim \text{Beta}(\mu,\nu).$$

The Beta-distribution acts as a prior over the color-distribution of a bag (each bag is parameterized by θ_b) and captures higher-level statistical invariants that are shared across all bags of a particular Box. The implicit assumption is that all θ are drawn from the same Beta-distribution with parameters [1] μ, ν.

Learning the "structure" of a box boils down to finding good parameters μ and ν. Rather than using point-estimates for the parameters, the hierarchical model is extended by another level that models the probabilistic belief over the parameters μ and ν. To do so, prior-distributions are placed on each parameter (the hyper-priors):

$$\mu \sim \text{Beta}_{\text{std}}(1,1)$$
$$\nu \sim \text{Exponential}(1)$$

The prior over μ is a uniform Beta-distribution (using the standard shape-parameterization). The prior over ν is an Exponential-distribution with parameter 1, which corresponds to a weak prior over small positive real numbers. Through Bayesian inference, these priors can be combined with the data D of the observed bags to form a posterior-belief $p(\mu,\nu|D)$.

When faced with a single observation y_{new} from a new bag, the posterior-belief over the color-distribution of the new bag is given by:

$$p(\theta_{\text{new}}|y_{\text{new}},D) = \int_{\mu,\nu} p(\theta_{\text{new}}|\mu,\nu)p(\mu,\nu|y_{\text{new}},D),$$

where $p(\mu,\nu|y_{\text{new}},D)$ can be computed using Bayes' rule and depends on $p(\mu,\nu|D)$. For instance, after observing the uniformly-colored bags in Box 1 the prior probability over θ is most dense around $\theta = 0$ and $\theta = 1$. A single observation of a novel bag is then sufficient to collapse the posterior to one of the two possibilities. The prior knowledge has narrowed down the hypothesis space dramatically, which means that the correct hypothesis can be found rapidly. See Figure 1.3 which shows how the distributions on the different levels of the hierarchy evolve in light of new observations from the different boxes.

The crucial point of this modeling-example is that hierarchical probabilistic models are well suited for implementing structure learning since the priors and hyper-priors naturally capture higher-level statistical invariants that are shared across a range of instances. Additionally, hierarchical Bayesian inference provides a computational mechanism to extract and exploit structural invariants using such models.

1.2 Literature overview

This section provides a literature overview over structure learning phenomena observed in animals and humans and the most influential computational ideas on formalizing structure

1 Note that for mathematical convenience the Beta-distribution here uses an alternative parameterization where μ is the mean and ν can be interpreted as a "virtual sample size" parameter - this allows to easily find suitable prior-distributions for μ and ν. The standard, shape-parameterization uses two parameters $\alpha, \beta > 0$ that can roughly be interpreted as the number of *virtual* observations

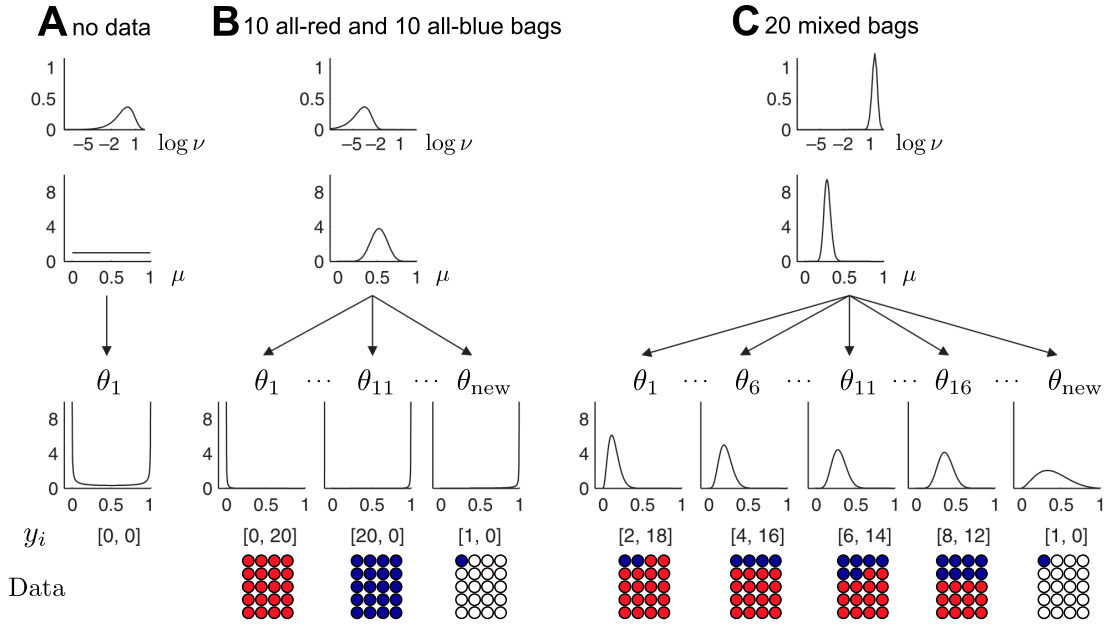

Figure 1.3: Modeling Bags & Marbles with a hierarchical Bayesian model. **A** (Hyper-)prior distributions over the parameters μ and ν as well as the resulting prior over θ, the color-bias within a single bag. The prior assigns more probability mass to uniformly colored bags of both colors. **B** Posterior distributions over μ and ν after having observed 10 all-red and 10 all-blue bags. The bottom row shows the corresponding posteriors over θ after having observed an all-red bag (left), an all-blue bag (center) and a single blue marble (right). Importantly, the single blue marble leads to a posterior over θ that assigns almost all probability mass to blue marbles. Note that the distribution over the mean-parameter μ is centered around 0.5 (same number of red and blue marbles across all observed bags) and small values of ν indicate large color-uniformity within a bag. **C** Posterior distributions over μ and ν after having observed 20 mixed-color bags (5 bags of each kind shown in the very bottom of panel C). The distribution over μ is now peaked below 0.5, capturing the fact that all bags contain more red than blue marbles and the concentration parameter ν now indicates a large probability of mixed-color bags. Correspondingly, the posterior over θ after observing a single blue marble does no longer collapse to a delta over blue marbles, but becomes rather smeared out. Figure originally published in [Kemp et al., 2007] and reproduced with permission. The copyright of the figure remains with the authors of the original publication and Blackwell Publishing Ltd, 9600 Garsington Road, Oxford.

learning. The interested reader is referred to the following suggested literature, which provides well-written, in-depth reviews on structure learning.

- **Structure Learning in Action** [Braun et al., 2010a]. Structure learning in animal cognition, cognitive neuroscience and motor learning. This review is most related to the issues discussed in this thesis.
- **How to grow a mind: statistics, structure, and abstraction** [Tenenbaum et al., 2011]. Cognitive neuroscience perspective on structure learning (including categorization, child- and language-development). Review is focused on the "structured-probabilistic-models" approach to cognition (and structure learning).
- **Learning latent structure: carving nature at its joints** [Gershman and Niv, 2010].

Review focused on structure learning in the context of learning causal models and the corresponding facilitation of reinforcement-learning.

- **Approaches to cognitive modeling** [Kousta, 2010]. Editorial in *Trends in Cognitive Scineces* that features two very influential viewpoints on cognition (including the learning and exploitation of structure). The editorial contains two lead articles, one on each viewpoint, and several comments on the articles by experts in the field: McClelland et al. *Letting structure emerge: connectionist and dynamical systems approaches to cognition* [McClelland et al., 2010] (this article also provides a good introduction and overview on the connectionist view) and Griffiths et al. *Probabilistic models of cognition: exploring representations and inductive biases* [Griffiths et al., 2010].

1.2.1 Structure learning in animal cognition

Facilitated learning of novel tasks with a similar structure has been previously investigated in cognitive science. One of the earliest instances was Harlow in 1949 [Harlow, 1949] who presented monkeys with the choice over two stimulus-objects (e.g. cylinders or cubes), but only one of the objects would lead to a reward. Within a block, the same stimulus-pair was shown at different locations—across blocks, the stimulus-objects were varied. Harlow noticed that initially the monkeys took many trials to figure out the correct response, but in later blocks their success rate was 95% already on the second trial of each block. The same qualitative results results were later found in many variations of this paradigm, for instance also with rats and odor-scented flower pots as stimuli [McKenzie et al., 2014]. This facilitation of learning in later blocks of experiments was termed 'learning to learn' which Harlow defined as the *formation of a learning-set* [Schrier, 1984, Bailey and Thomas, 2001, Langbein et al., 2007], but is sometimes also referred to as *transfer learning*.

Further experiments unraveled that facilitated learning of novel tasks critically depends on certain features of the training- and test-stimuli. For instance, in 1969 Mackintosh & Little [Mackintosh and Little, 1969] found that facilitation is substantial for *intradimensional* shifts compared to *extradimensional* shifts. In their task, pigeons were presented with colored shapes where the color was predictive of the reward. When presenting novel colors (intradimensional shift), pigeons could rapidly adapt. In contrast, adaptation was much slower when presenting trials where the reward depended on the shape of the object (extradimensional shift). The same conclusions were drawn in other experimental studies that investigated adaptation to intra- versus extradimensional shifts [Roberts et al., 1988, Trobalon et al., 2003, Garner et al., 2006], suggesting that some animals are able to learn and exploit abstract concepts such as color or shape. In experimental psychology, structure learning is often thought of as learning a general rule or set of rules that govern a task—e.g. the color but not the shape of an object determines its reward. Animals were also found to be able to learn quite complex rules, for instance by exposing the animals to conditional discrimination and categorization tasks [FUJITA, 1983, Preston et al., 1986, Young and Wasserman, 1997, Thompson and Oden, 2000, Wasserman et al., 2001] (e.g. discrimination based on stimulus color if the stimuli are presented on a dark background versus discrimination based on stimulus shape if the stimuli are presented on white background [Reznikova, 2007]). Other instances of complex rule learning by animals are studies of complex domain-specific behavior in expert animals, such as tool use and intuitive physics in corvids [Weir et al., 2002, Cheke et al., 2011] and non-human primates

[Whiten et al., 2005, Seed et al., 2009] and use of natural or artificial syntax in songbirds [Abe and Watanabe, 2011, Beckers et al., 2012]. Note that in most studies mentioned before, structure learning refers to learning a set of rules, but in general the term structure learning is broader than just learning "hard" rules and also refers to learning of "soft", statistical rules such as correlations or covariations, see Section 1.2.3 for a more detailed discussion.

1.2.2 Structure learning in human cognition

Learning-to-learn phenomena have also been observed in adult humans [Duncan, 1960, Hultsch, 1974, Halford et al., 1998] and children [Inhelder and Piaget, 1964, Carey, 1985, Brown and Kane, 1988]. The underlying structures and abstractions learned and used by humans have been particularly investigated in category learning [Markman, 1989, Rosch, 1978, Anderson, 1991, Ashby and Maddox, 2005, Perfors and Tenenbaum, 2009], language acquisition [Carey and Bartlett, 1978, Chomsky, 1965] as well as causal learning and object-name-generalization in child development [Carey and Bartlett, 1978, Jones and Smith, 2002, Shultz, 2003, Gopnik et al., 2004]. Children and adults can learn the meaning of a new word, novel concepts and relations between objects very quickly from sparse and noisy data, which is only possible if the corresponding inference problem is sufficiently constrained by the underlying structure [Kemp and Tenenbaum, 2008]. Structure learning has been proposed before to explain how humans are able to generalize well in light of sparse and ambiguous data when judging similarity between animals or numbers [Tenenbaum and Griffiths, 2001a, Kemp et al., 2004]. Humans have been shown to be able to learn and use many different kinds of structures, such as partitions, ordered chains, hierarchies, trees, grids or rings [Kemp and Tenenbaum, 2008, Kemp, 2008]. Additionally, adults have been shown to be able to flexibly choose the most appropriate representation for a given problem—for instance, the authors in [Novick and Hurley, 2001] presented college students with short verbal scenarios that encoded information of a certain structure. Participants were then asked to choose the most efficient structure (matrix, network or hierarchy) for organizing the information in each of the scenarios. The authors found that choices were very uniform across participants and reflected the structure of the information encoded in each scenario very well.

Humans and animals excel at visual object recognition, a domain that is closely related to categorization. The problem of recognizing objects based on visual input alone is solved so well and seemingly effortlessly by humans that the massive underlying computational processes are easily overlooked or underestimated [Logothetis and Sheinberg, 1996]. However, on a retinal level, the exact same visual information is hardly encountered multiple times throughout a lifetime. Even slight changes to an object's pose, illumination, position or deformation can have huge effects on the object's raw visual appearance. So how can any system learn to robustly recognize objects? The key is to exploit invariants on many different levels [DiCarlo et al., 2012]—for instance by extracting invariant features and discarding irrelevant information in early processing stages. While some of these invariant-extractors are innate to humans and have been shaped by evolutionary processes, other invariant-extractors are acquired in early stages of development (for example oriented edge detectors in cats are developed in early developmental stages [Hirsch and Spinelli, 1970, Stryker et al., 1978]) or learned on demand [Pinto et al., 2010]. For instance, consider the abstract visual concept of a chair: many details such as pose, illumination, color, material or even particular shape are variations of the abstract object and

are considered irrelevant information for recognizing that the object is a chair. However, there is some crucial *structural* information that clearly separates chairs from other object categories such as tables. Importantly, learning these category-invariants is a structure learning problem, but rather than learning a hard, explicit set of rules, visual object recognition and categorization relies upon flexible structure-learning mechanisms to learn soft constraints and (high-level) statistical regularities. Some researchers even argue that such high-level constraints crucially drive conscious, explicit visual perception in a top-down fashion [Navon, 1977, Hochstein and Ahissar, 2002].

Another feat of biological intelligence is the ability to discover and exploit causal structure in the environment. This important instance of structure learning has been studied before in cognitive neuroscience [Tenenbaum and Griffiths, 2001b, Steyvers et al., 2003, Gopnik et al., 2004, Griffiths and Tenenbaum, 2005, Kemp and Tenenbaum, 2009]. Causal structural knowledge is particularly useful when interacting with an environment, since causal knowledge allows to make better predictions about the outcome of certain actions and also helps in identifying relevant targets for strategy-exploration and -refinement. For instance, consider three factors involved in scientific publishing: quality of research, impact factor, advancement of the field. High-quality research will cause both, a larger impact factor and increased advancement of the scientific field and thus all three variables are correlated. However, since the quality of research is the cause and the other two factors are two independent effects (conditionally independent given the cause), the only way to increase the advancement of the field is to increase the quality of research. An erroneous assumption about the causal structure based on correlation alone might easily lead to the exploration of a strategy that tries to increase scientific advancement by increasing the impact factor alone. Knowledge about causal structure is particularly relevant for control and decision-making (but can also simplify inference problems, especially over latent variables) and is thought to be crucial for efficient exploration strategies when learning through interaction in rich, high-dimensional environments, as in reinforcement learning [Gershman and Niv, 2010]. This idea has been investigated in [Gershman et al., 2009], where the authors tested participants on a reinforcement learning task that required both hands. The authors found that if the reward could be decomposed into separate components for each hand, participants exploited this reward-structure to guide their learning and exploration. Another study [Acuna and Schrater, 2010] presented participants with a bandit task where a latent variable coupled the rewards of the different arms of the bandit. Participants' behavior in the task was suboptimal from a pure, reward-maximizing perspective but could be explained with a model that performs Bayesian inference over the underlying coupling structure, which is equivalent to inferring the causal structure of the task.

1.2.3 Structured probabilistic models

Historically, structure learning phenomena were first observed in experimental psychology as facilitated learning in novel variations of previously experienced tasks. The speed-up in learning of new tasks was mainly attributed to the ability to learn a general rule or set of rules that govern the tasks at hand (also referred to as transfer learning or the formation of a learning set). This learning of "hard", explicit rules can be considered a special case of structure learning, but in general structure learning also refers to learning "soft", statistical rules or sets of statistical regularities and constraints, such as means and variances, correlations, lower-dimensional

manifolds or latent, stochastic causes. Signatures of such structure learning can be found in perceptual tasks, for instance when robustly perceiving complex objects [Kersten et al., 2004], when inferring whether multiple sensory inputs are coupled by a common latent cause [Körding et al., 2007, Sato et al., 2007] but also in human sensorimotor learning [Braun et al., 2010a]. However, many studies that report these phenomena do not address the underlying structure learning problem on a computational level. From a theoretical perspective it is highly desirable to shed some light on the computational processes and principles involved, particularly because structure learning might be key to enabling efficient learning and acting as well as the ability to generalize well in high-dimensional, noisy environments. An important theoretical question is whether there is a common, universal computational principle that underlies different structure learning phenomena, ranging from complex perception, inductive reasoning and decision-making to categorization, language acquisition and concept learning.

There is a growing body of literature that argues in favor of universal structure learning mechanisms [Gershman and Niv, 2010, McClelland et al., 2010, Tenenbaum et al., 2011, Braun et al., 2010a], even though there is a dispute about the corresponding computational principles. Most of the proposed computational approaches to structure learning follow one of two strands of computational neuroscience research: structured probabilistic models of cognition [Tenenbaum et al., 2011] versus connectionist models of cognition [McClelland et al., 2010]. The main idea behind structured probabilistic models is that cognition and intelligent behavior can be interpreted on a computational level as statistical inference (Bayesian inference) using structured probabilistic models, typically expressed as graphical models [Koller and Friedman, 2009]. Central to the approach is that both perception but also action can be phrased as inference problems [Knill and Richards, 1996, Kersten et al., 2004, Körding and Wolpert, 2006]. In high-dimensional natural environments these inference problems can easily become intractable and solutions can only be found efficiently if the corresponding hypothesis space is sufficiently constrained, for instance by assuming certain (biased) priors—e.g. the "inductive bias" [Griffiths et al., 2010]—but also by the constraints that are imposed through the topological structure of the probabilistic model [Tenenbaum et al., 2006]. The term "structure learning" has been frequently used in the context of learning causal relationships by inferring the correct topological structure of a Bayesian network (a probabilistic graphical model) [Tenenbaum and Griffiths, 2001b, Gopnik et al., 2004, Gershman and Niv, 2010]. While this type of structure learning can be considered a special case within the framework of learning the structure of a probabilistic model, there are some indications that humans use specialized computational principles for detecting causal relationships, most importantly active interventions [Lagnado and Sloman, 2004] but also other cues such as temporal order of correlated events [Lagnado and Sloman, 2006].

Following Marr's levels of analysis [Marr, 1982], structured probabilistic models primarily address the computational level. However, there is a large body of work that aims to connect these computational ideas all the way down to their neural implementation. In particular, Bayesian inference is the key computational principle behind structured probabilistic models. The universality of Bayesian inference to tackle many different kinds of computational problems that appear in human perception and action has famously led to the formulation of the "Bayesian brain hypothesis" [Knill and Pouget, 2004, Doya, 2007], where the brain is essentially conceptualized as a biological (Bayesian) inference machine. Interestingly, signatures of Bayesian inference in human neural substrate have been found, for instance neural correlates of hypothesis sampling and evidence accumulation [Gold and Shadlen, 2007, Ploran et al., 2007, Liu and Pleskac, 2011]. See [Tervo et al., 2016] for a recent review of work that aims at

bridging the gap between computational principles of probabilistic structure learning and their corresponding neural implementation. While Bayesian models have been successfully applied to describe human and animal behavior in numerous situations, there is some valid criticism to the Bayesian brain hypothesis [Kwisthout et al., 2011, Kwisthout and van Rooij, 2013]. The two most severe shortcomings of the Bayesian framework are (in)tractability [Gigerenzer et al., 2008] and the flexibility of the framework (almost anything could be modeled, given the right priors and likelihood models). The intractability argument refers to the fact that certain computations that arise in the Bayesian framework (e.g. marginalization, which is needed for instance in order to compute the normalization constants) have an exponential complexity in the cardinality of the random variable(s) or the number of random variables. However, improving the tractability of Bayesian computation is a very active field of research and many approaches involving approximation or sampling-schemes exist [Chater et al., 2006, Sanborn et al., 2010]. In neuroscience, *hierarchical predictive coding* [Rao and Ballard, 1999, Friston, 2002, Friston, 2005] has become popular as a framework for empirically tractable Bayesian inference, even though the implications of that tractability are not yet fully understood [Kwisthout and van Rooij, 2013]. The argument that Bayesian computation is so flexible that almost any system could be modeled requires further investigation into low-level brain function. In particular, clear neural correlates of Bayesian operations would need to be identified (substantial progress toward that direction has been made [Liu and Pleskac, 2011, Ploran et al., 2007]).

Connectionist models of cognition are based on the idea that biological brains are complex dynamical systems. Accordingly, behavior is thought to reflect the operation of subcognitive dynamical processes such as the propagation of activation and inhibition across neuron-populations or the adjustment of connection strength between neurons [McClelland et al., 2010]. One of the most prominent computational frameworks in connectionist models is *dynamic field theory* [Thelen et al., 2001, Johnson et al., 2008]. In contrast to the computational, top-down/function-first approach of structured probabilistic models, connectionist models of cognition are driven by a bottom-up/mechanism-first approach that is heavily guided by neurophysiological and connectomic findings. However, there is also work that investigates connectionist models on a computational level, for instance by comparing connectionist models with Bayesian inference [McClelland, 1998]. Connectionism is tightly coupled to the so called "emergentist" point-of-view, where structure observed in human and animal cognition emerges from the (dynamic) operation of the underlying subcognitive processes. Under this point of view, structured probabilistic models of cognition, are often interpreted as symbolic approaches (cognition is modeled directly at the level of manipulating symbols and symbolic structures such as propositions or rules), which would require some kind of pre-specification of the corresponding symbolic structures, thus being somewhat in contrast to the *emergence* of structure. Recent work on *non-parametric hierarchical Bayesian models* (NPHBM) addresses this criticism, as these types of structured probabilistic models become increasingly more complex with more observation-data. NPHBMs have been used to successfully model the emergence of structured behavior and achieve human-level performance on a range of cognitive tasks including concept learning [Lake et al., 2015], causal learning [Griffiths and Tenenbaum, 2006] and parsing motion in the environment [Gershman et al., 2015] (a review is provided in [Tenenbaum et al., 2011]). At this point it is unclear whether the view of cognition as structured probabilistic inference and connectionist models of cognition can be reconciled under a common computational framework. More on the current dispute is summarized in *Trends in Cognitive Sciences'* editorial "Approaches to cognitive modeling" [Kousta, 2010], which features one lead-article from each

approach (McClelland et al. [McClelland et al., 2010] and Griffiths et al. [Griffiths et al., 2010]) that also comments on the companion article, plus additional letters by experts in the field who comment on the articles from different viewpoints.

The work in this thesis is heavily influenced by structured probabilistic models of cognition. The experimental studies of the thesis investigate the role of structured probabilistic models and Bayesian inference in human sensorimotor tasks (Chapter 6) and whether Bayesian principles can explain how humans choose among several learned structures (Chapter 4 and 5). The theoretical part of the thesis (Chapter 7) sheds some light onto how hierarchical probabilistic models could emerge from the structure of the tasks at hand in conjunction with the environment of a system, without the need of pre-specification of the model's particular structure. This way, symbolic representations, that arise as the results of information-efficient computation, become intrinsically grounded in an "emergentist" fashion.

1.2.4 Structure learning in motor control

The facilitation of adaptation to novel but related tasks (learning to learn) has been observed in many studies on human sensorimotor learning [Welch et al., 1993, Bock et al., 2001, Seidler, 2004, Seidler, 2007, Adolph, 2008, Braun et al., 2009a]. This is in agreement with the observation that many motor-tasks exhibit rich structural regularities that can be exploited for efficient motor learning and adaptation [Wolpert et al., 2011]. Similar to learning-to-learn phenomena in perceptual tasks, where intra- vs. extra-dimensional shifts make a critical difference [Trobalon et al., 2003], learning to learn in motor tasks crucially depends on the kind of task-variation. Broadly speaking, task similarity is critical for transfer [Shea and Morgan, 1979, Cohen et al., 2005]. For instance, the authors in [Mulavara et al., 2009] exposed participants to one of two training-tasks: either walking on a treadmill or a balancing task. During training, participants experienced a range of visual distortions (by using distortion glasses). Subsequently, both groups were tested on a walking task that required obstacle avoidance while wearing distortion lenses. The group that was trained on the treadmill task showed facilitated adaptation to the obstacle-avoidance task compared to the group that was trained on the balancing-task, suggesting that task similarity plays a critical role. Interestingly, the authors also found that facilitation was better for the treadmill-group that had worn multiple distortion lenses during training, compared to a control group that performed the treadmill-training with a single distortion lens only. This finding is consistent with other studies which found that increased task variability during practice leads to facilitated retention and transfer of motor skills [Shea and Morgan, 1979, Roller et al., 2001, Abeele and Bock, 2001, Braun et al., 2010a]. Further experimental work seemed to suggest that learning in *highly* variable environments cannot be achieved or that, at best, average responses can be learned in such environments [Karniel and Mussa-Ivaldi, 2002, Wigmore et al., 2002, Davidson and Wolpert, 2003, Yousif and Diedrichsen, 2012]. This view has recently been reconsidered in light of carefully designed sensorimotor experiments [Braun et al., 2009b, Genewein et al., 2015a]. In particular, Braun et al. [Braun et al., 2009b] showed that both components, task-variability but also similarity in task-structure are critical for structure learning in motor tasks. The authors used a virtual-reality setup to expose participants to a reaching-task under different kinds of visuomotor perturbations. During training one group of participants experienced visuomotor rotations of random angles, while the other group was confronted with random linear perturbations (a random combination of rotations,

shearings and scalings). Subsequently, both groups were tested on a rotation-perturbation with a fixed angle. The rotation-group showed significantly facilitated adaptation compared to the random-transformation-group. Importantly, the random-transformation-group did not show an increased performance or adaptation-speed compared to a naïve group that did not experience any perturbations, even though the random-transformation group had experienced many pure rotation-perturbations during training.

On a computational level, the sensorimotor system is faced with two problems: parameter adaptation and structure learning [Wolpert and Flanagan, 2010]. The first problem is finding an optimal set of parameters for the current task at hand, given the learned structure. One popular framework for this component of motor learning is *adaptive control* [Astrom and Wittenmark, 1994]. The second problem that the motor system is faced with is the structure learning problem, that is the problem of inferring the correct structure of a set of tasks. For instance, consider using a robotic (or biological) arm with a golf club attached in order to send a golf ball on a desired trajectory. The structure of such a control task is typically described by equations that describe the dynamics of the arm and the club-ball interaction. Since multiple different clubs can be used, the particular club enters as an (unknown) parameter of the (known) model. Parameter adaptation can then be phrased as an optimization problem (solved for instance by optimal feedback control [Diedrichsen et al., 2010, Diedrichsen, 2007]), which relies upon a structure learning mechanism that is capable of learning the structure of the task at hand. A graphical intuition of the interplay of the two learning problems is given in Figure 1.4. Each axis in Figure 1.4A corresponds to one parameter-variable of the motor-controller (e.g. the synaptic strength between two neurons). Adapting to a novel task (that is a novel golf club) without any additional knowledge requires an exhaustive search through this two-dimensional parameter-space. However if the structure of the task has been learned (the thick black line), for instance after observing that the optimal parameters for many variations of the task (different clubs) lie on a sub-manifold, adaptation to a novel task that shares the same structure can be facilitated by constraining the search to the learned sub-manifold. This reduces the effective dimensionality of the search-problem (from a two-dimensional to a one-dimensional search in this example). The benefits of structure learning become immediately clear: searching over a reduced set of options, in this case by reducing the dimensionality of the search problem, significantly simplifies the search problem, regardless of the specific implementation of the search. In very high-dimensional search spaces, such as synaptic weights for millions of neurons, dimensionality reduction (more generally: restriction of the search space) becomes critically essential. If a reduction is not possible, the well known *curse of dimensionality* (a term coined by Richard E. Bellman [Bellman, 1957]) leads to an exponential explosion of potential parameter-settings that need to be evaluated, thus prohibiting any efficient implementation of a search process.

In the same study as described before ([Braun et al., 2009b]), Braun et al. also found that initial exploration for tasks that do not share exactly the same structure is heavily guided by the learned structure (see Figure 1.4). These results were later confirmed in another sensorimotor experiment by Kobak and Mehring [Kobak and Mehring, 2012]. Structure learning thus provides a basis for efficient exploration. Similarly, Gershman and Niv [Gershman and Niv, 2010] argue that (causal) structure learning and the resulting reduction of the size of the hypothesis space is essential for reinforcement learning in natural environments. Structural knowledge allows "*animals and humans to focus their attention on those objects and events that are key to obtaining reinforcement, and learn only about these while ignoring other irrelevant stimuli*". Additionally, they argue, structural knowledge about how actions affect the environment can be

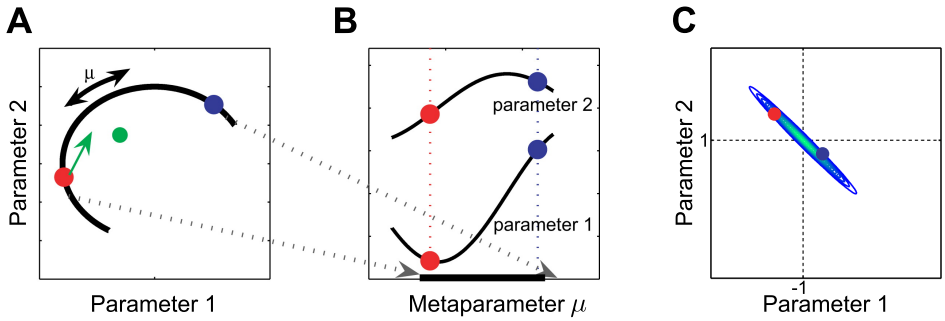

Figure 1.4: Structure learning vs. parameter adaptation. **A** Imagine a controller that should learn how to control a robotic arm in order to swing a golf-club and send a golf-ball onto a desired trajectory. The controller has many parameters and learning corresponds to finding a desirable setting of these parameters. For simplicity, the figure shows only two dimensions the parameter-space. Without any additional knowledge, finding the best set of parameters requires an exhaustive search through the whole parameter-space. However, consider the situation where many variations of the task have been learned already (for instance by varying the length of the golf-club and the size of the club's head) and all these optimal parameters lie on a manifold (the thick black line). **B** Adaptation to a new variation (a novel golf-club) is then greatly simplified, because the search can be reduced to a search along the learned manifold, thus reducing the effective dimensionality of the search space. Even parameters that lie outside the manifold (green dot) might benefit from the manifold as the search strategy can be guided by the learned manifold. Extracting such a manifold corresponds to structure learning because the manifold captures invariants and higher-order regularities of a family of tasks. **C** The structural relationship between two parameters must not always strictly lie on a manifold but can also be captured by a stochastic relationship. In the panel, the two parameter-dimensions are related through a (noisy) correlation—in particular through a two-dimensional Gaussian with a correlation coefficient close to -1 (contours show iso-probability lines and a higher saturation with green indicates a higher probability). Similar to how a manifold constrains the effective search-space, a stochastic relationship between the two parameter-dimensions can be used as a prior to constrain the inference process by placing prior-probability mass in regions that follow the learned higher-order statistical invariants. This, in turn, allows for better inference in light of noisy or sparse data. The figures in panels A and B were first published in Current Biology [Braun et al., 2009b] under the Creative Commons license (https://creativecommons.org/licenses/by/3.0/) and reproduced with permission.

utilized to decompose tasks into smaller and more manageable components, leading to better exploration-strategies for policy-learning and -adaptation.

1.2.5 Neural correlates and mechanisms of structure learning

One challenging open problem in structure learning research is to bridge the gap between high-level computational principles and the corresponding neural hardware that implements these computations. This section gives a very brief overview of such attempts. Below is a list of suggested publications for further reading:

- **Toward the neural implementation of structure learning** [Tervo et al., 2016]. Considerations for identifying neural substrate that represents learned hierarchical structure and structure learning mechanisms from a computational point-of-view, particularly

focused on non-parametric hierarchical Bayesian models (NPHBMs).

- **Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective** [Botvinick et al., 2009] Structure learning mostly in the context of learning abstractions over temporal sequences for hierarchical reinforcement learning (HRL). The paper suggests specific brain regions that the computational framework of HRL might map onto and discusses how the framework might complement existing psychological models of hierarchically structured behavior.

In the structured-statistical-models community, non-parametric hierarchical Bayesian models provide a promising basis for explaining a very broad range of (hierarchical) structure learning phenomena observed in humans and animals. The generality of the framework is very appealing from a computational viewpoint, since many different phenomena are unified under a common principle. The downside, however, is that the specific neural implementation of these different phenomena is very likely to be spread across many different brain regions and, even more challenging, the brain might make use of different implementations of the computational principle. However, there is experimental evidence in face recognition [Tsao, 2014], action planning [Koechlin, 2016] and language processing that suggests that the different layers of an acquired hierarchy tend to be embedded in spatially segregated circuits. This could be a key-advantage for designing experiments since learning at each level of the hierarchy can be observed but also perturbed in the corresponding circuit. One of the main-questions is how to identify neural representations of abstract symbols in structured abstractions. It is highly likely that such variables are encoded at the level of neural ensembles rather than single cells. To complicate matters further, recent experimental work suggests that stable variables need not always be encoded in the form of stable neural ensemble activity patterns [Bathellier et al., 2012] but can also be represented by dynamically varying patterns of ensemble activity [Druckmann and Chklovskii, 2012, Sadtler et al., 2014].

Braun et al. [Braun et al., 2009b] (Supp. Mat.) suggest that the brain and in particular the motor system can be thought of as a controller "with certain dials or 'free variables', for example the synaptic configurations in the motor cortex". These free variables fluctuate but their fluctuation can be adjusted to lie on a manifold suitable for solving a given task [Rokni et al., 2007]. Learning a task corresponds to *constraining* the neural fluctuations (weights in the space of synaptic configurations) to a particular manifold. Learning the structure of several variations of the task or a family of related tasks adds further constraints such that synaptic weights have to be adjusted to match the intersection of manifolds optimal for these tasks. Interestingly this leads to the prediction that more demanding tasks lead to more stable representations because more task constraints reduce the dimensionality of the manifold and thus reduce drift in synaptic strengths [Rokni et al., 2007]. Mc Kenzie et al. [McKenzie et al., 2014] found the existence of a hierarchy of neural manifolds in the context of a transfer learning study, however the precise relationship to task structure remains unclear. In a recent study by Stadtler et al. [Sadtler et al., 2014], the authors trained non-human primates to alter the neuronal dynamics of their primary motor cortices to move a cursor on a screen to a target (using a closed-loop brain-computer interface). This allowed to investigate the precise relation between cursor movements and neural activity in the task. The authors found that the animals could easily adjust their neuronal dynamics along an intrinsic manifold but showed great difficulty learning to produce activity patterns outside the manifold. In the context of hierarchical Bayesian models, the restriction of activity patterns to specific manifolds (or

intersections of manifolds) could be interpreted as the neural instantiation of structured priors. However, further experimental work is required to establish precise relationships between neural activity and abstract task variables.

Some authors relate the phenomenon of *savings* with causal structure learning [Gershman and Niv, 2010]. Historically, savings has been first observed and studied in the context of classical Pavlovian conditioning of motor responses. In eyelid conditioning for instance, a tone (the conditioned stimulus CS) is paired with periorbital electric stimulation (the unconditioned stimulus US). After an *acquisition* phase, the tone, in absence of the electrical stimulation, leads to closing of the eyelid. These conditioned responses can be unlearned during *extinction* training (where the CS is not paired with the US) [Schneiderman et al., 1962]. During *reacquisition*, savings is apparent through much faster re-learning compared to the original learning. The idea is that there is a simple structure learning mechanism where the animal or human learns that there is a latent variable in the environment that governs whether a conditional stimulus and unconditional stimulus are paired or not. Once this causal structure has been learned, only few trials are required to infer the correct state of the latent variable, thus leading to an increased rate of re-learning [Gershman and Niv, 2010] Savings is interesting because it has been studied extensively for decades, including investigations on the level of neural circuitry. However there is some dispute as to whether savings is actually related to structure learning. The authors in [Medina et al., 2001] used a combination of a sophisticated computer simulation of the cerebellum and reversible lesions (in rabbits) to investigate the mechanisms of savings that operate in the cerebellum during eyelid conditioning. Their results suggest that a site of neural plasticity outside the cerebellar cortex, most probably in the cerebellar nucleus, is protected from the full effects of extinction training. The remaining residual plasticity in this area can later contribute to the savings observed during re-learning. The authors further conclude that the phenomenon of savings does not necessarily imply learning about a latent task-variable but could be fully explained by two different sites of plasticity, where one site can be shielded from the effects of extinction training.

Note that the authors of [Zarahn et al., 2008] point out that there are two effects that need to be considered for faster re-adaptation: one, aftereffects which describe a persistence of the adapted state into re-adaptation and two, savings (a faster rate of re-adaptation). For investigating structure learning, only the latter effect is of interest. According to the authors, many experimental studies, particularly in the domain of visuomotor adaptation, mix the two effects to a certain degree by not ensuring that the starting states for initial adaptation and re-adaptation are identical. To ensure this, experimenters can either *wash out* the aftereffects with a sufficient number of zero-perturbation trials (or unpaired trials in case of a conditioning experiment) or insert *counter-perturbation trials*.

Another instance of structure learning for acting and decision-making where the neural basis has been partially investigated is hierarchical reinforcement learning (HRL) [Botvinick, 2012]. Reinforcement learning is perhaps one of the most successful instances of a transfer from machine learning to neuroscience and computational psychology. For instance, the temporal-difference (TD) learning paradigm provided a framework for interpreting temporal profiles of dopaminergic activity [Botvinick et al., 2009] (encoding a kind of prediction error of reward, exactly like the TD-error). One computational problem with reinforcement learning is that it does not scale well, i.e. for large task domains with many world states and possible actions, reinforcement learning algorithms quickly become intractable ("the curse of dimensionality"). One approach

to tackle this shortcoming is the use of temporal abstractions (also known as options, skills, operators or macro-actions). In this framework, sets of interrelated (atomic) actions are sequentially grouped together into more abstract macro-actions. This naturally leads to a hierarchy of temporal abstractions, hence the name hierarchical reinforcement learning. These temporal abstractions have also been observed in experimental psychology when investigating human and animal goal-directed behavior and have been linked to prefrontal cortical function [Botvinick, 2008]. The authors in [Botvinick et al., 2009] investigate how one particular implementation of HRL, based on an actor-critic architecture with options, could map onto brain regions and fit with the model's predictions about connectivity structure. In an actor-critic architecture, the actor selects optimal actions based on the current state and a corresponding internal estimate of the value-function. The critic adjusts the value-estimates based on the actually observed reward. Both, the actor and the critic use the temporal difference error (reward prediction error) to update either the policy or the value-function estimate. Importantly, actor-critic architectures with TD error updates have been related to neural structures by some researchers before (see [Joel et al., 2002] for a review). One proposal is to identify the actor with the dorsolateral striatum and the critic with the ventral striatum and the mesolimbic dopaminergic system. The reward prediction error (TD error) is conveyed through dopamine. See [Botvinick et al., 2009] on how HRL in an actor-critic architecture with options predicts extensions on the basic actor-critic scheme that can partially be mapped to other neural structures (dorsolateral prefrontal cortex, pre-supplementary motor area) and importantly leads to predictions for behavioral experiments but also predictions for connectivity patterns between the areas involved.

1.3 Setting and scope of the thesis

Bayesian inference has been proposed as a universal computational mechanism that quantitatively explains behavior and phenomena across a variety of cognitive and sub-cognitive tasks that the human brain is faced with ("Bayesian brain hypothesis" [Knill and Pouget, 2004, Doya, 2007]). A natural question is thus whether and how the structure learning problem can be cast within the Bayesian framework. In cognitive neuroscience, hierarchical probabilistic models have been found to be very suitable for capturing learned structure in the form of (structured) prior distributions, potentially on multiple levels of a hierarchy, and human behavior was found to be quantitatively consistent with Bayesian inference processes in such hierarchical models [Lee and Mumford, 2003, Lucas and Griffiths, 2010, Tenenbaum et al., 2011]. When work on this thesis began in 2012, structure learning and hierarchical Bayesian inference in the human motor system had not been extensively investigated from a computational point-of-view. While many learning-to-learn phenomena in sensorimotor tasks had been reported, structure learning was mostly attributed to higher-level, cognitive processing. Among others, Wolpert et al. started in the previous decade to heavily investigate the idea that the human motor system performs Bayesian computations on a low, sub-cognitive level. One famous result is their 2004 paper [Körding and Wolpert, 2004], where they showed that the human sensorimotor system integrates prior knowledge with feedback uncertainty in a statistically optimal (Bayesian) fashion. Around 2010 Wolpert, Mehring, Braun and Turnham (and others) conducted a series of experiments that showed that humans are capable of extracting and exploiting non-trivial structure in highly-variable sensorimotor tasks. Importantly, they proposed and quantitatively

investigated (hierarchical) Bayesian inference as a computational explanation of the phenomena observed [Braun et al., 2009b, Braun et al., 2009a, Braun et al., 2010b, Turnham et al., 2011]. The starting-point of this thesis is a direct continuation of these ideas. In particular, the tasks used in the previously mentioned studies could not fully answer the question whether the observed structure learning phenomena are indeed due to low-level, sub-cognitive motor processing. Additionally, the studies (except [Turnham et al., 2011]) did not use noisy or ambiguous feedback, which requires a (Bayesian) integration of learned prior knowledge with the observed feedback and would thus allow to test whether humans are capable of tuning sensorimotor integration processes to the structure of the environment. Turnham et al. could show that humans can alter their prior-beliefs over sensorimotor transformations thus altering the prior used for sensorimotor integration. However, in their task sesnsorimotor integration was required between pairs of trials which allows for rather long time-scales and potentially higher-level cognitive processing as part of the feedback integration process. In contrast, one goal for this thesis was to study structure learning in sensorimotor integration with a task that requires rapid, on-line feedback integration within a single trial.

Another question that had received little attention so far, but is undoubtedly important, is how humans select among learned structures when faced with a novel task or ambiguous feedback that is compatible with multiple structures. A promising idea is to treat the question as a model selection problem and, correspondingly, apply the framework of Bayesian model selection to the problem. In particular, this idea lead to the design of experimental studies that test whether Bayesian model selection can quantitatively describe human behavior in sensorimotor model-selection tasks.

The third issue, which tackles a large, open question in neuroscience is how structured representations *emerge* from interaction with the environment and whether there is a common underlying computational principle. While this question has been investigated heavily from a mechanistic point-of-view in connectionist models of cognition [McClelland et al., 2010], the structured-probabilistic-models community has started to explore this question only recently, mostly by suggesting nonparametric hierarchical Bayesian models (NPHBMs) [Tenenbaum et al., 2011] that grow in complexity in a purely data-driven fashion. An interesting observation from a computational viewpoint is that learning structure allows to capture abstract, transferable knowledge, thus structure learning can be interpreted as a process of abstraction that extracts general invariants. In order to form abstractions, relevant structural information must be separated from irrelevant information. There are interesting ideas and insights on the problem of separating structure from noise originating from physics and information-theory (particularly lossy compression)—the details are omitted from this high-level overview, see more on the topic in Section 2.3. The idea that *'learning is compression'* and the resulting connection to information theory (which itself has deep ties to statistical mechanics) is not new and dates back to the early days of cybernetics [Wiener, 1948], pioneered by Wiener, Shannon, van Neumann and others. It is appealing to study conceptual similarities and theoretical overlap when viewing structure learning as (lossy) compression.

Research goals

The aim of this thesis is to contribute towards a **quantitative** understanding of the **computational** models and principles underlying structure learning, that is the extraction and transfer

of higher-level statistical invariants, with a focus on **human sensorimotor processing**. In particular,

1. to investigate whether and how **Bayesian inference** principles in conjunction with **hierarchical probabilistic models** can explain structure learning phenomena in sensorimotor tasks with ambiguous feedback, which requires integration of the feedback-information with prior-knowledge. The goal is to investigate whether and how such sensorimotor integration processes can be tuned to the (statistical) structure of the task(s) at hand.
2. to answer the question of how humans, when faced with a novel problem, **select among several different structures** that have been learned previously and whether this question can be cast and answered within the Bayesian framework.
3. to shed some light on the principles that lead to the formation of structured knowledge representations. What are theoretical benefits of structure learning with hierarchical Bayesian models? Can the intuition that the learned structure reduces the size of the search-space over parameters or hypotheses be further formalized and, importantly, can the process of extracting invariants be viewed from a compression point-of-view?

2 Results and Discussion

2.1 Sensorimotor structure learning

The major idea behind structure learning is that the learned structure constrains inference processes such that the set of candidate-hypotheses that need to be evaluated becomes significantly reduced in size, due to the structural restrictions. This, in turn, allows for more efficient information processing, often rendering intractable, high-dimensional inference problems tractable, especially in the light of sparse, ambiguous data. An illustration of this idea is shown in Figure 1.4 where the learned structure is a sub-manifold in potentially high-dimensional data. The sub-manifold describes optimal parameter-settings for many variations of the same task or a family of related tasks. Once the structure is known, adaptation to a novel task is facilitated because the search for the optimal parameter-setting can be constrained to the lower-dimensional manifold.

2.1.1 Experimental design and findings

In the experimental study shown in full detail in Chapter 6, the idea that the structure of a task can be expressed as regularities in parameter space is translated into a sensorimotor task. The goal is to test whether humans are capable of learning and exploiting structural invariants and whether this can be modeled with a hierarchical Bayesian model. In particular, the question is whether humans can tune their low-level sensorimotor integration processes to the structure observed in the environment. In the study, we designed a sensorimotor reaching task in 3D virtual reality (see [Genewein and Braun, 2012] for details of the setup). Participants had to perform a reaching movement with a virtual cursor, representing their hand-position, from a start-position to a fixed target. After movement onset the virtual cursor was shifted horizontally and vertically by a random distance in the plane that is perpendicular to the forward-backward direction of the movement. Participants did not observe this shift, nor did they see their cursor during movement, but they were shown brief visual feedback (150ms) about half-way through the movement such that they could adapt to the visuomotor shift and hit the target with the shifted cursor. Crucially, the shifts followed higher-level statistical regularities that could be learned and exploited to improve the chances of hitting the target.

Following the main-idea of the illustration in Figure 1.4, where the structure of the task forms a one-dimensional manifold in the two-dimensional parameter space, we introduced structural invariants to our task by tying the two dimensions of the visuomotor shift, see Figure 2.1. In particular, the shifts in each trial of our task were drawn from a fully correlated two-dimensional Gaussian. Additionally, we introduced a control group that performed the same experiment under an uncorrelated Gaussian distribution over the shifts with the same mean and variance (per dimension) as the correlated group. To allow participants to learn the structure

of the distribution over the shifts we used *full-feedback* trials, where the visual feedback of the shifted cursor shown to participants consisted of a small sphere, thus providing precise feedback in both dimensions of the shift. To test for the learned structure that participants were using, we introduced *partial feedback* trials, where the visual feedback consisted of a bar-like stimulus that gave precise feedback in one dimension but was completely uninformative in the other dimension as the bar covered the whole workspace width or height respectively. If the correlation-structure was perfectly known, having precise feedback about one dimension of the shift is sufficient to compensate both dimensions of the shift and be able to hit the target accurately. In contrast, if the correlation structure is not known to participants, the corrective movement in the uninformative dimension can at best rely on simple statistics by compensating for the mean-shift that has been observed for the corresponding dimension. Importantly, while the uncorrelated group was trained in absence of the correlation structure, the test-trials with partial feedback were identical to the test-trials in the correlated group, that is shifts were drawn from the correlated 2D Gaussian. Humans have been shown previously to be able to learn the mean of a Gaussian visuomotor shift and use this mean for compensation in the absence of visual feedback information. To test for this and investigate the speed of learning simple statistics (mean shift) compared to higher-level invariants (correlation structure) we introduced *no-feedback* trials that provided no visual feedback such that participants had to rely on their prior expectations.

By comparing the performance of the correlated and uncorrelated group, we found that participants in the correlated group did indeed learn the correlation structure and exploited it for facilitated sensorimotor integration of noisy and ambiguous feedback with prior knowledge in partial feedback trials. However, we found that learning the correlation structure was much slower compared to learning the mean-shift and even after 4000 trials, spread over four sessions on four different days, learning of the correlation was still incomplete for most participants. We could also rule out the possibility that participants learned a cognitive rule (the correlation structure can be expressed as a simple cognitive rule and knowledge of the rule might render the partial feedback trials a lot easier) which suggests that participants of the correlated group were able to adjust low-level feedback-integration processes of their sensorimotor system. If instead they simply discovered a cognitive rule, their learning progress would not be expected to be gradual (as observed in the experiment) but there should be an abrupt improvement upon discovering the cognitive rule. Additionally, participants should be able to express such a cognitive rule after completing the experiment which we found not to be the case. In fact, participants were mostly surprised after revealing the remarkably simple cognitive rule to them (after completion of the experiment). To be sure, we tested another control group, similar to the correlated group, with the only difference being that this control group was verbally instructed about the correlation structure before starting the experiment and reminded again before starting the second session. We found that verbal instructions about the cognitive rule did not lead to any performance improvement.

Participants' behavior in the task, in particular their structure learning, was successfully modeled with a hierarchical Bayesian model where we assumed a 2D Gaussian prior-distribution over the shifts with unknown mean and covariance-matrix. Learning in this model corresponds to placing a prior-distribution over the mean-vector and covariance-matrix—the so-called hyper-prior. In our case, we used an inverse normal-Wishart distribution and updated its parameters using an on-line, Bayesian inference scheme based on the observations in each training-trial (full feedback). We could show that such a model could quantitatively explain the final cursor

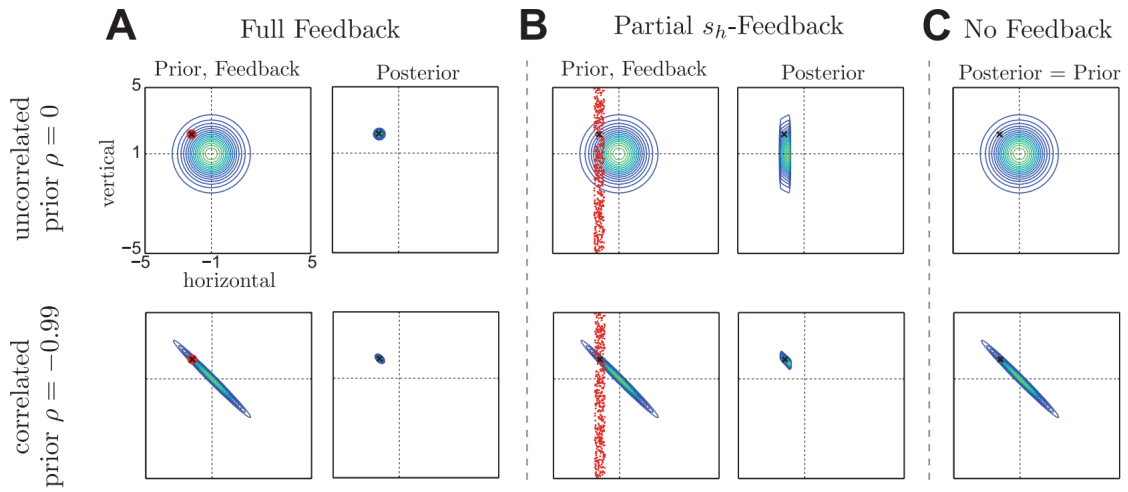

Figure 2.1: Task setup and description of the structure learning experiment, described in full detail in Chapter 6. **A** Full feedback trials (training trials). In each trial, a two-dimensional visuomotor shift is drawn from the distribution shown by the iso-probability lines. Participants performed a blind reaching movement under the visuomotor shift with brief visual feedback halfway through the movement. In this trial type, the visual feedback consisted of a small red sphere, providing precise information about the shift. The corresponding posterior looks almost identical under both a correlated and an uncorrelated prior. **B** Partial feedback trials (test trials). In this trial type, the visual feedback consisted of a flickering horizontal or vertical bar, which provides precise feedback in one dimension and no feedback in the other dimension (since it covers the whole workspace). In these trials, the posterior distributions that integrate prior information with feedback information look very different under the correlated and the uncorrelated prior. Intuitively, if the correlation has been learned, reliable feedback about one dimension of the shift provides the same information about the other dimension of the shift. **C** No feedback trials. These trials were used to assess participants' belief over the mean shift. This figure was first published in PLoS Computational Biology [Genewein et al., 2015a] and is published under the Creative Commons license (https://creativecommons.org/licenses/by/4.0/) and reproduced with permission.

displacements (with respect to the target), observed from participants in the experiment. In particular, we trained the exact same model either on correlated or uncorrelated trials and found the resulting behavior to be very similar to the correlated and uncorrelated groups of participants respectively. By placing a medium prior-weight on a zero-mean vector and a large prior-weight on an uncorrelated covariance-matrix, we could reproduce structure-learning progress of the model with the same time-scale as observed in the correlated group.

2.1.2 Contributions and Novelty

In our experiment, we designed a sensorimotor task that required participants to integrate ambiguous feedback with prior knowledge. Importantly, the task was designed such that the prior knowledge could potentially include rich structural regularities that go beyond simple statistics such as means and variances. We tested whether participants could tune their sensorimotor integration processes to reflect these structural regularities and compared against a hierarchical Bayesian model of computation. With respect to sensorimotor learning, the work of Braun,

Wolpert, Mehring et al. has pioneered the idea that the human motor system is capable of structure learning in highly variable environments and that Bayesian inference in hierarchical models can explain these phenomena on a computational level. In particular, in [Braun et al., 2009b] Braun et al. found that humans can learn the structure of a visuomotor rotation perturbation in a two-dimensional reaching task by experiencing many rotations of different, random angles—a setting that had previously been thought to prevent learning and adaptation. In their work they speculate that if the human motor system was able to learn the structure of the task by learning that there is a latent random variable (the angle of the rotation-perturbation) that governs the task, then Bayesian inference can explain human behavior in the task. Following up on this idea, Braun et al. show in [Braun et al., 2010b] that humans do not just learn simple mappings from sensory input to motor output but instead they simultaneously learn structural invariants that facilitate generalization to novel tasks. Additionally they show that this behavior in their task cannot be explained by regression in conjunction with reinforcement learning, but it can be explained with a hierarchical Bayesian model and probabilistic inference. In [Turnham et al., 2011] Turnham, Braun and Wolpert designed a two-dimensional reaching task under a linear visuomotor transformation that exhibits certain statistical regularities. They measured participants prior-beliefs over the transformation and found that, over the course of the experiment, these prior-beliefs started to reflect the structural regularities of the visuomotor transformations. Participants' behavior was modeled with a Bayesian model by placing a prior-belief over the parameters of the transformation matrix and updating this belief in light of new data.

In contrast to [Braun et al., 2009b] and [Braun et al., 2010b] the on-line visual feedback provided in our study was noisy and potentially highly ambiguous which required participants to integrate the noisy feedback with their prior expectations. The human motor system has been previously shown to be capable of integrating noisy feedback with prior knowledge in a statistically optimal fashion—by performing a Bayesian integration of the two sources of information [Körding and Wolpert, 2004]. In our study we could show that this integration process itself can be tuned to higher-level statistical regularities exhibited by the task. Prior to the experimental studies mentioned above, a common assumption was that structure learning does not happen on a low, sensorimotor-level but was rather attributed to higher-level, cognitive processes. Our our study adds another piece of evidence that structure learning also happens on the sensorimotor-level. In particular, we could show that an explicit, cognitive rule could not explain the structure learning behavior observed in our experiment. Similarly, Kobak and Mehring [Kobak and Mehring, 2012] also tested explicitly for low-level structure learning in a reaching task with either horizontal or vertical perturbations and found that structure learning in their experiment changed involuntary visuomotor reflexes and therefore cannot be exclusively a high-level cognitive phenomenon.

2.2 Model selection

Structure learning is thought as the extraction of higher-level statistical invariants. These invariants can, for instance, be represented by structured probabilistic models (hierarchies of priors). For any system that can extract and store multiple structural models, regardless of whether it is biological or artificial, an immediately following question is: how should the system decide which structural model to use when faced with a novel task or environment?

In probabilistic terms, for some observations y and multiple structural models M, what is the probability of a particular model $M = m$ given the observations y

$$p(m|y) = \frac{p(y|m)p(m)}{p(y)}, \tag{2.1}$$

with $p(y) = \sum_m p(y|m)p(m)$. $p(m)$ is the prior probability for model m and $p(y|m)$ denotes the likelihood function of the model given the data [1].

In some situations it is required to deterministically select one model (for instance when comparing two competing hypotheses to explain the observed data) which leads to so-called *Bayesian model selection* (also known as Bayesian model comparison) [MacKay, 2003]. In Bayesian model selection, two models are compared by their posterior probability ratio (following Eq (2.1))

$$\underbrace{\frac{p(m_1|y)}{p(m_2|y)}}_{\text{posterior odds}} = \underbrace{\frac{p(y|m_1)}{p(y|m_2)}}_{\text{Bayes factor}} \underbrace{\frac{p(m_1)}{p(m_2)}}_{\text{prior odds}}. \tag{2.2}$$

If the ratio of posterior probabilities (left-hand side of Eq (2.2)) is larger than one, m_1 is selected, if the ratio is smaller than one, m_2 is selected. Quite commonly, the prior probabilities for each model $p(m_1)$ and $p(m_2)$ are assumed to be equal, in which case Bayesian model comparison reduces to evaluating the ratio of marginal likelihoods known as the *Bayes factor*. The marginal likelihood (also called model evidence) is obtained by *"integrating out"* the parameters of the model (hence the term, *marginal* likelihood)

$$p(y|m_i) = \int p(y|\theta,m_i)p(\theta|m_i)d\theta. \tag{2.3}$$

The first term, $p(y|\theta,m_i)$ is the likelihood of parameters and model given the data and the second term $p(\theta|m_i)$ is the prior-distribution over model-parameters θ induced by model m_i.

One of the most interesting aspects about Bayesian model selection is that it naturally penalizes overly complex models, thus leading to an automatic regularization that prevents over-fitting without the need for an additional, explicit regularizer (see MacKay [MacKay, 2003] for an excellent introduction and discussion). Intuitively, there are two viewpoints to understand why Bayesian model selection leads to automatic regularization:

- Compared to a simple model, which can only explain a narrow range of observations, a complex model spreads its probability mass over a large range of potential observations. If an observation falls into a region where both, the simple and the complex model can explain the data, the probability mass of the simple model tends to be more concentrated compared to the complex model that must spread its probability mass more thinly. In such a case, the simple model is the more likely explanation of the data.

1 Note that $p(y|m)$ is often called "the likelihood of the data given the model", however this is somewhat sloppy since the data is given and the likelihood is a function of the model-parameter in this case. More formally the likelihood of the model given the data is defined as the probability of observing the data under the model $\mathcal{L}(m|y) = p(y|m)$.

- Since the crucial quantity for Bayesian model selection is the marginal likelihood, *all* parameter-settings of a model are taken into account to compute the average (marginal) likelihood. While a complex model is likely to have very good parameter settings for the given observations (thus leading to a large likelihood), it is also likely to have many bad parameter settings (leading to a very low likelihood). In contrast, a simpler model might not achieve an equally large maximum likelihood, given the data, but might still have a better likelihood *average*. An overly complex model with too many degrees of freedom and too many different parameter-settings will thus be penalized by Bayesian model selection.

The phenomenon that Bayesian model selection naturally prefers simpler models over more complex models, in case both models explain the data equally well, is known as *Bayesian Occam's razor*. Note that by integrating out the parameters, the model comparison only depends on the models given the observed data. This means that the parameter-spaces of the two models can be completely different in terms of dimensionality and domain. Many other frameworks for model or hypothesis comparison cannot handle such cases. On the other hand, this freedom and flexibility comes at the cost of evaluating the integral in the marginal likelihood, Eq (2.3). Often there exist closed-form, analytical expressions for the marginal likelihood. If this is not the case, the integral can be approximated (Approximate Bayesian computation [Toni et al., 2009]), but there are also theoretically well-grounded approximations to the Bayes factor such as the Bayesian Information Criterion (BIC) [Schwarz et al., 1978] or the Akaike Information Criterion (AIC) [Akaike, 1974], which involve the maximum value of the (log) likelihood and the number of parameters of the models.

In cases where deterministic model selection is not required or even undesirable, the marginal likelihood is still a good choice for constructing a stochastic model selection mechanism, because the automatic regularization, that is introduced through the marginalization, can be retained. A common way for such stochastic model selection is to use a softmax-selection rule, which can naturally be extended to more than two models

$$p(m_i|y) = \frac{e^{\alpha p(y|m_i)p(m_i)}}{\sum_j e^{\alpha p(y|m_j)p(m_j)}}, \tag{2.4}$$

where α plays the role of an *inverse temperature* and controls the degree of stochasticity ($\alpha \to 0$ corresponds to uniform posterior probabilities whereas $\alpha \to \infty$ leads to a deterministic selection of the maximum). Stochastic model selection is then implemented by sampling a model according to the posterior probabilities given by Eq (2.4).

2.2.1 Experimental design and findings

To test whether human model selection behavior in sensorimotor structure learning tasks is quantitatively consistent with Bayesian model selection, we designed two experimental studies, presented in full detail in Chapter 4 and Chapter 5 respectively. While the experimental design of the first study focuses on the "integrating out" aspect that arises due to the marginalization in Eq (2.3), the second study focuses on the "Bayesian Occam's razor effect" and explicitly tests for the preference of the simpler model in light of data that can be explained equally well by a simple and a complex model.

Model selection study I: A sensorimotor paradigm for Bayesian model selection

In this study, we designed a reaching task with a one-dimensional visuomotor shift in virtual reality (see [Genewein and Braun, 2012] for details of the setup). Participants performed a reaching movement from a start-position to one of two targets, where a cursor represented their virtual hand position. After movement onset, the virtual cursor disappeared and was shifted laterally (left-right direction). Halfway through the movement brief visual feedback (100ms) was shown to participants, which allowed them to correct their movement in order to compensate for the shift and hit the target. In each trial, a shift was drawn randomly from a distribution $p(s|m_i)$, where m_i indicates one of two models that each induce different distributions over the shifts. The models were randomly sampled with uniform probability in each trial. For instance, in one experimental condition the first model m_1 corresponded to a zero-mean Gaussian with medium variance and would thus mostly lead to small cursor displacements. In contrast, the second model m_2 was a bi-modal mixture of two Gaussians with one mode at $\pm$2.5cm each and a relatively small variance—this model would thus produce large positive and negative shifts with high probability. Crucially though, every possible shift observed could in principle be explained by either model. In training trials, where participants could learn about the statistics of each model, participants could hit one of two targets (with their compensatory movement in the left-right direction). The targets were equally sized and displaced in the up-down direction at the end of the workspace (forward-backward direction). Each target was associated to one of the two models and participants were informed whether they had hit the correct target, corresponding to the model where the observed shift was actually drawn form.

To test for model selection behavior, we introduced ambiguous-feedback trials where the visual feedback did not represent the shifted cursor, but consisted of a bar with a certain horizontal width. Participants were instructed that the actual cursor could be anywhere within the bar-stimulus but could not lie outside the bar (the bar acts as an occluder). Participants were asked to select the model that they thought to be more probable by moving to one of the two targets corresponding to each of the models respectively. Crucially, in these test-trials the targets covered the whole horizontal workspace width such that participants were only required to report their belief about the model but not their belief about the most probable parameter. This design forced participants to "integrate out" all shifts that were compatible with the bar stimulus and then pick the model that they thought was more probable to have produced the stimulus. For instance (with respect to the example from above), if the bar-stimulus was rather short, then the unimodal, zero-mean Gaussian is the more probable explanation whereas for a bar-width of $5cm$, both models were equally likely.

We compared participants model selection behavior with a theoretical model based on stochastic Bayesian model selection (stochastic selection according to softmax-probabilities based upon the marginal likelihoods) in two experimental conditions (where the variance of the unimodal Gaussian was varied between conditions). We found that participants' behavior could be well explained by the theoretical model and we could further rule out some alternative explanations based on simple heuristics.

Model selection study II: Occam's razor in sensorimotor learning

Our second model selection study focuses explicitly on the "automatic regularization" aspect of Bayesian model selection (Bayesian Occam's razor). Informally, Occam's razor states that if two hypotheses explain the observed data equally well, then the simpler hypothesis, that makes fewer additional assumptions, should be preferred. In terms of model selection, different models correspond to different hypotheses and the number of assumptions relates to the models' degrees of freedom (number and domain of model parameters). To test for the preference of simpler models we designed a *sensorimotor regression* task where participants were shown a number of noisy samples of an underlying function and were then asked to draw their best guess of the underlying function. We used the same virtual reality setup as in our other experimental studies (see [Genewein and Braun, 2012] for details on the setup). Importantly, we trained participants on two different generative models of the underlying (noise-free) functions. To do so, we indicated the generating model at the beginning of each training-trial with a color cue. After participants had drawn their regression trajectory, the true underlying function was revealed. To test model selection behavior, we introduced test-trials where the true generative model was unknown to participants and the actual function underlying the noisy observations was not revealed after they had drawn their regression trajectory. In these trials, we inferred which of the two models participants had chosen from their drawn trajectory.

To generate the underlying trajectories but also the noisy samples with two models of different complexity, we used Gaussian processes with squared-exponential kernels with different length-scales as generative models. The length-scale parameter allows to control the model complexity and, intuitively, leads to either smooth trajectories (for simple models that have a large length-scale) or wiggly trajectories (for more complex models with a lower length-scale). Interestingly, the log marginal likelihood for a Gaussian process (GP) has a closed analytic form and decomposes into a data-dependent term and a term that is independent of the data and represents model complexity

$$\log p(m_i|y) \propto -\underbrace{\frac{1}{2} y^\mathsf{T} \Sigma_{\lambda_i}^{-1} y}_{\text{goodness-of-fit}} - \underbrace{\frac{1}{2} \log |\Sigma_{\lambda_i}|}_{\text{model complexity}} , \tag{2.5}$$

where y indicates the (noisy) observations, Σ is the covariance matrix of the GP, which is obtained by evaluating the kernel function (in this case: a squared-exponential kernel) with length-scale parameter λ_i that depends on the model m_i. Note how the complexity term, the log determinant of the covariance matrix, essentially corresponds to the entropy of a multidimensional Gaussian (the entropy of the GP). The GP framework allowed us to design stimuli (noisy observations), where the goodness-of-fit term was equal under both models, such that the difference in marginal likelihood solely depended on the model complexity. In these trials we found participants' behavior to be very consistent with Bayesian Occam's razor, as implemented by Bayesian model selection, which (strongly) prefers the simpler model with the lower model complexity in such a case. Additionally, we compared participants' choice behavior in trials with an arbitrary goodness-of-fit term and found that their behavior could be quantitatively well described by stochastic Bayesian model selection (stochastic softmax based on marginal likelihoods)—thereby confirming our results of the previous model selection study.

We conducted a control experiment to rule out that the preference of the simpler model can

be explained by the reduced physical effort of drawing a smooth versus a wiggly trajectory. Additionally we conducted a control experiment that showed that participants do not simply prefer smooth trajectories, regardless of model complexity. To this end, we changed the generative Gaussian processes such that the simpler model produced trajectories with larger spatial frequency but with little variation between different samples (by using a sinusoid mean for the GP of the simple model). We found that in this setting participants still selected according to model complexity (preferring the simpler model in equal-error trial), even though the corresponding trajectories for the simpler model now implied a higher spatial frequency and also larger physical effort of drawing (compared to the more complex generative model).

2.2.2 Contributions and Novelty

Biological organisms in natural environments are faced with at least two kinds of uncertainty: one, state-estimation based on noisy sensory feedback and two, prediction of sensory consequences of actions. Internal models are thought to play a central role in solving these problems [Shadmehr and Mussa-Ivaldi, 1994, Wolpert and Ghahramani, 2000, Todorov, 2004, Shadmehr et al., 2010, Franklin and Wolpert, 2011, Narain et al., 2013b, Narain et al., 2013a]. While a lot of research has been targeted at how biological organisms learn and adapt the parameters of these internal models, the question of how to select among models has received less attention. Undoubtedly this question is important, particularly in cases where different internal models correspond to different hypotheses or lead to very different predictions. Importantly, the same question also applies to the structure learning problem, where the question is how to select among multiple learned structures where each structure has a range of parameter values that need to be chosen as well. The process of structure learning itself is equivalent to learning the models. Model selection in humans has been studied previously in the context of disambiguating competing explanatory hypotheses, for instance in the so called *ventriloquist problem* [Sato et al., 2007, Körding et al., 2007] where humans have to discriminate whether a visual and an auditory signal stem from one common source or from two different sources. Learning of the individual causal models (corresponding to the structure of the task) is commonly phrased as an inference problem over potential causal structures [Körding et al., 2007, Gershman and Niv, 2010], whereas selecting between the different models (common cause versus different causes) in a novel instance of the task is a problem of model selection. In the context of motor learning, it has been argued before that the human ability to generate accurate and appropriate motor behavior under many different and often uncertain environmental conditions can hardly be explained by a "single controller", which would need to be highly complex in order to allow for all possible scenarios. Rather, modular approaches were proposed in which multiple controllers coexist, each controller suitable for a small set of contexts (similar to a mixture of experts known from supervised learning). One concrete framework for such a modular architecture is the MOSAIC model for sensroimotor learning and control [Haruno et al., 2001]. The MOSAIC model consists of several modules that each predict the next sensory state with a forward model but also contain an inverse model to produce a control signal in order to reach a desired sensory state. The different modules (and their corresponding control signals) are weighted by their (forward) prediction error, more precisely by the likelihood of each model given the (new) state of the system. In particular, [Haruno et al., 2001] propose to normalize these likelihoods with a softmax function, similar to Eq (2.4). The paper also suggests (in accordance with findings from psychophysics) to include prior contextual information through so-called responsibility

predictors that play the role of prior-weights in the softmax function. Such a softmax selection scheme is equivalent to the scheme used in our experiments. In a broader context, a softmax-temperature of one (as used by the authors of the previously mentioned publication) directly corresponds to Bayes' rule, thus adding to the importance of Bayesian computational principles in sensorimotor learning and decision-making.

In our first model selection study, we investigated Bayesian model selection in the context of a sensorimotor integration task where feedback uncertainty (resulting from ambiguous visual feedback) had to be integrated with prior knowledge (prior expectations over the visuomotor shift). We designed a paradigm that explicitly requires "integrating out" all model-parameters that are compatible with the observed stimulus by asking participants to report their belief about the model but not their belief over the most probable parameter. Previous sensorimotor integration tasks either focused on parameter adaptation under a single model or required participants to report their belief about both, the model and the most likely parameter simultaneously. In our second model selection study, we focused on the intrinsic regularization properties of Bayesian model selection (Bayesian Occam's razor). The standard textbook task to illustrate model complexity and over-fitting is (polynomial) regression [Bishop, 2006], where a curve must be fit to a set of noisy observations. The main complication is that increasingly complex models will lead to an increasingly better fit on the observed data, but at the same time overly complex models are very likely to have low predictive power for novel data-points. As explained previously, Bayesian model selection naturally penalizes overly complex models, thus avoiding over-fitting without the need for explicit, additional regularizers. We translated this textbook task of model selection into an elegant sensorimotor paradigm (using Gaussian processes instead of polynomials), which, to the best of our knowledge, has not been done before. We could explicitly test for model preference in situations where a simple and a complex model explained the same data equally well and, importantly, our notion of "equally well" was mathematically firmly grounded. Additionally, the experimental task was backed up by a theoretical model whose predictions could be quantitatively compared to human behavior. Note that function learning has been studied previously (see [Lucas et al., 2015] for a recent review). Classically, studies have investigated how people learn relationships between continuous variables, such as linear, quadratic, cubic, exponential or circular relationships. Our work differs from these studies because in our task participants had to learn families of functions (each Gaussian process corresponds to a distribution over functions). Importantly, the main point of our study was not to test for function learning but for model selection, in particular for the preference of simpler models in ambiguous situations. In [Narain et al., 2014] Narain et al. trained participants in a sensorimotor task on several functional relationships (the function governed a temporal relationship between a stimulus cue and the appearance of a target that had to be hit). In their task they found that participants were able to learn several different kinds of functional relationships (linear, quadratic and cubic) and interpolate well in regions of the function where no data had been observed before. In ambiguous situations, humans selected the simplest fitting function, thereby showing a signature of Occam's razor similar to the behavior observed in our experiment.

2.3 Extraction of invariants as lossy compression

This thesis adopts the viewpoint that structure learning corresponds to the extraction of higher-level (statistical) invariants in order to cope with noise and ambiguity more efficiently. Learning structure thus becomes the process of forming abstractions that allow to generalize well in light of novel data. Consider the distinction between models and parameters described in the previous section on sensorimotor model selection: ultimately the goal is to find optimal parameters, but each model captures regularities over the parameter-space by inducing prior distributions over parameters. The process of finding optimal parameters can thus be separated into finding the correct model and then finding the best set of parameters under a given model. The computational advantage of such a procedure is that finding the optimal model is typically a search problem over a small number of alternatives. The subsequent search over optimal parameters can often be greatly simplified by a model, since good models can significantly narrow down the search space over potential parameters. A complex search problem can thus be divided into multiple search problems of (potentially significantly) lower complexity. This is particularly interesting in the context of *bounded rational* agents, that is agents that have to act optimally under limited computational resources. Structure learning is then the problem of learning good models (good prior distributions over parameters).

In the third part of the thesis (Chapter 7), the idea that structure learning is related to the formation of abstractions and reduction in computational complexity is investigated from an information processing point-of-view. In particular, structure learning is viewed from an information-theoretic angle, where the formation of abstractions is formalized under the framework of lossy compression (rate-distortion theory). In a nutshell (explained in more detail in the following sections), structure learning is the problem of separating structure from variation (or noise) which can be viewed as essentially the same problem as in lossy compression, where the goal is the separation of relevant from irrelevant information. Interestingly, this information-theoretic viewpoint also allows to formalize the idea that structure learning allows for more efficient information processing (by reducing computational complexity) and thus allows agents with limited computational capacities to perform well.

The next sections first introduce an information-theoretic framework for bounded rational decision-making with close ties to thermodynamics and statistical mechanics. Then, the connection to rate-distortion theory (lossy compression) is shown mathematically and discussed conceptually. The result will be an optimization principle for decision-making that leads to the emergence of natural abstractions. Finally, this principle is extended to perception-action systems and decision-making hierarchies. In particular the optimization principle that results from the latter extension leads to the emergence of hierarchies of abstraction that can be directly mapped to the model-parameter case described before.

Contributions and Novelty

The mathematical and conceptual connection between the free-energy principle for decision-making [Ortega and Braun, 2013] and rate-distortion theory in the light of the formation of (behavioral) abstractions has been first reported by the author of this thesis at the NIPS 2013 workshop on *Planning with Information Constraints* [Genewein and Braun, 2013] (this work was extended in [Genewein et al., 2015b]). Later, the same connection was presented in S. Still's

Lossy is lazy [Still, 2014]. The idea that biological organisms as well as artificial agents can be cast as information processing machines, that are bound by fundamental limits of information theory, dates back to the early days of Cybernetics [Wiener, 1948]. Similarly, the idea of conceptualizing a decision-maker as an information theoretic channel (from observations to actions) with limited capacity was pioneered by Sims under the name of *rational inattention* (see for instance [Sims, 2003]). Rational inattention is a straightforward application of rate-distortion theory and is thus conceptually very similar to the work presented in the following sections (based on [Genewein and Braun, 2013, Genewein et al., 2015b]). However, the interpretation in terms of the formation of natural levels of abstraction that are induced through limitations in channel capacity, reflecting the structure inherent in the utility function of the task, has not been reported before. There are strong ties to the work of Tishby, Polani, Rubin et al. (the information bottleneck method [Tishby et al., 1999], hierarchical behaviors: getting the most bang for your bit [Van Dijk et al., 2009], trading value and information in MDPs [Rubin et al., 2012], the information theory of decisions and actions [Tishby and Polani, 2011], G-learning [Fox et al., 2015]) but also to the work of Still and Crutchfield [Still and Crutchfield, 2007], Kappen (KL-Control [Kappen, 2007], [Kappen et al., 2012]), Path Integral Control [Braun et al., 2011] and Relative Entropy Policy Search [Peters et al., 2010, Daniel et al., 2012].

The basic rate-distortion principle for information-theoretic bounded rationality was extended in [Genewein et al., 2015b] (see Chapter 7) towards more complex systems involving multiple random variables. These extensions lead to interesting optimality principles whose analytical solutions provide insights and predictions for the bounded-optimal design of perception-action systems and decision-making hierarchies with multiple levels of abstraction (summarized in Section 2.3.2). Both, the derivations of the analytical solutions but also the conceptual interpretations in terms of decision-making, backed by illustrative example simulations, have not been reported before.

2.3.1 Information-theoretic bounded rationality

What is bounded rationality?

In classical decision-making, the goal is to find an action a^* that maximizes a utility function $U(a)$

$$a^* = \arg\max_a U(a).$$

A decision-maker, or agent, following this equation is termed a *rational* agent. In stochastic environments, where the actual outcome of an action can differ from the desired outcome or where the utility-function itself is stochastic, the goal is to maximize expected utility, leading to the well-known *maximum expected utility* (MEU) principle [Ramsey, 1931, Von Neumann and Morgenstern, 1944, Savage, 1954]. In such a setting, the optimal policy is in general also stochastic (expressed as $p^*(a)$), including deterministic actions selection as a special case. In the following, the case of optimal decision-making *with context* is considered. In particular, a context, or world state w is given and the goal is to find the optimal action a in response to that world state

$$a^*_w = \arg\max_a U(w,a). \tag{2.6}$$

The problem with MEU as described by the previous equation is that for many problems, the action-space is too large to be evaluated. An agent that is bound by limited computational capacity can often not afford to search through the whole action-space in order to (deterministically) pick the single best action. However, within its computational limitations, an agent can still try to find the best possible action (even though it is likely to be globally suboptimal). Following the pioneering work of Simon [Simon, 1955, Simon, 1972] such an agent is called a *bounded rational* agent. Consider, for instance an artificial or human chess-player: a brute-force search over possible moves given the current state of the board easily leads to a computational explosion, since the utility of a move can only be evaluated by taking all possible future games and their outcomes into account. Instead, bounded-rational agents use clever search-schemes, approximations and other short-cuts such as exploiting partial similarity of game-states in order to find a good next move. While such agents are not guaranteed to always find the optimal move in each situation, they can still find very good moves, given their limited computational resources. This is, perhaps, the main intuition behind bounded rational decision-making—the goal is not to try and find the single best action for a given situation but rather to find a "good enough" solution that can actually be computed. In so called *needle-in-a-haystack* problems, where only one single action is acceptable, a bounded rational approach is not helpful. However, many natural tasks, including most sensorimotor tasks, do not suffer from such a degeneration—yet, in many classical optimization frameworks these tasks are often treated as if there was only one solution that needs to be found. For instance, consider a human grasping a cup of tea from a table. While there might be one particular trajectory (or set of parameters for such a trajectory) that leads to the highest possible utility (safe grip, lowest energy consumption, etc.), there will also be many other solutions that are slightly suboptimal but still satisfying. In such a situation the bounded rational strategy of "picking the first solution that is good enough" can lead to critical reductions in computational demand while at the same time only having slight impact on the performance of the agent.

One of the main challenges behind the idea of bounded rational decision-making is how to formalize the idea mathematically, in particular how to formalize the notion of computational limitations. Most of the common approaches fall into one of two categories:

- **Implicit boundedness:** the decision-maker does not reason about its computational limitations and acts according to some internal principle. The limitations only become apparent from an external point-of-view. For instance, approximate computation schemes and heuristics fall into this category. One of the main challenges is how to formally describe such a decision-maker.
- **Meta-reasoning:** The agent reasons about its computational limitations and adjusts its decision-making strategy accordingly. One problem is that the meta-reasoning process itself is also subject to the computational limitations of the agent which requires another level of "meta-meta-reasoning", which is again subject to computational constraints and thus leads to an infinite depth of meta-levels.

The following section introduces an information theoretic framework for bounded rational decision-making that explicitly formalizes the notion of computational resources and leads to an optimization principle which describes implicitly bounded decision-makers. It can be used to compute bounded-optimal policies that are exhibited by such decision-makers. Further, the principle can also be used to derive a sampling scheme for designing an implicitly bounded

decision-maker, where the internal boundedness of the agent is given by a limitation in the number of samples.

Information-theoretic bounded rationality

The information-theoretic framework for bounded rational decision-making and inference introduced in this section is presented and discussed in detail by Ortega and Braun [Ortega and Braun, 2010, Ortega and Braun, 2013, Ortega et al., 2015]. It is strongly related to recent advances in the information theory of perception-action systems [Todorov, 2007, Still and Crutchfield, 2007, Still, 2014, Friston, 2010, Peters et al., 2010, Daniel et al., 2012, Daniel et al., 2013, Kappen et al., 2012, Tishby et al., 1999, Tishby and Polani, 2011, Tkačik and Bialek, 2014]. The fundamental premise is that any change of behavior incurs computation. Since the computational resources of a bounded rational agent are limited, change in behavior is also subject to these limitations. The important question is how to quantify change of behavior. In the information-theoretic framework, change in behavior is measured as the amount of adaptation from a prior behavior $p_0(a)$ (before observing) to a posterior behavior $p(a|w)$ (after observing w and performing deliberation). Formally, the amount of adaptation is measured with the Kullback-Leibler (KL) divergence between prior and posterior. The limitation in computational resources of a bounded rational agent can thus be mathematically formalized as an upper-bound on this KL divergence.

The goal of a bounded rational decision-maker is to maximize expected utility under the posterior policy $p(a|w)$, while at the same time not exceeding the KL-bound. Formally, this leads to the following constrained optimization problem

$$p^*(a|w) = \arg\max_{p(a|w)} \sum_a p(a|w)U(w,a) \quad \text{s.t.: } D_{\mathrm{KL}}(p(a|w)||p_0(a)) \leq K, \tag{2.7}$$

where D_{KL} denotes the KL divergence and is upper-bounded by some constant K. The constrained problem can be turned into an unconstrained optimization problem

$$p^*(a|w) = \arg\max_{p(a|w)} \underbrace{\sum_a p(a|w)U(w,a)}_{\text{expected utility}} - \frac{1}{\beta} \underbrace{D_{\mathrm{KL}}(p(a|w)||p_0(a))}_{\text{computational demand}}. \tag{2.8}$$

The bounded-optimal policy $p^*(a|w)$ is the result of trading-off a large expected utility against low computational demand. The Lagrange multiplier β translates the computational effort (in bits or nats) into a cost of computation (in the same units as the utility function, utils). Because of the strong mathematical and conceptual similarities of the optimization problem to the minimization of a *free energy* difference in thermodynamics, β is referred to as *inverse temperature* (more details on the connection to thermodynamics are in [Ortega and Braun, 2013]). Note that different values of β correspond to different values of the upper bound K and thus the computational resources of the bounded rational agent are governed by the parameter β.

The unconstrained optimization problem in Eq (2.8) has a closed-form solution (also well known from thermodynamics)

$$p^*(a|w) = \frac{1}{Z} p_0(a) e^{\beta U(w,a)}, \tag{2.9}$$

where the normalizing constant $Z = \sum_a p_0(a)e^{\beta U(w,a)}$ is known as the partition sum. The inverse temperature β acts as a rationality parameter as it governs the trade-off between large expected utility and low computational demand:

- $\beta \to 0$: no computational resources, posterior policy is equal to prior policy (see Eq (2.9)).
- $\beta \to \infty$: unlimited computational resources (KL term drops in Eq (2.8)), fully rational MEU agent is recovered as a special case and deterministically picks globally best action according to Eq (2.6).
- $0 < \beta < \infty$: bounded rational agent trades off expected utility against the cost of computation

The solution to the optimization problem (Eq (2.9)) can be translated into an elegant rejection-sampling scheme to construct an implicitly bounded agent [Ortega et al., 2015]. According to the scheme, the agent proposes actions according to $p_0(a)$ and then (stochastically) tests them in a rejection step against a fidelity constraint. Importantly, acceptance probability is governed by β which thus directly controls the (expected) number of samples that the agent can afford (higher β means more computational resources which means more samples). Informally, the sampling scheme translates into: search through the action space according to $p_0(a)$ and accept (with high probability) the first action that satisfies the fidelity constraint. Such an agent is not explicitly concerned with not to violating the KL constraint and does not reason about its own computational limitations. Remarkably, the scheme can be shown to produce samples exactly from the bounded-optimal posterior policy $p^*(a|w)$, which minimizes the bounded rational trade-off given by Eq (2.8) that explicitly takes into account limited computational resources through the KL constraint.

So far, the information-theoretic framework for bounded rationality has been introduced in the context of decision-making. However, inference can also be cast as a special form of decision-making, where the utility is a (log) likelihood and a "prediction" plays the role of an action. The framework can thus be used to tackle both, decision-making but also inference problems. Assume that a log-likelihood function is used as the utility $U(w,a) = \log q(w|a)$. Plugging this into Eq (2.9) and setting $\beta = 1$ yields Bayes' rule as a special-case solution:

$$p^*(a|w) = \frac{p_0(a)e^{\log q(w|a)}}{Z} = \frac{q(w|a)p_0(a)}{\sum_a q(w|a)p_0(a)}.$$

With values of β smaller than one, agents become more conservative in the sense that after observing a new data-point, the posterior is moved away less from the prior compared to the $\beta = 1$ case (less KL-divergence). On the other hand, values of β larger than one lead to an accelerated discrepancy between prior and posterior as data comes in. For $\beta \to \infty$ the *maximum likelihood* principle is recovered.

The optimal prior leads to lossy compression

The basic free energy principle of bounded rationality (Eq (2.8) formalizes a fundamental trade-off between large expected utility and low computational cost. The principle assumes a prior or initial policy over actions $p_0(a)$ (before observation and computation). Through computation the prior is transformed into a posterior policy. One interesting question is: what is the optimal

prior $p_0(a)$? This question is answered in full depth in Chapter 7. The answer and the resulting consequences are outlined in this section.

When considering a single world-state w, the optimal prior $p_0(a)$ is trivial to find: it is given by the optimal posterior $p^*(a|w)$ (that deterministically maximizes utility). A more interesting question is which distribution over actions a is optimal *on average* across all w? This question leads to the following optimization problem

$$\arg\max_{p(a|w),p_0(a)} \sum_{w,a} p(w)p(a|w)U(w,a) - \frac{1}{\beta}\sum_{w} p(w)D_{\mathrm{KL}}(p(a|w)||p_0(a)). \tag{2.10}$$

The optimization over the prior can be solved independently and is given by the marginal:

$$p_0^*(a) = \sum_{w} p(a|w)p(w) = p(a). \tag{2.11}$$

Plugging this result back into Eq (2.10) finally leads to

$$p^*(a|w) = \arg\max_{p(a|w)} \underbrace{\sum_{w,a} p(w,a)U(w,a)}_{\text{expected utility}} - \frac{1}{\beta} \underbrace{I(W;A)}_{\text{computational demand}}, \tag{2.12}$$

where $I(W;A)$ denotes the *mutual information* between the random variables W and A. The mutual information is defined as the average KL-divergence between $p(a)$ and $p(a|w)$. The new optimization problem given by Eq (2.12) again formalizes a trade-off between high expected utility and low computational demand, however, the computational demand is now measured through the mutual information (an average KL-divergence). The inverse temperature β still plays the role of translating computational effort into a *cost of computation* and thus governs the trade-off between the two terms.

The variational problem for bounded rational decision making given by equation (2.12) turns out to be mathematically equivalent to the variational problem in rate-distortion theory [Cover and Thomas, 1991, Tishby et al., 1999], the information-theoretic framework for lossy compression. In rate-distortion theory, the goal is to transmit information over a communication channel of limited capacity. If the incoming information-rate is larger than the channel capacity (measured by the mutual information), some of the information must be discarded, leading to a loss of information. A distortion function $d(w,a)$ measures the distortion between the input symbol w and the output symbol a and quantifies how well the output resembles the input. An ideal, lossless communication channel can achieve zero-distortion, whereas a channel with insufficient (limited) capacity can at best minimize distortion. This leads to the familiar optimization problem where expected distortion is to be minimized subject to the constraint that the mutual information is upper-bounded. By realizing that the utility function in Eq (2.12) plays the role of a negative distortion (which leads to a maximization instead of a minimization and to a flip in the sign of β), it becomes apparent that the two problems are mathematically equivalent. However, there is also a deeper, conceptual connection: a bounded rational agent must process information about w in order to act well. Since its computational resources are limited, the agent cannot afford to fully process w and must discard some information. The goal then becomes, exactly as in lossy compression, to discard the most irrelevant information and process the most relevant information. Thus, bounded rational decision making and lossy compression can be

viewed as two sides of the same coin. This interpretation of rate-distortion theory has been explored before under the name *rational inattention* [Sims, 2003, Sims, 2010]—however, it has not been linked to structure learning.

Similar to the free-energy case in the previous section, the optimization problem in Eq (2.12) has a closed-form solution—in the form of two self-consistent equations:

$$p^*(a|w) = \frac{1}{Z} p(a) e^{\beta U(w,a)} \tag{2.13}$$

$$p(a) = \sum_w p^*(a|w) p(w), \tag{2.14}$$

with the partition sum $Z = \sum_a p(a) e^{\beta U(w,a)}$. The solutions can be obtained from arbitrary initializations by iterating the two equations until convergence—this scheme is well known in information theory as the Blahut-Arimoto algorithm [Blahut, 1972, Arimoto, 1972, Cover and Thomas, 1991] and is guaranteed to converge to the correct solution.

As previously, $\beta \to 0$ corresponds to an agent with no computational resources which means that all posterior policies $p^*(a|w)$ are exactly equal to to the marginal $p(a)$ (which means that the posterior has effectively become independent of w, thus requiring no computation). This leads to a maximum abstraction, where all w are treated as if they were the same. On the other extreme $\beta \to \infty$, the KL-term in Eq (2.12) drops, thus essentially decoupling the optimization problems into independent problems for each w. The corresponding solution is to deterministically select the best action a^*_w for each world state. For $0 < \beta < \infty$, the agent trades off large expected utility against low mutual information. Large mutual information values occur if the agent's behavior is very world-state specific, that is if a certain action a_i is picked almost deterministically in response to a certain w_i, but the same a_i is (almost) never picked for a different w_j. In such a case the action becomes very informative about the world state, leading to a large mutual information. In contrast, a bounded rational agent must save mutual information by using the same or very similar policies for a whole range of w-values. This leads to a natural emergence of abstractions where certain subsets of w are treated as if they were the same (by responding with the same policy to all elements of a subset). Importantly, the abstractions are induced by the structure of the utility function and their granularity is governed by β which corresponds to the computational capacity of the agent. A more detailed discussion and an intuitive, illustrative example is presented in [Genewein et al., 2015b], shown in full detail in Chapter 7.

One shortcoming of the principle introduced in this section is that the optimization problem only describes a particular level of abstraction, yet it seems that humans can easily view a problem on multiple levels of abstraction, thereby flexibly choosing the most appropriate level of granularity. This shortcoming is addressed by some early work on extending the principle to decision-making hierarchies that involve multiple decision-nodes. An overview is given in the following section.

2.3.2 Decision-making hierarchies

The emergence of abstractions due to the computational limitations of a bounded rational agent is very interesting from a structure learning point-of-view since learning structure can be

thought of as the formation of abstractions (the extraction of higher-level statistical invariants). The principle introduced in the previous section adds a novel and interesting theoretical angle to viewing structure learning as the formation of abstractions. In particular, the principle suggests that structure learning might be crucially driven by limitations in computational capacity of agents. Additionally, the information-theoretic principle is expressed as a concrete optimization problem that might shed some light into the question of which computational principles can explain the (emergent) design of structured representations.

In this section, the basic rate-distortion principle for bounded rational decision-making is extended to simple two-stage hierarchies that involve an additional random variable. Depending on some conditional independence assumptions, this additional variable can be interpreted as playing either the role of a percept in a perception-action chain or as a high-level model in a model-parameter hierarchy. Importantly, the fundamental trade-off between high utility and low computational demand is applied consequently to the corresponding three-variable systems. This leads to interesting optimality principles for perception-action systems and hierarchies of abstraction.The full details are shown in Chapter 7 and briefly summarized in the following.

Perception-action systems and likelihood functions synthetization

In this section, the information-theoretic bounded rationality principle is applied to a two-stage information processing system, where the first stage can be interpreted as a perceptual channel and the second stage is a decision-making stage, referred to in the following as the action-stage. Consider the following setting: given a world state w, an agent with limited computational resources should produce an action a in order to maximize the utility $U(w,a)$. However, in contrast to the previous section, the world state is not directly accessible to the decision-maker, but can be transformed into an internal percept x by a perceptual stage. The percept can then be used to drive the subsequent action-stage. Formally, the following conditional independences hold:

$$p(w,x,a) = p(w)p(x|w)p(a|x),$$

which can also be represented as the graphical model $W \rightarrow X \rightarrow A$. The classical approach is to treat perception as an inference problem and compute a posterior belief over the world state, given the current percept, using Bayesian inference:

$$p(w|x) = \frac{p(x|w)p(x)}{p(w)}. \tag{2.15}$$

The subsequent decision-maker should then maximize the expected utility under the posterior belief over w

$$U(x,a) = \sum_{w} p(w|x)U(w,a). \tag{2.16}$$

Since the decision-maker has limited computational resources, it can be modeled with the information-theoretic framework for bounded rationality according to Eq (2.12)—from Eq (2.13) it follows that the bounded-optimal decision-making stage is given by

$$p^*(a|x) = \frac{1}{Z}p(a)e^{\beta_2 U(x,a)} \tag{2.17}$$

With this approach, the likelihood function $p(x|w)$ of Eq 2.15 must be specified in some way. Commonly, the likelihood function is chosen independently of the action-stage of the system in order to maximize predictive power. That is, w should be represented as faithfully as possible through x. Note that the assumption that the perceptual stage is also subject to limitations in computational capacity prevents a one-to-one mapping from w to x. The consequent application of the information-theoretic principle, shown in the following, will lead to a perception-action system where this likelihood function $p(x|w)$ is well defined and introduces a tight coupling between perception and action.

Within the information-theoretic framework for bounded rationality, gains in expected utility must be traded off against the effort of computation, which is measured by the mutual information. For the perception-action system this means that both the computation of the action-stage but also the computation of the perceptual stage must be taken into account, leading to the following objective:

$$p^*(x|w), p^*(a|x) = \underset{p(x|w),p(a|x)}{\arg\max} \sum_{w,x,a} p(w)p(x|w)p(a|x)U(w,a) - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(A;X), \quad (2.18)$$

where the mutual information terms measure the (average) computational demand of the perception- and the action-stage respectively and the inverse temperatures β_1, β_2 translate computational demand into cost of computation. The optimization problem given by Equation (2.18) has a closed-form solution, given by a set of self-consistent equations:

$$p^*(x|w) = \frac{1}{Z_x} p(x) e^{\beta_1 \Delta F(w,x)} \quad (2.19)$$

$$p^*(a|x) = \frac{1}{Z_a} p(a) e^{\beta_2 \sum_w p(w|x)U(w,a)} \quad (2.20)$$

where $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{1}{p(w)} p^*(x|w)p(x)$—compare the intuitive formulation of perception as Bayesian inference in Eq (2.15). $\Delta F(w,x)$ is the free energy (difference) of the action stage (compare Eq (2.8))

$$\Delta F(w,x) = \sum_a p^*(a|x)U(w,a) - \frac{1}{\beta_2} D_{\mathrm{KL}}(p^*(a|x)||p(a)) \quad (2.21)$$

and acts as a utility for the perceptual stage (see Eq (2.19) and compare it against Eq (2.9)). This means that the perceptual stage maximizes the free energy trade-off of the action-stage (in a bounded rational fashion). Additionally, $p(x) = \sum_w p(w)p^*(x|w)$ and $p(a) = \sum_{w,x} p(w)p^*(x|w)p^*(a|x)$. Importantly, the solution given by the self-consistent equations specifies the (bounded-optimal) likelihood function $p^*(x|w)$ for inference over the world-state w. The bounded-optimal decision-maker (Eq (2.20)) is identical to the bounded rational decision-maker the was intuitively specified in Eq (2.17), where inference and decision-making were decoupled and the decision maker was maximizing the posterior expected utility (Eq (2.16)) in a bounded rational fashion.

By inspecting the solution for the bounded optimal likelihood function $p^*(x|w)$ in Eq (2.19), it can be seen that the free energy trade-off of the downstream action stage (that is trading off high expected utility against low computational demand according to Eq (2.21)) plays the role of a utility for the perception stage. Crucially, this means that bounded-optimal perception has the goal of enabling the action-stage to operate as efficiently as possible. This is achieved

by a perceptual channel that extracts the most *relevant* information for acting from w, thus maximizing the downstream free energy (Eq (2.21)) and leading to a tight coupling between perception and action. This viewpoint, derived from the information-theoretic optimality principle, is in contrast to a more classical approach that assigns perception the role of representing the world-state w as faithfully as possible (given the model's limitations) leading to a decoupling between perception and action. See the publication [Genewein et al., 2015b] shown in full detail in Chapter 7 for an intuitive, illustrative example of bounded-optimal perception and action.

Distributed lossy compression - hierarchies of abstractions

In this section, the information-theoretic framework for bounded rationality is used to model distributed information processing, where the total information processing load is split up onto two processing stages. These two stages can be interpreted as two levels of an information processing hierarchy, where the high-level decision narrows down the search space for the low-level decision.

In the rate-distortion case given by Eq (2.12) which involves only the random variables W and A, a bounded-rational decision maker is presented with world-states w and computes actions a (or a posterior policy $p(a|w)$) in order to maximize the (expected) utility $U(w,a)$. The computational resources of the agent are bounded by an upper limit on the mutual information

$$I(W;A) \leq K. \tag{2.22}$$

In the serial action-perception case presented in the previous section (Eq (2.18)), the basic setup was extended by an internal variable X, playing the role of a percept. This lead to a serial chain of two processing stages: a perceptual stage that transforms w into x and an action-stage that computes a from the percept x. However, in this setup, the overall information processed is upper-bounded by the capacity of either stage, meaning that the overall information processed cannot be larger than either of the information throughputs of the two serial stages:

$$I(W;A) \leq \min\{I(W;X), I(X;A)\}. \tag{2.23}$$

This means that in order to achieve $I(W;A) = K$, both $I(W;X) \geq K$ and $I(X;A) \geq K$ or in other words: in order to achieve the same expected utility, the serial perception-action case requires at least twice the amount of information processing compared to the rate-distortion case that consists of a single processing stage only. There might be reasons why the perception-action case is still favorable, for instance by using different computational hardware for perception and action, but without additional constraints the perception-action case should in theory not be preferred over the single-processing stage case described by the rate-distortion variational problem (Eq (2.12)).

In the following, the information-theoretic framework for bounded rationality is used to construct an information processing hierarchy which allows to split a total computational load onto two stages. This architecture can in principle achieve the same expected utility as the single-stage architecture (rate-distortion) while not requiring an increased amount of information processing (which is not true for the perception-action case). Additionally, the parallel

hierarchy produces higher-level (statistical) abstractions that reduce computational effort on the lower level of the hierarchy and could be interesting for transfer learning.

To construct the parallel hierarchy of information processing, an intermediate random variable M is introduced. In contrast to the perception-action case, the joint-distribution factorizes as follows:

$$p(w,m,a) = p(w)p(m|w)p(a|w,m).$$

This opens up the possibility for a distribution of information processing load on two stages, a high-level and a low-level stage:

$$\underbrace{I(W,M;A)}_{\text{total processing}} = \underbrace{I(W;M)}_{\text{high-level}} + \underbrace{I(W;A|M)}_{\text{low-level}}. \tag{2.24}$$

The reason why the first stage is referred to as *high-level* is because it corresponds to a high-level decision that narrows down the search space over actions for the low-level decision of the second stage (more on this in the following paragraph). Crucially, the distribution of information processing also holds when considering the mutual information between W and A, which is the quantity of interest for achieving a certain expected utility

$$I(W;A) \leq \min\{I(W;M), I(M;A)\} + I(W;A|M). \tag{2.25}$$

This means that in order to achieve $I(W;A) = K$, the individual terms of the sum in the equation above can be below K and as long as they sum up exactly to K, no extra computation compared to the single-stage case occurs.

In the following, the variable M is referred to as the *model*. The model induces prior distributions over the action A according to $p(a|m)$. The (low-level) decision-maker uses these prior distributions in addition with information about w to compute a policy $p(a|w,m)$. The distribution $p(a|m)$ induced by the model can reduce the effective search-space over actions which lowers the computational demand for the low-level stage. Uncertainty over a can thus be reduced in two steps of computation: the high-level stage (the model) turns $p(a)$ into $p(a|m)$ and the low-level stage further refines $p(a|m)$ into $p(a|w,m)$. The following three components must be specified to design such a decision-making hierarchy:

$$\text{Model selector:} \quad p(m|w) \tag{2.26}$$

$$\text{Model priors:} \quad p(a|m) \tag{2.27}$$

$$\text{Low-level decision-maker:} \quad p(a|w,m) \tag{2.28}$$

The difficulty lies in choosing the correct model given a world state ($p(m|w)$), but also in designing "good" models $p(a|m)$. Consequent application of the information-theoretic framework for bounded rationality will lead to a full, formal specification of all three components of the hierarchy. In the framework, the basic principle is that gains in expected utility must be traded off against the cost of computation—in this case, computation is required to determine the correct model (or a distribution over the models) and another computational step is required to find an action given the prior induced by the model. Formally, the following optimization

problem needs to be solved:

$$p^*(m|w), p^*(a|w,m) = \underset{p(m|w),p(a|w,m)}{\arg\max} \sum_{w,m,a} p(w)p(m|w)p(a|w,m)U(w,a) - \frac{1}{\beta_1} I(W;M) - \frac{1}{\beta_2} I(W;A|M) \tag{2.29}$$

A closed-form solution exists, leading to a set of self-consistent equations:

$$p^*(m|w) = \frac{1}{Z_m} p(m) e^{\beta_1 \Delta F(w,x)} \tag{2.30}$$

$$p^*(a|m) = \sum_w p(w|m) p^*(a|m,w) \tag{2.31}$$

$$p^*(a|w,m) = \frac{1}{Z_a} p(a|m) e^{\beta_2 U(w,a)} \tag{2.32}$$

where $p(w|m)$ is given by Bayes' rule $p(w|m) = \frac{1}{p(w)} p^*(m|w)p(m)$. $\Delta F(w,x)$ is the free energy (difference) of the low-level stage (compare Eq (2.8))

$$\Delta F(w,x) = \sum_a p^*(a|w,m)U(w,a) - \frac{1}{\beta_2} D_{\mathrm{KL}}(p^*(a|w,m)||p(a|m))$$

and acts as a utility for the high-level stage, that is the high-level stage maximizes the free energy trade-off of the low-level stage in a bounded rational fashion (since the free energy appears in the exponential term of the solution for the high-level model selector in Eq (2.30), compare against Eq (2.9)). Additionally, $p(m) = \sum_w p(w)p^*(m|w)$.

The solution to the optimization problem in Eq (2.29) fully specifies all three components of the two-level hierarchy:

- a model selector $p^*(m|w)$
- model priors $p^*(a|m)$
- the low-level decision-maker $p^*(a|w,m)$

Note that all three components of the hierarchy are tightly coupled through the self-consistent equations and that the structure of the hierarchy is governed entirely by the utility-function, the distribution over world-states $p(w)$ and the computational limitations of the two stages (controlled by β_1, β_2). The hierarchy thus emerges from the utility-function and the computational limitations. Importantly, since computation is split up onto two stages, it is most economic to put the most re-usable computation into the high-level stage (stored in the priors $p^*(a|m)$. Therefore models induced by the high-level stage can be regarded as abstractions (narrowing down the search space over more specific actions) and the hierarchical architecture becomes a two-level hierarchy of abstractions.

The emergence of abstractions on the upper level of the hierarchical model, resulting from the information-theoretic optimality principle, provides a novel theoretical angle on the computational principles underlying structure learning and the emergence of structured representations. The section ends by going full-circle and discussing the introductory Bags & Marbles example from the viewpoint of the information-theoretic decision-making hierarchy. In the example there

were two boxes, each with different higher-level invariants (one box had uniformly colored bags, the other box always contained mixed-color bags with more red than blue marbles). Assume that the example is extended by showing a new bag, where it is unknown whether it belongs to the first or the second box. Again, a new marble is drawn from the new bag and turns out to be blue, and again the question is: what is the color-distribution of the remaining marbles in the bag? Answering this question can be mapped exactly onto the parallel decision-making hierarchy. For each box, the statistical invariants are captured on the level of *models*, corresponding to the upper level of the hierarchy. In particular, there are two models m_1 and m_2—each model is a Beta distribution (with parameters μ and ν modeling the distribution over the binomial parameters θ_b across bags). The marble observed from the novel bag is referred to as w and the prediction for the color-distribution over the new bag (again, a binomial parameter) is a. The first question is then, what is the probability of each model m_i given the observed marble: $p(m|w)$. Each model, induces a prior over possible binomial parameters $p(a|m)$. In conjunction with the new observation w, this prior is turned into the posterior $p(a|w,m)$. The three distributions involved correspond exactly to the components of the hierarchical model. Note that with a single observation, the predictions might be quite unspecific, since both models m_i are compatible with such an observation. But consider the case where a second marble is drawn from the novel bag and it turns out to be red (this observation is concatenated into w). This will instantly lead to a very low $p(m_1|w)$ and a very large $p(m_2|w)$, reflecting the intuition that this bag is very likely to be from the second box that contains mixed-colored bags.

3 Conclusions

3.1 Summary and Outlook

The main goal of this thesis is to contribute towards a quantitative understanding of structure learning. To this end, three concrete sub-goals were formulated

1. Investigate structure learning in a human sensorimotor integration experiment.
2. Study human structure selection behavior from a model-selection point-of-view.
3. View structure learning from an information-theoretic / cybernetic angle and shed some light on the theoretic principles underlying the emergence of structured representations for acting.

With respect to the first subgoal, we conducted an experiment to test if humans are able to tune their low-level sensorimotor integration processes to structural regularities of the environment. We found this to be the case and could quantitatively model behavior with a hierarchical Bayesian model. The experimental paradigm we designed tests for a correlation structure as the higher-level invariant, however, the paradigm is quite flexible and could easily be extended to test for more complex structures. In future experiments it could be interesting to test whether there are certain types of structure that humans can learn easier or faster (for instance correlations are very natural structures that humans are commonly faced with). Perhaps there are also some types of structure that humans fail to learn on a sensorimotor level. One interesting observation from our experiment is that the time-scale to learn the higher-level (correlation) structure was quite different from learning simple statistics, such as the mean-shift in our experiment. A natural question based on this observation is whether there are different learning-mechanisms at work and if that is the case, whether these different learning mechanisms can be attributed to different neural hardware.

To test for model selection behavior in sensorimotor tasks, we designed and conducted two experiments—one that uses a sensorimotor integration paradigm and another one that translates a textbook regression task into a sensorimotor paradigm. We found that human model or structure selection behavior is quantitatively consistent with a choice scheme based on Bayesian model selection. However, in contrast to strict Bayesian model selection that deterministically selects the model with the larger posterior probability, we found that participants stochastically selected between different models. This behavior is in quantitative agreement with a slightly modified, stochastic Bayesian model selection scheme, where the selection-probability for a model is given by the models' posterior probability. Interestingly though, the theoretical basis for this scheme is the model evidence or marginal likelihood (exactly as used in Bayesian model selection). The marginal likelihood has two interesting theoretical aspects: one, the "integrating out" of parameter-settings that are compatible with the observed, ambiguous stimulus and two, the intrinsic penalization of overly complex models (Bayesian Occam's razor). We tested for

both aspects with the two different experiments and found that they were both strongly (and quantitatively) reflected in human model selection behavior. From a purely theoretical perspective, a remaining question is whether humans really select one of the models (stochastically) and act accordingly or whether they perform a Bayesian weighting of the models according to their posterior probabilities. Distinction between these two options is complicated by the fact that an elegant scheme to practically implement Bayesian averaging could be to sample one of the two models based on their posterior probabilities and then act according to the sampled model. Such a scheme could be interpreted as a bounded rational implementation of Bayesian averaging. One prediction that requires further investigation is that if the stochastic model choice is based on a softmax-rule, then for Bayesian averaging the softmax temperature should be close to one, whereas for a selection scheme closer to Bayesian model selection, the temperature should be larger (less stochasticity). Additionally, it must be ensured that the softmax temperature differing from one is not the result of perceptual or motor uncertainty that is induced through the design of the experiment.

A very promising direction to pursue in future work is the further exploration of the theoretical groundwork laid in this thesis. Although not yet fully fleshed out and extended to the multi-variable case (that could explain deeper hierarchies), the information-theoretic principle for bounded rationality leads to interesting insights and novel angles on structure learning. While the principle was known and has been used for modeling bounded rational inference and decision-making in previous work, this thesis adds at least two interesting directions for future research. One, the thesis relates the free-energy principle (Eq (2.8)) to lossy compression and the emergence of abstractions. One prediction of this rate-distortion principle for acting is the emergence of natural levels of abstraction. Depending on the utility function of the task, these levels of abstraction are quite stable over a range of computational limitations (different β values) with sudden phase-transitions in-between. It would be interesting to design experiments with utility functions that exhibit a similar structure and then test whether humans use similar abstractions and whether the same levels of abstraction can be found in human behavior when varying the computational resources (for instance by varying the amount of noise in the stimuli or by varying reaction-time limits in non-trivial tasks). In such an experiment it would also be interesting to compare human behavior against the theoretically optimal behavior given by the rate-distortion curve. From an evolutionary point-of-view it would make sense that humans can flexibly adjust their behavior to be close to the information-theoretic optimum (given certain limitations on computational capacity). There are definitely hints for such information-optimal behavior mostly studied in multiple-alternative choice tasks (pioneered by Hick [Hick, 1952, Hyman, 1953], leading to the Hick-Heyman law) with reaction-time limits or explicit, external evidence accumulation. The other interesting direction to pursue further is to refine and study the hierarchical decision-making architectures, that is the serial perception-action systems (Eq (2.18)) and the parallel model-parameter hierarchies (Eq (2.29)). There are a few challenges and immediate open problems: solving the self-consistent equations for the multi-variable cases is tricky as there is no Blahut-Arimoto scheme with a convergence proof. Iterating the (four) self-consistent equations seems to stably converge to similar solutions in our simulations, however different initializations sometimes lead to different solutions, thus the optimization-problems are non-convex and global optima can no longer be guaranteed. Additionally, the iteration-scheme only works for discrete (tabular) distributions since the self-consistent equations do not lead to closed-form expressions in case of parametric distributions. In order to compare against human behavior, but also in case of most artificial intelligence

scenarios, continuous versions of the principles are needed. While the optimization principle and its solutions can easily be written down for the continuous, parametric case, the iteration scheme cannot. There are some recent, successful applications that use parametric distributions and perform gradient ascent with respect to the free-energy or rate-distortion objective ([Leibfried and Braun, 2015, Leibfried and Braun, 2016, Grau-Moya et al., 2016]). This approach could potentially be extended to the cases involving three variables (perception-action chain and model-parameter hierarchy). To overcome the technical problems involved when using parametric distributions but also discrete problems with large state- and action-spaces, the sampling scheme (briefly described before, but in more detail in Chapter 7) is another option that seems worth exploring. Extending the basic scheme to the hierarchical cases could provide a good starting point for analyzing and comparing against human behavior but also for constructing artificial systems that exhibit signatures of structure learning behavior.

An intuitive idea behind structure learning in motor tasks, illustrated in Figure 1.4, was that the structure of a task specifies a certain sub-manifold in a high-dimensional parameter space. The sub-manifold reduces the effective size of the search-space over parameters and thus alleviates the search-problem. In the information-theoretic principle for bounded rationality the effective size of the search-space is limited by the bound on the KL-divergence between prior and posterior (or in the hierarchical case between the prior induced by the model and the low-level posterior). From a theoretical point-of-view it is interesting to investigate whether this KL restriction can also be thought of as a restriction to some sort of sub-manifold in a different space (perhaps using information-geometric arguments) and whether the two intuitions are essentially the same or whether there are crucial differences. Intuitively, the KL-restriction could perhaps be shown to be the most effective restriction of a search-space that does not introduce any further (implicit) assumptions (actually all additional assumptions should be explicit in the prior). Pushing research towards this direction might also help in finding appropriate computational frameworks for identifying structured representations and hierarchical structures in neural activity patterns.

This thesis has looked into structure learning from a "structured probabilistic models" point-of-view. A different view on structure learning is to consider learning of functional abstractions. For instance, learning how a summation operation works can be a very useful abstraction. However, this kind of abstraction is not well captured as a statistical regularity but rather by learning a mathematical function or operand. Functional abstractions become particularly useful when combined in a compositional framework. For instance, arbitrarily complex arithmetic functions can be composed from a very small set of basic operations. One of the most promising frameworks for studying these kinds of abstractions are artificial grammars that specify rules of how to combine elements of a basis set (alphabet) into more complex, composed expressions. Recent advances try to combine such compositional systems with ideas from structured probabilistic models [Lake et al., 2015, Lake et al., 2016]. The main idea is to phrase the grammar as simple probabilistic programs that specify how to combine the base elements probabilistically (probabilistic program synthetization) and then perform inference using these probabilistic program representations. The basis-elements in the alphabet of such a system must be carefully chosen and it seems promising to study how information-theoretic constraints, similar to the ones explored in the information-theoretic principle for learning abstractions, could drive the formation of such a basis set. Additionally, the probabilistic programs involved often show a hierarchical structure, where elements of the basis set are combined into more complex elements that are again combined and so forth. The principles behind the parallel hierarchy of abstrac-

tions might provide a basis for tuning or learning good probabilistic programs similar to how the principle leads to the formation of “good and useful” models for the tasks at hand.

3.2 Statement of contributions

This cumulative thesis is composed of four peer-reviewed research papers. The author of this thesis is the main-author of all four papers. The following list provides an overview of the contributions of the author of this thesis to each of the papers and also briefly states the contributions of each of the coauthors, respectively. Authors are referred to by their initials, "TG" is the author of this thesis and "DAB" is the main supervisor of this thesis.

- **A sensorimotor paradigm for Bayesian model selection**, T. Genewein, D. A. Braun, *Frontiers in Human Neuroscience* (2012). Full text provided in Chapter 4. TG and DAB conceived the project. TG and DAB designed the experiment. TG implemented the experiment and recorded participants. TG analyzed the data and implemented and performed simulations. TG and DAB wrote the paper.
- **Occam's Razor in sensorimotor learning**, T. Genewein, D. A. Braun, *Proceedings of the Royal Society B: Biological Sciences* (2014). Full text provided in Chapter 5. TG and DAB conceived the project. TG and DAB designed the experiment. TG implemented the experiment and recorded participants. TG analyzed the data and implemented and performed simulations. TG and DAB wrote the paper.
- **Structure learning in Bayesian sensorimotor integration**, T. Genewein, E. Hez, Z. Razzaghpanah, D. A. Braun, *PLoS Computational Biology* (2015). Full text provided in Chapter 6. TG and DAB conceived the project. TG and DAB designed the experiment. TG, EH and ZR implemented the experiment and recorded participants. TG, EH and ZR analyzed the data. TG implemented and performed simulations. TG and DAB wrote the paper. EH and ZR were involved in the project as student assistants.
- **Bounded rationality, abstraction and hierarchical decision-making: an information-theoretic optimality principle**, T. Genewein, F. Leibfried, J. Grau-Moya, D. A. Braun, *Frontiers in Robotics and AI* (2015). Full text provided in Chapter 7. TG and DAB conceived the project. TG and JG implemented and performed simulations. TG and FL did analysis and derivations. TG, JG, FL and DAB wrote the paper.

4 A sensorimotor paradigm for Bayesian model selection

For a color-version of the plots in this chapter please see the digital version of this thesis or the original publication [Genewein and Braun, 2012].

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 October 2012
doi: 10.3389/fnhum.2012.00291

A sensorimotor paradigm for Bayesian model selection

***Tim Genewein*[1,2]* *and Daniel A. Braun*[1,2]**

[1] *Max Planck Institute for Biological Cybernetics, Tübingen, Germany*
[2] *Max Planck Institute for Intelligent Systems, Tübingen, Germany*

Edited by:
Sven Bestmann, University College London, UK

Reviewed by:
Amir Karniel, Ben-Gurion University, Israel
Robert Van Beers, VU University Amsterdam, Netherlands

***Correspondence:**
Tim Genewein, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany.
e-mail: tim.genewein@tuebingen.mpg.de

Sensorimotor control is thought to rely on predictive internal models in order to cope efficiently with uncertain environments. Recently, it has been shown that humans not only learn different internal models for different tasks, but that they also extract common structure between tasks. This raises the question of how the motor system selects between different structures or models, when each model can be associated with a range of different task-specific parameters. Here we design a sensorimotor task that requires subjects to compensate visuomotor shifts in a three-dimensional virtual reality setup, where one of the dimensions can be mapped to a model variable and the other dimension to the parameter variable. By introducing probe trials that are neutral in the parameter dimension, we can directly test for model selection. We found that model selection procedures based on Bayesian statistics provided a better explanation for subjects' choice behavior than simple non-probabilistic heuristics. Our experimental design lends itself to the general study of model selection in a sensorimotor context as it allows to separately query model and parameter variables from subjects.

Keywords: Bayesian model selection, sensorimotor control, structural learning, hierarchical learning, sensorimotor integration

INTRODUCTION

For biological organisms in uncertain environments, at least three important problems arise: one, the estimation of the state from noisy sensory feedback (e.g., the state of body parts). Two, the prediction of sensory consequences of actions and three, the selection of desirable actions, which builds upon the state estimate as well as the capability to predict consequences (Wolpert and Ghahramani, 2000; Todorov, 2004; Shadmehr et al., 2010; Franklin and Wolpert, 2011). Internal models are thought to play a central role in solving these problems (Shadmehr and Mussa-Ivaldi, 1994; Wolpert et al., 1995; Kawato, 1999; Tin and Poon, 2005). For estimation, internal models make use of sensory feedback to update prior beliefs about unobserved variables. Forward models predict sensory consequences of one's own actions, which allows not only to bridge delays in the sensorimotor loop, but also to distinguish between self- and externally generated motion (Poulet and Hedwig, 2006; Imamizu, 2010). To solve the problem of action selection, the theory of optimal feedback control has been used as one of a number of frameworks that study how internal models are harnessed in control (Todorov and Jordan, 2002; Diedrichsen, 2007; Chen-Harris et al., 2008; Izawa et al., 2008; Braun et al., 2009a; Diedrichsen and Dowling, 2009; Nagengast et al., 2009).

Besides the question of how biological organisms adapt internal models when the environment changes over time, another important question is how they learn new internal models and select between existing models (Shadmehr et al., 2010). There is a large body of evidence that shows that learning of predictive models happens on many different time scales and levels of abstraction (Newell et al., 2001; Smith et al., 2006; Wolpert and Flanagan, 2010). In a number of recent studies (Braun et al., 2009a,b) it was shown, for example, that the motor system can learn structural invariants when faced with randomly changing environments that share a structural similarity. In particular, in these tasks subjects had to both adapt parameters of internal models to environments with known structure and to learn new structures and their parameters from exposure to environments with different variability pattern. Here we are interested in the mechanism by which the motor system selects between different structures, that is the selection between different models that can take on different parameter settings.

In cognitive science, a number of studies has shown that human model selection in categorization or language learning tasks can be well described as Bayesian model selection (Holyoak, 2008; Kemp and Tenenbaum, 2008; Tenenbaum et al., 2011). Bayesian models have also been very successful in explaining human perceptual and sensorimotor learning of parameters in environments with known structure (van Beers et al., 1999; Ernst and Banks, 2002; Knill and Pouget, 2004; Körding and Wolpert, 2004; Körding and Wolpert, 2006; Braun et al., 2009a; Girshick and Banks, 2009). However, if there are several structures, and each structure has a range of parameter values, then the full problem of model and parameter selection arises. For perceptual learning this has been studied, for example, in case of the ventriloquist problem, where subjects have to discriminate whether a visual and an auditory signal stem from one source or from two different sources (Körding et al., 2007; Sato et al., 2007). Here we study Bayesian model selection in the context of a *sensorimotor integration* task that allows for ambiguous stimuli which are compatible with different model classes. The goal of our study is to develop an experimental paradigm for model selection and to

test whether human sensorimotor choices in such a setting are quantitatively consistent with Bayesian model selection.

RESULTS

Subjects controlled a cursor from a start position to one of two targets in a 3D virtual reality setup—see **Figure 1A**. At the start position the cursor was always displayed and represented subjects' veridical hand position. However, during the movement a random lateral shift s was applied to the cursor with respect to the hand position. Importantly, throughout most of the movement, the cursor was hidden and there was only a brief time interval of sensory feedback of the shifted cursor position. In each trial, the shift s was randomly sampled from one of two possible distributions with 50:50 probability. In the first part of the experiment (first 500 trials), the two distributions were given by a Gaussian $P(s|M_1^{\sigma_1})$ and a mixture of Gaussians $P(s|M_2)$—see **Figure 2A**. In the second part of the experiment (last 500 trials), the standard deviation of the first distribution was increased, so that the two distributions were given by $P(s|M_1^{\sigma_2})$ and $P(s|M_2)$—see **Figure 2B**.

From the point of view of model selection, the two distributions over the shift correspond to two different models M_1 and M_2. Subjects could indicate their choice of model by selecting one of the two targets. Subjects could exploit the observed shift to infer the correct target. Their belief about the shift s was reported by a compensatory horizontal movement. In all trials, the upper target represented the selection of model M_1 and the lower target represented the selection of model M_2. After completing a reaching movement, participants were informed about the correctness of their beliefs by showing the shifted cursor and hiding the incorrect target. Subjects were instructed about the relationship between shift and target selection. For example, in the first part of the study they were told that small shifts are mostly associated with the upper target and larger shifts with the lower target—see "Materials and Methods" for details. They could use the first 100 trials of each part of the experiment to acquaint themselves with the decision criterion.

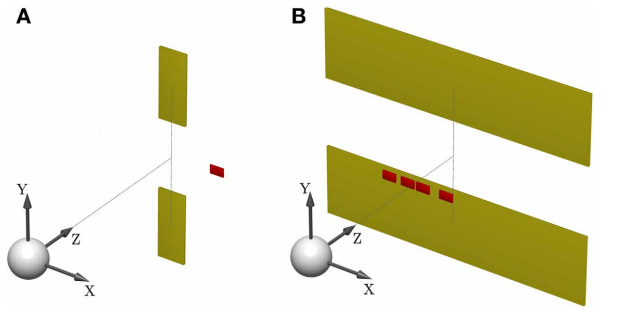

FIGURE 1 | Sensorimotor task setup. (A) Shows the schematic of a *standard trial*. One of two targets (yellow) has to be hit after observing a *shifted* cursor (red) for a short feedback duration. The white sphere depicts the start position. **(B)** Illustrates a *probe trial* where the visual feedback is ambiguous and consists of a densely sampled array of cursors, centered horizontally and vertically. The target width covers the whole lateral workspace, in order to make horizontal compensations obsolete. A photograph of the experimental apparatus is provided in **Figure 15**.

To test for model selection with different degrees of feedback uncertainty, we used *probe trials* where participants were shown ambiguous feedback. The feedback presented in these trials was an array consisting of uniformly and densely sampled rectangles that represented all the possible cursor locations—see **Figure 1B**. The array width d could take on one of three different values with equal probability: small ($d = 3$ cm), medium ($d = 5$ cm), and large ($d = 8$ cm). The larger the array the higher the uncertainty about the cursor position, and therefore the higher the uncertainty about the underlying shift s. In these probe trials subjects only reported their belief about the model without indicating the presumed shift. This was achieved by increasing the target width to the full size of the lateral workspace, which made horizontal compensations unnecessary. As in the standard trials, participants reported their belief about the correct model M by choosing either the upper or the lower target, however, in probe trials they did not receive any feedback on whether their choice was correct or not. Probe trials occurred intermixed with standard trials after the first 100 trials of each part of the experiment.

In the probe trials of the first part of the experiment we found that the probability of choosing model $M_1^{\sigma_1}$ decreases with increasing width of the ambiguous feedback array: for small array widths subjects preferred model M_1, whereas for large ambiguous feedback arrays they preferred model M_2. This can be seen in **Figure 3B**. The correlation between the array width and the choice probability was statistically significant for all subjects ($p < 0.01$, Fisher's exact test). For the second part of the experiment, subjects' model selection probabilities are depicted in **Figure 3C**. The correlation between array width and choice probability was no longer significant ($p > 0.1$ for all subjects, Fisher's exact test). This is because, for ambiguous feedback arrays with medium and large uncertainty ($d = 5, 8$ cm) subjects were now indifferent between model M_1 and M_2. Importantly, when comparing the model selection probabilities of the first part of the experiment and the second part of the experiment shown in **Figure 3A**, the model selection probabilities changed significantly ($p < 0.05$, ranksum test) for ambiguous stimuli with large uncertainty ($d = 8$ cm). For an array width of $d = 3$ or 5 cm there was no significant change in the choice probabilities for the two variance conditions of model M_1 ($p > 0.05$ ranksum test). Importantly, this implies that the choice probabilities of selecting model M_2 are very different for the same stimulus ($d = 8$ cm) depending on the complexity of model M_1.

We tested five different explanatory schemes to describe subjects' choice behaviors: model selection with Bayes factors, model selection with Bayesian policy inference of the discrimination functions learned in the standard trials, and three heuristic explanations that are non-probabilistic.

EXPLANATION 1: BAYES FACTORS

Given prior probability $P(M_i)$ over the two models M_1 and M_2, Bayes' rule describes how to assign posterior probability $P(M_i|d)$ after observing array width d in a probe trial, such that $P(M_i|d) \propto P(d|M_i)P(M_i)$—where $P(d|M_i)$ measures how well the array width d can be explained on average by the shifts that are compatible with model class M_i. This average is also called the *marginal likelihood* or *evidence* and can be computed as

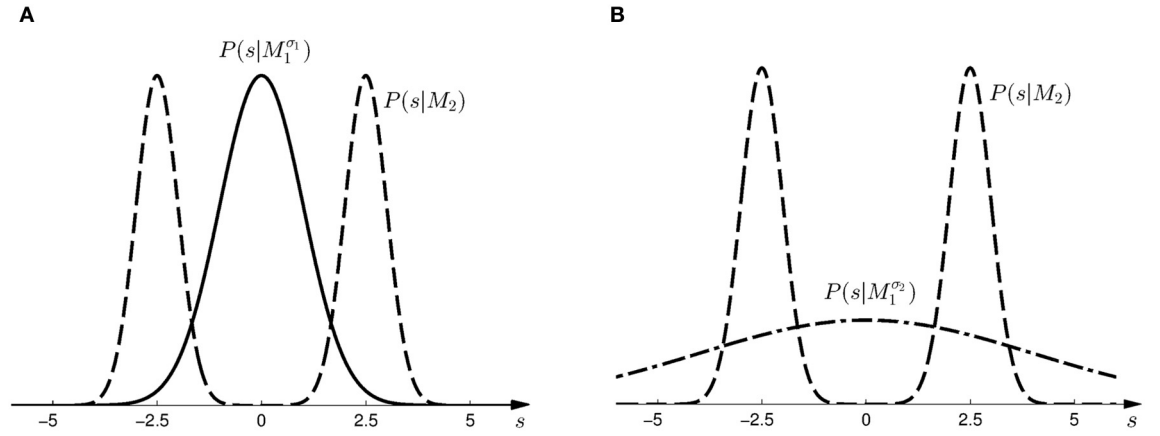

FIGURE 2 | Prior distributions for the horizontal shift *s*. (A) Shows the two prior distributions induced by model $M_1^{\sigma_1}$ (Gaussian distribution with $\sigma_1 = 1$, solid line) and M_2 (mixture of two Gaussians, dashed line) in the *first part* of the study. **(B)** Shows the two prior distributions induced by model $M_1^{\sigma_2}$ (Gaussian distribution with $\sigma_1 = 4$, dash-dot line) and M_2 (same mixture of two Gaussians, dashed line) in the *second part* of the study.

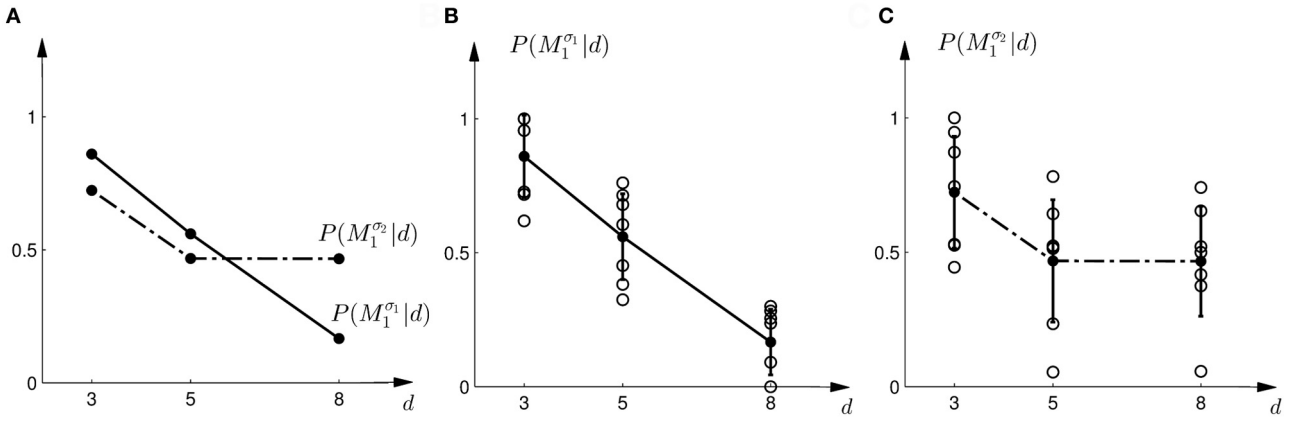

FIGURE 3 | Subjects' choice behavior in probe trials. (A) Shows the average of the experimentally observed choice probabilities for choosing M_1 for three different observed array widths d of the ambiguous feedback array. The mean is taken over all subjects. The solid line corresponds to the choice probabilities observed in the first part of the study, the dash-dot line shows the choice probabilities observed in the second part of the study, where the variance of M_1 was increased. **(B)** Shows the experimentally observed probability of choosing $M_1^{\sigma_1}$. Circles represent individual subjects' choice probabilities, the solid line shows the *average* over participants with standard deviation error bars. With increasing array size, the probability of choosing $M_1^{\sigma_1}$ decreases. **(C)** Shows the experimentally observed choice probabilities $P(M_1^{\sigma_2}|d)$ for the second part of the experiment, where $\sigma_2 > \sigma_1$. Compared to the first part, the probability of choosing $M_1^{\sigma_2}$ for the *large* array size $d = 8$ is significantly increased.

$P(d|M_i) = \int ds P(d|s, M_i)P(s|M_i)$, where each shift s contributes the likelihood $P(d|s, M_i)$ weighted by the prior $P(s|M_i)$ shown in **Figure 2**. $P(d|s, M)$ is an observation model that explains how likely it is to observe an array of width d if the true shift is s. In our experiment both models have the same observation model, that is $P(d|s, M) = P(d|s)$, which assigns equal probability to all array widths d greater or equal than a given shift s up to a maximum width $d_{\max}$. This uniform distribution over d can be seen in **Figure 4A** for different given shifts s. When $P(d|s)$ is used as a likelihood model, however, it is considered as a function of s with a particular fixed observation d. The likelihood model then indicates how likely all the different shifts s would be as an explanation of the observed array width d. The likelihood model as a function of s can be seen in **Figure 4B**.

Based on the likelihood model in **Figure 4B** and the priors shown in **Figure 2**, **Figure 5** explains how the model evidence can be computed for different observations d. **Figure 6** shows the model evidence for the different widths d of the ambiguous feedback array for all three models $M_1^{\sigma_1}$, $M_1^{\sigma_2}$, and M_2. As can be seen in the bottom row of **Figure 6**, for small width ($d = 3$ cm) of the ambiguous stimulus array, the evidence for M_1 (for both σ_1 and σ_2) is higher than for M_2. This is because model M_1 places a high probability mass on small shifts centered around zero, whereas model M_2 does not—see **Figure 2**. For ambiguous feedback arrays of medium uncertainty (array width $d = 5$ cm), the evidence of all models is very similar, that means they all can explain a medium-size range of possible shifts equally well. For ambiguous feedback arrays with large uncertainty (array width

$d = 8$ cm), the evidence of $M_1^{\sigma_1}$ is lower than the evidence for M_2, because model M_2 places more probability mass on larger feedback array widths, which ultimately results from the higher probability placed on large shifts—see **Figure 6**. However, when the standard deviation of model M_1 is increased to σ_2 in the second part of the study, both models can explain ambiguous feedback with large uncertainty equally well.

Once we have computed the model evidence for all possible observations and models, we can use it to make predictions about subjects' choice probabilities between the models. As we have equal prior probabilities $P(M_i)$ for the models in our experiment, the model class M_i that assigns a higher marginal likelihood $P(d|M_i)$ to the observation d is predicted to be preferred. Model selection is then determined by the *Bayes factor* (Kass and Raftery, 1993) between the two models, that is $P(d|M_1)/P(d|M_2)$. Based on a softmax decision rule, we can then predict subjects' choice probabilities as

$$P(a = M_1) = \frac{1}{1 + e^{-\alpha \log \frac{P(d|M_1)}{P(d|M_2)}}},$$

where $a = M_1$ implies moving up to choose model M_1 and $a = M_2$ implies moving down to choose model M_2. We assumed $\alpha = 1$ throughout. Thus, if the Bayes factor is larger than one, subjects should be more likely to choose Model 1. Conversely, if the Bayes factor is smaller than one, subjects should be more likely to choose Model 2. The choice probabilities resulting from the softmax rule are shown in **Figure 7A**. In the case of small variance σ_1, this predicts that the probability of choosing model M_1 should decrease with the increase of the uncertainty of the ambiguous feedback. In the case of large variance σ_2, this predicts that the probability of choosing model M_1 is very similar to the probability of choosing model M_2 for medium and large feedback arrays. Especially for the large ambiguous feedback array of width

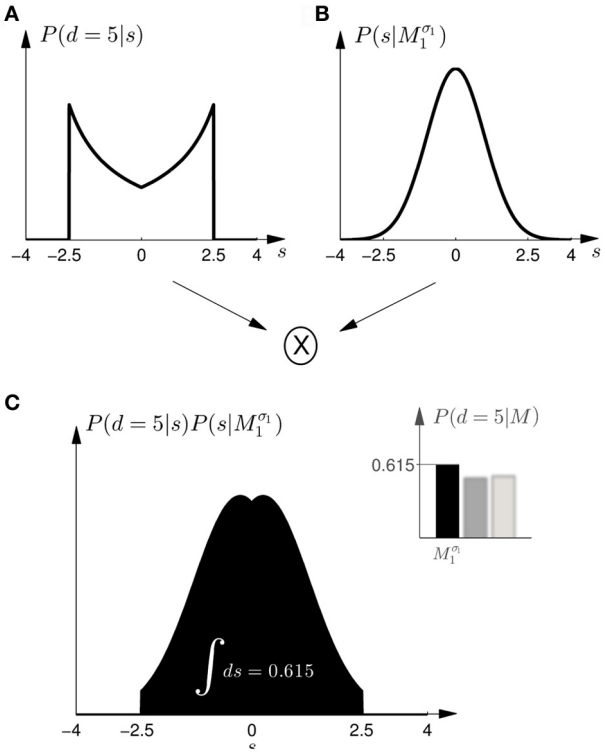

FIGURE 5 | Computation of the model *evidence* $P(d|M)$ when observing an ambiguous cursor array of width d—shown for model $M_1^{\sigma_1}$ and $d = 5$, so the possible shifts of the hand position range from −2.5 to 2.5 cm. (A) Shows the *likelihood* $P(d = 5|s)$ of observing the ambiguous cursor array of width $d = 5$ for different shifts s. **(B)** Shows the *prior* $P(s|M_1^{\sigma_1})$ over the shift s according to model $M_1^{\sigma_1}$. **(C)** Depicts the product of the distributions in panels **(A,B)** and illustrates the integration over the shifts s. This computation leads to a single value for the model evidence $P(d|M_1^{\sigma_1})$. The model evidence can be similarly computed for all models as shown in the inlet panel.

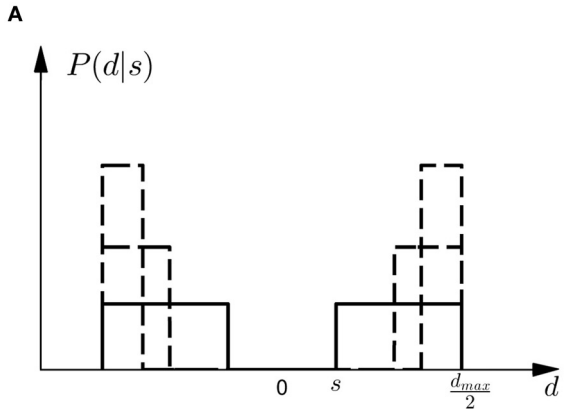

FIGURE 4 | Observation model. (A) Shows the probability $P(d|s)$ of observing an array of width d given the shift s. All array widths up to a maximum $d_{\max}$ have equal probabilities but arrays that are too small to contain the shift s have zero probability. The figure shows three different cases for three different shifts s with a solid line and two dashed lines, respectively. It is important to notice that in case of small shifts probability mass is spread quite evenly, whereas in the case of large shifts most of the probability mass is placed close to the maximum array width $d_{\max}$. **(B)** Shows the likelihood $P(d|s)$ as a function of s. For an array width d it is more likely that the shift s that caused the observation is close to half the array width $d/2$.

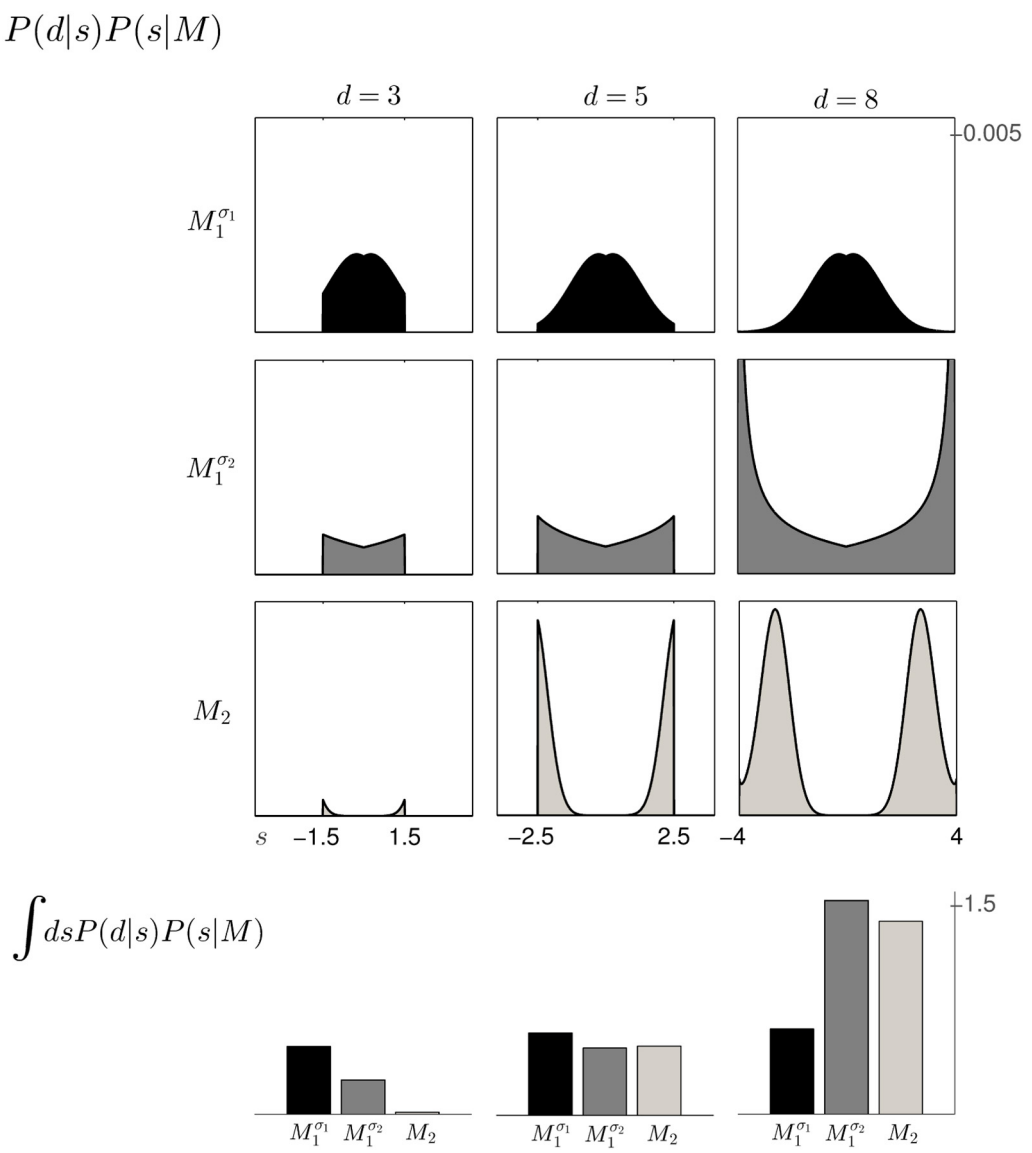

FIGURE 6 | The 3 × 3 panel shows the product of parameter prior $P(s|M)$ and observation likelihood $P(d|s)$ for the three possible array widths $d \in \{3, 5, 8\}$ cm and the three different models $M_1^{\sigma_1}$ (black), $M_1^{\sigma_2}$ (dark gray) and M_2 (light gray). The **bottom row** shows the *model evidence* $P(d|M)$ of observing a feedback array of width d for the different models, which is obtained by integrating out the shift s for each of the three models within the same column. For feedback arrays of large width ($d = 8$ cm), model M_2 is a better explanation than model $M_1^{\sigma_1}$. However, when increasing the standard deviation of M_1 to σ_2 both models have similar marginal likelihood for large width. For a detailed illustration on the computation of the model evidence $P(d|M)$ see **Figure 5**.

$d = 8$ cm—the prediction implies that for the *same* stimulus the choice probabilities for selecting model M_2 are very different depending on the complexity of model M_1.

The comparison to the actual probabilities of model selection observed in the experiment are presented in **Figures 7B,C**. In line with the predictions shown in **Figure 7A** the probability of choosing model $M_1^{\sigma_1}$ decreases with increasing width of the ambiguous feedback array for the first part of the experiment: for small array widths subjects preferred model M_1, whereas for large ambiguous feedback arrays they preferred model M_2. Similarly, the predictions explain the choice probabilities of the probe trials in the second part of the experiment, where for small array widths model M_1 is preferred, and for larger array widths subjects are indifferent between the two models. The predictions achieved a negative log-likelihood of $L = 1170$ with respect to the data.

EXPLANATION 2: BAYESIAN POLICY INFERENCE

Instead of learning different prior distributions $P(s|M_i)$ over the shifts for the two models M_1 and M_2, subjects could directly learn optimal responses $P(a = M_1|s)$ to the shifts in the standard trials and a single prior $P(s)$ over the possible shifts. They could learn, for example, that for small shifts they should mostly

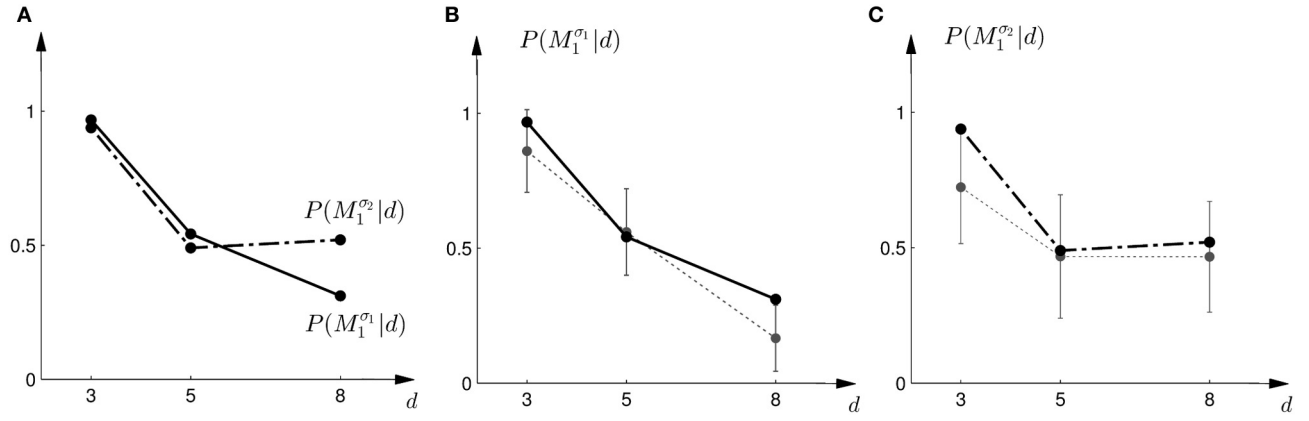

FIGURE 7 | Predictions and data: Bayes factors. (A) Shows the predicted choice probabilities for choosing M_1 as a function of the observed array width d under the assumption of a softmax choice rule. The solid line corresponds to the predicted choice probabilities for the first part of the study, the dash-dot line shows the predicted choice probabilities for the second part of the study, where the variance of M_1 was increased. **(B)** Shows the predictions for the first part of the study (solid line) and the experimental results for three different widths d of the ambiguous feedback array. The dotted line shows the mean and standard deviation across subjects of the experimentally observed probabilities of choosing $M_1^{\sigma_1}$. **(C)** Shows the predictions (dash-dot line) as well as the experimentally observed choice probabilities (dotted line represents mean with standard deviation error bars across subjects) for choosing $M_1^{\sigma_2}$ in the second part of the study.

move to the upper target, that is $a = M_1$, and for large shifts they should mostly move to the lower target, that is $a = M_2$. When faced with an ambiguous stimulus in a probe trial, they could then integrate over the responses $P(a|s)$ by considering all possible shifts weighted by the plausibility of each shift. This plausibility is given by the posterior distribution $P(s|d)$ that results from inferring the underlying shift after observing an array of width d. The choice probability in the probe trial can then be computed by the integral

$$P(a = M_1|d) = \int ds\, P(s|d) P(a = M_1|s).$$

This choice rule has been previously proposed as a stochastic Bayesian rule for control in (Ortega and Braun, 2010) to solve adaptive control problems. In this framework it is assumed that a number of primitive strategies are known that are suitable for different environments. When knowledge of the environment is not available a probabilistic superposition of the primitive strategies results in a stochastic strategy that conforms with Bayesian statistics. In our experiment the different environments correspond to trials with different shifts. The basic strategies coping with these shifts could be learned in the standard trials together with a prior $P(s)$ over all possible shifts. The unconditional prior $P(s)$ is given by the superposition of the two conditional priors shown in **Figure 2**, that is $P(s) = \frac{1}{2}P(s|M_1) + \frac{1}{2}P(s|M_2)$. In the probe trials ambiguity is induced about the underlying shift. The possible underlying shifts can be inferred through the posterior $P(s|d)$ that is given by $P(s|d) \propto P(d|s)P(s)$, with the same likelihood model $P(d|s)$ as described in the previous section and displayed in **Figure 4B**.

To test this model, we first investigated subjects' choice behavior in standard trials. In particular, we examined subjects' probability $P(a = M_1|s)$ of choosing model M_1 or M_2 for different shifts s. The response curves $P(a = M_1|s)$ for a typical subject can be seen in **Figures 8A,B**. Panel (A) shows the response curve for the first part of the experiment, and Panel (B) shows the response curve for the second part of the experiment. In both cases subjects showed a high probability for choosing model M_1 for small shifts. In the first part of the experiment this probability decreases for large shifts, implying the selection of model M_2. In the second part of the experiment the probability of selecting model M_1 decreases for larger shifts, but then increases again with very large shifts. These response curves are in agreement with the prior distributions shown in **Figure 3**, as in the first part of the experiment model M_1 was only associated with small shifts, whereas in the second part of the experiment model M_1 could also be associated with very large shifts. Learning the response functions $P(a = M_1|s)$ is therefore equivalent to learning the conditional priors $P(s|M_i)$. The fitted response curves for all subjects can be seen in **Figure 9**.

In the probe trials the underlying shift is unknown and therefore the policy $P(a = M_1|s)$ cannot be applied directly, as it requires knowledge of the shift s. Using Bayesian policy inference, the action is then determined by a probabilistic superposition that is weighted by the posterior probabilities of the shifts. This superposition allows predicting directly the choice probabilities for the probe trials. The predicted choice probabilities are shown for a typical subject in **Figures 8C,D**. Panel (C) shows the subject's predicted probability of choosing model M_1 for three different array widths d in the first part of the experiment. It can be seen that the choice probability decreases for large observed array widths. Similarly, Panel (D) shows the subject's predicted probability of choosing model M_1 in the second part of the experiment. In this case the choice probability for model M_1 is elevated for the small array width, but is close to one half for the two larger array sizes. The comparison to the actual choice probabilities of all subjects

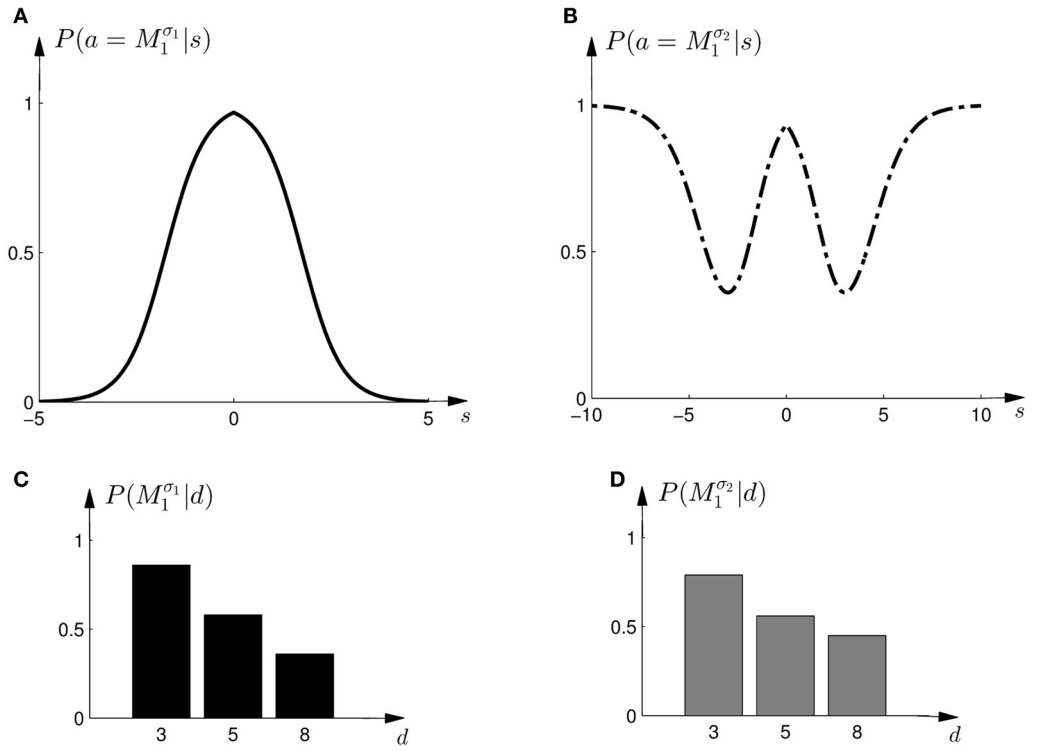

FIGURE 8 | Construction of predicted choice probabilities following Bayesian policy inference. (A) Shows the fitted choice probabilities of selecting model M_1 in standard trials when the shift s is known for a typical subject in the first part of the experiment. Model M_1 is mostly selected for small shifts. **(B)** Shows same as panel **(A)** but for the second part of the experiment. **(C,D)** Show the predicted choice probability of selecting model M_1 for this subject in a probe trial with observed array width d. The probe trial choice probabilities are obtained by a probabilistic superposition of the standard trial choice probabilities shown in panels **(A,B)**.

observed in the experiment are presented in **Figure 10**. The predictions achieved a total negative log-likelihood of $L = 1097$ with respect to the data.

EXPLANATION 3: THE "AVERAGE SHIFT"-HEURISTIC

The response curves $P(a|s)$, that describe behavior in the standard trials in dependence of the observed shift s, could also be used for non-probabilistic heuristic strategies in probe trials. One such strategy could be to simply assume the average shift when faced with an ambiguous stimulus, which corresponds to the location in the middle of the cursor array in the probe trial—see **Figures 11A,B**. In this case the choice probabilities are determined by

$$P(a = M_1|d) = P(a = M_1|s = 0).$$

However, this strategy predicts constant choice probabilities that do not vary with the observed array size. Moreover, this predicts that subjects should choose model M_1 most of the time, as it explains shifts in the middle best—see the prediction for a typical subject in **Figures 11C,D**. This prediction is in clear contradiction to the observed choice probabilities that change depending on the feedback array width. **Figure 12** shows the predictions of the "average shift"-heuristic compared to the actual choice probabilities of all subjects. The predictions of the "average shift"-heuristic achieved a total negative log-likelihood of $L = 2689$ with respect to the data.

EXPLANATION 4: THE "BIGGEST SHIFT"-HEURISTIC

Another non-probabilistic heuristic that could be employed based on the response curves of the standard trials, is to assume always the largest possible shift for any given cursor array in the probe trial. Accordingly, the location of the assumed shift would correspond to the edge of the array with total width d, such that the edge corresponds to the half-width $d/2$. The corresponding choice probabilities are determined by

$$P(a = M_1|d) = P(a = M_1|s = d/2).$$

The predictions of the "biggest shift"-heuristic can be seen in **Figures 11E,F** for a typical subject. As in the two Bayesian models, the predicted choice probability of model M_1 decreases with increasing array width for the first part of the experiment. For the second part of the experiment the "biggest shift"-heuristic predicts a slightly increased probability of choosing model M_1 for the small array width and almost indifferent choice probabilities for the two larger array widths. **Figure 13** shows the actual choice probabilities of all subjects compared to the predictions. While the "biggest shift"-heuristic predicts the right trend in the first part of the experiments, it considerably underestimates the actual

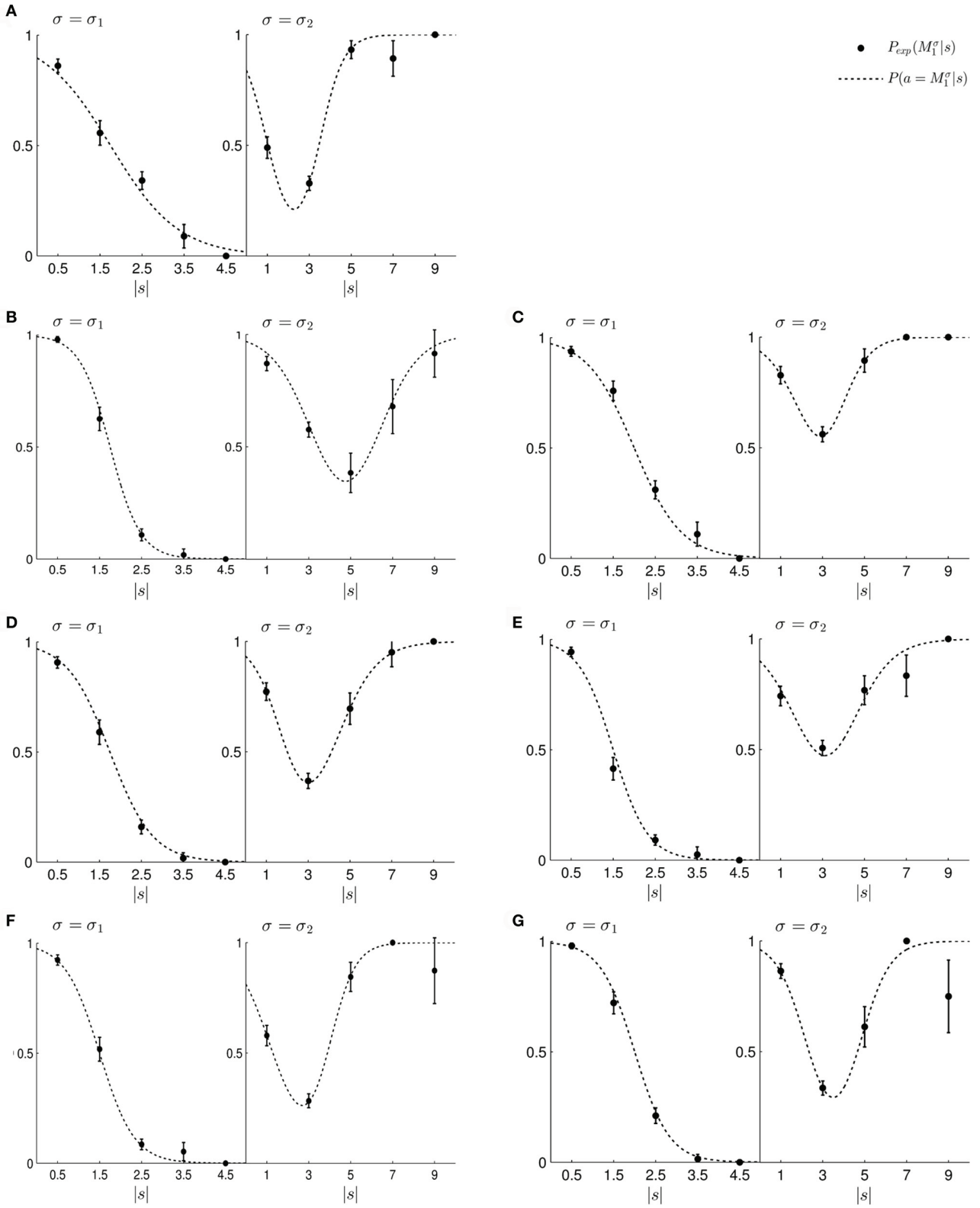

FIGURE 9 | (A–G) Show the fitted standard trial choice probabilities of selecting model M_1 for all seven subjects. The left panel shows performance during the first part of the experiment and the right panel shows performance during the second part of the experiment. In each panel, the filled circles show the subject's mean choice probability $P(M_1^\sigma)$, given a particular shift s along with standard deviation error bars. The choice probabilities for negative shifts have been mapped to the corresponding positive shift. The dotted line shows the fitted response curve that lies closest to the observed choice probabilities.

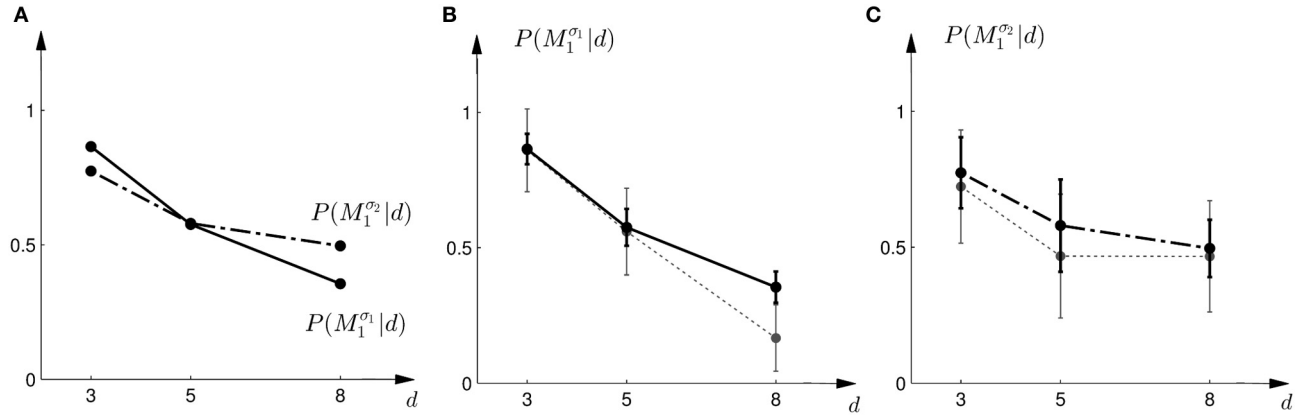

FIGURE 10 | Predictions and data: Bayesian policy inference. (A) Shows the predicted choice probabilities for choosing M_1 as a function of the observed array width d using response functions that have been fitted to the standard trial data. The solid line corresponds to the predicted choice probabilities for the first part of the study, the dash-dot line shows the predicted choice probabilities for the second part of the study, where the variance of M_1 was increased. Since the response functions have been fitted individually per subject, the predictions show the mean across all subjects. The corresponding standard deviation error bars are shown in panels **(B,C)**. **(B)** Shows the predictions for the first part of the study (solid line) and the experimental results for three different widths d of the ambiguous feedback array. The dotted line shows the mean and standard deviation across subjects of the experimentally observed probabilities of choosing $M_1^{\sigma_1}$. **(C)** Shows the predictions (dash-dot line) as well as the experimentally observed choice probabilities (dotted line) for choosing $M_1^{\sigma_2}$ in the second part of the study.

choice probabilities. The predictions for the second part of the experiment lie within the standard deviation of the experimental data. The predictions of the "biggest shift"-heuristic achieved a total negative log-likelihood of $L = 1321$ with respect to the data.

EXPLANATION 5: THE "HALFWAY SHIFT"-HEURISTIC

The probe trial shifts are grossly underestimated by the "average shift"-heuristic and overestimated by the "biggest shift"-heuristic. Accordingly the choice probabilities for model M_1 are either too high or too low, especially in the first part of the experiment. We therefore considered a "halfway shift"-heuristic that would always assume a shift halfway between the middle and the edge of the cursor array. Accordingly, the choice probabilities of the "halfway shift"-heuristic lie inbetween the extremes of the other two heuristics

$$P(a = M_1|d) = P(a = M_1|s = d/4).$$

Figure 14 shows the actual choice probabilities of all subjects compared to the predictions of the "halfway shift"-heuristic. While the error bars of the experimental data and the theoretical curves overlap, it can be seen that the heuristic generally overestimates the choice probability of choosing model M_1. The predictions of the "halfway shift"-heuristic achieved a total negative log-likelihood of $L = 1203$ with respect to the data.

DISCUSSION

We designed a three-dimensional visuomotor integration experiment where we could distinguish between parameter variables and model variables, such that the parameter variable was represented by lateral visuomotor shifts in one dimension and the model variable was represented by two targets in the other dimension that were associated with different distributions over the shifts. In particular, we designed probe trials that did not require subjects to compensate these shifts, such that the shift variable could be "integrated out" when they reported their belief about the underlying model class. This allowed us to directly compare subjects' choice probabilities to the selection probabilities predicted by five different schemes of model selection: Bayesian model selection based on Bayes factors, Bayesian policy inference over response curves that were fitted to the standard trials, and three non-probabilistic heuristics that were also based on the standard trial response curves. We found that the Bayesian model selection procedures explained our data best, whereas the three heuristics were worse in explaining choice behavior in the probe trials. By testing two sets of distributions over the shifts, for which the observed model selection probabilities agreed with the predictions of two different Bayesian model selection procedures, we achieved a proof of concept for this experimental paradigm.

The experimental paradigm differs from previous sensorimotor paradigms on Bayesian integration (Körding and Wolpert, 2004) in two important ways. First, by introducing a third dimension to the task we can simultaneously induce uncertainty over two random variables, one of which can represent a parameter variable and the other one a model variable. Previous studies (Körding and Wolpert, 2004) have shown Bayesian integration in visuomotor tasks where only uncertainty over parameters was investigated, meaning that subjects had to infer visuomotor shifts which were drawn from a particular distribution. It was shown that subjects combined information about the prior distribution of these shifts together with noisy sensory feedback in order to obtain an optimal estimate of the shift. By varying the reliability of the sensory feedback, the authors could show that subjects weighted prior and feedback in a Bayesian way—giving less weight to the feedback if reliability of the feedback was low. As there was only one distribution over

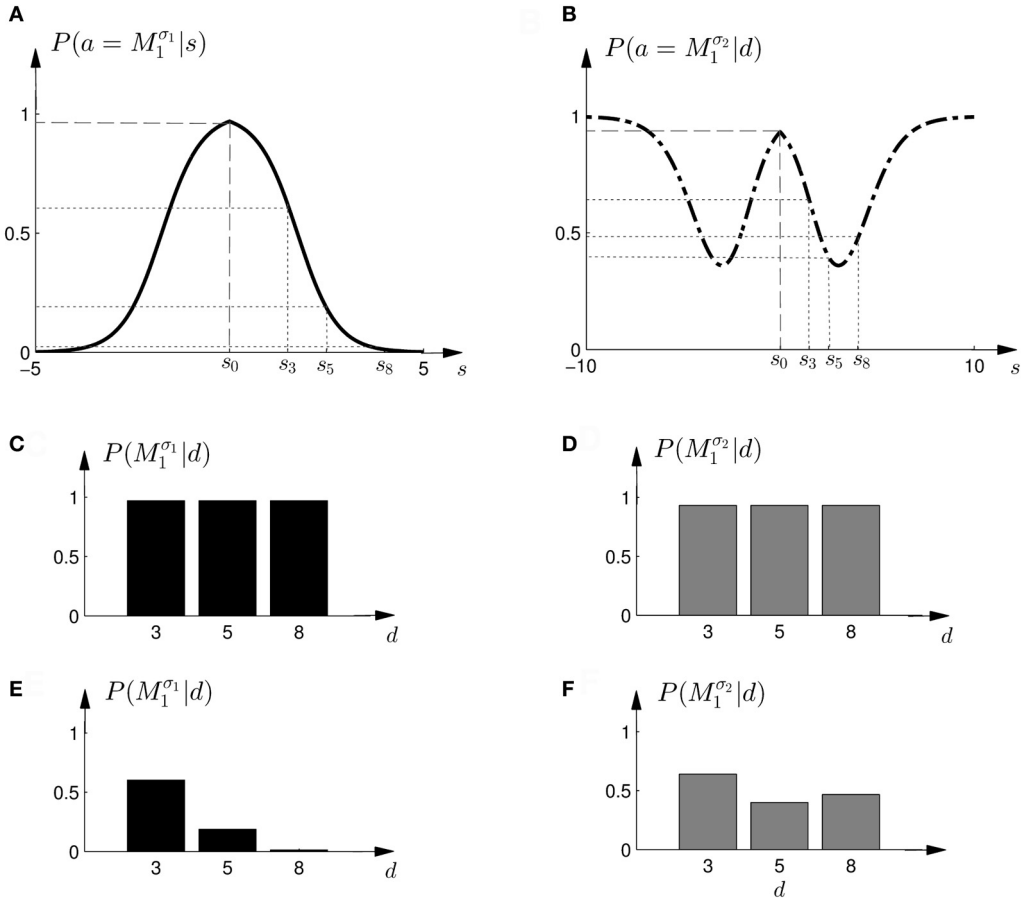

FIGURE 11 | Construction of predicted choice probabilities following heuristics. (A,B) Show the fitted choice probabilities of selecting model M_1 in standard trials of the same subject shown in **Figure 8**. For the "average shift"-heuristic the shift s_0 is always assumed in the probe trials. For the "biggest shift"-heuristic the shift is always assumed to be the largest possible shift that depends on the width of the feedback array and ranges from s_3 to s_8. **(C,D)** Show the predicted choice probability of selecting model M_1 for this subject in a probe trial with observed array width d under the "average shift"-heuristic. **(E,F)** Show the predicted choice probability of selecting model M_1 for this subject in a probe trial with observed array width d under the "biggest shift"-heuristic.

shifts, the authors could not test for Bayesian model selection in their experiment. By introducing the third dimension for the model variable, we could therefore naturally extend their paradigm.

Second, we developed a paradigm where we can separately query the model and parameter variables from subjects. We achieved this by enlarging the horizontal size of the targets in probe trials, such that lateral corrections, that are used to report the shift parameter, become obsolete. Moreover, we clamped horizontal movements in probe trials with a force channel to ensure that only the model variable is reported. These precautions are necessary, in particular if we assume that subjects report maximum a posteriori estimates, because the maximum of a joint distribution $\max_{s,M} P(s, M|d)$ is not necessarily the same as the maximum of the marginal distribution $\max_M \sum_s P(s, M|d)$. This asymmetry is one of the most important problems when designing a sensorimotor paradigm for model selection, in which the model variable has to be queried non-verbally.

As subjects were instructed verbally at the beginning of the experiment about the relationship between the cursor shifts and the two targets, the question arises in how far cognitive processes might have played a role during the experiment. We instructed subjects about the cursor-target relationship to speed up and simplify the learning process, as we were not primarily interested in standard trial performance, but in probe trial behavior. Subjects could use these instructions as a good first guess to discriminate between the two targets. Importantly, the verbal instructions relevant for standard trials did not eliminate any of the ambiguity faced in probe trials that needed to be resolved during the movement. In this sense, our task can be conceived as a generalization of previous sensorimotor tasks (Körding and Wolpert, 2004). Nevertheless we cannot rule out that cognitive processes played a role in the perception of the ambiguous stimuli during the probe trials and the subsequent discrimination between the two models, as cognitive and sensorimotor processes are often intertwined. However, even cognitive processes have been previously shown to be consistent with Bayesian inference.

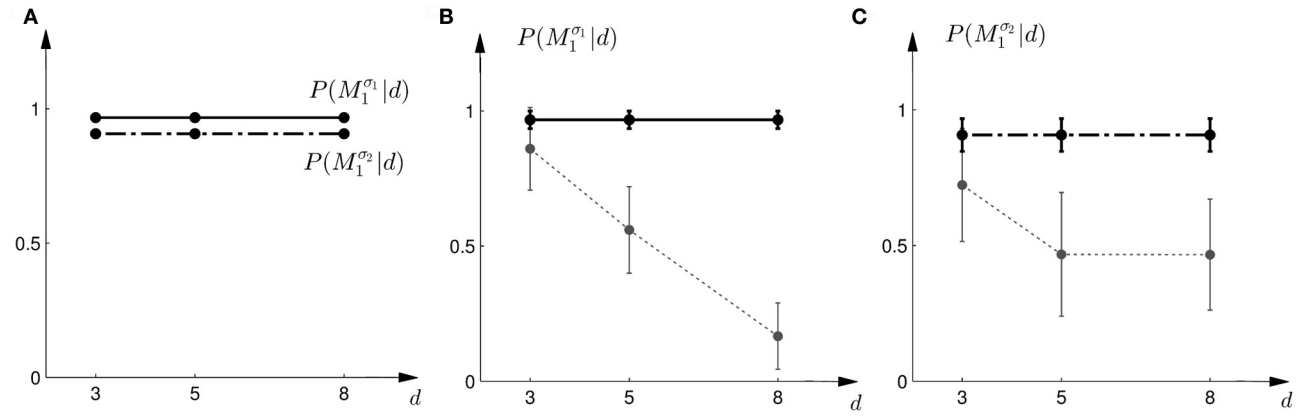

FIGURE 12 | Predictions and data: "average shift"-heuristic. (A) Shows the predicted choice probabilities for choosing M_1, assuming that subjects base their decision on the average shift when observing an array of width d. The predictions are based on response functions that have been fitted to the standard trial data. The solid line corresponds to the predicted choice probabilities for the first part of the study, the dash-dot line shows the predicted choice probabilities for the second part of the study, where the variance of M_1 was increased. Since the response functions have been fitted individually per subject, the predictions show the mean across all subjects. The corresponding standard deviation error bars are shown in panels **(B,C)**. **(B)** Shows the predictions for the first part of the study (solid line) and the experimental results for three different widths d of the ambiguous feedback array. The dotted line shows the mean and standard deviation across subjects of the experimentally observed probabilities of choosing $M_1^{\sigma_1}$. **(C)** Shows the predictions (dash-dot line) as well as the experimentally observed choice probabilities (dotted line) for choosing $M_1^{\sigma_2}$ in the second part of the study.

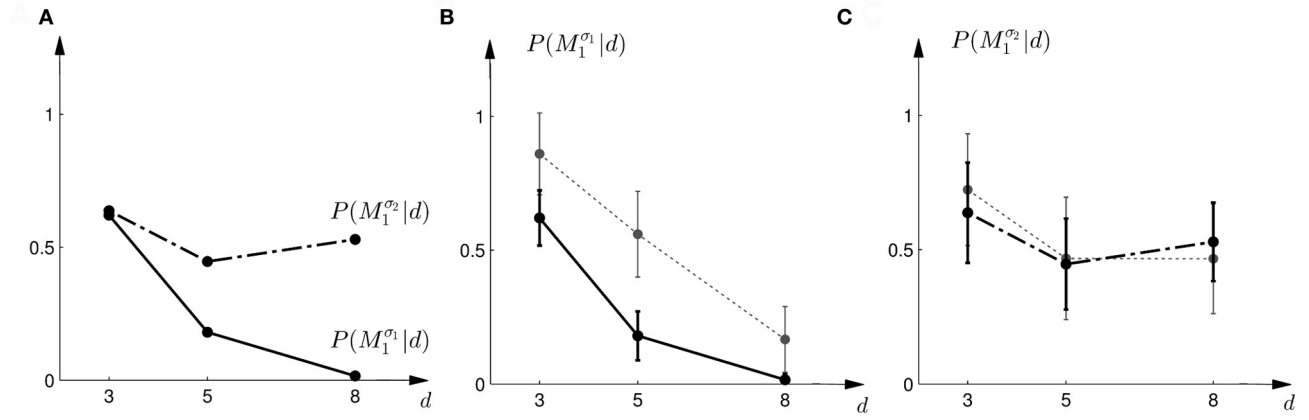

FIGURE 13 | Predictions and data: "biggest shift"-heuristic. (A) Shows the predicted choice probabilities for choosing M_1, assuming that subjects base their decision on a shift equal to the largest value that lies within the array of width d. The predictions are based on response functions that have been fitted to the standard trial data. The solid line corresponds to the predicted choice probabilities for the first part of the study, the dash-dot line shows the predicted choice probabilities for the second part of the study, where the variance of M_1 was increased. Since the response functions have been fitted individually per subject, the predictions show the mean across all subjects. The corresponding standard deviation error bars are shown in panels **(B,C)**. **(B)** Shows the predictions for the first part of the study (solid line) and the experimental results for three different widths d of the ambiguous feedback array. The dotted line shows the mean and standard deviation across subjects of the experimentally observed probabilities of choosing $M_1^{\sigma_1}$. **(C)** Shows the predictions (dash-dot line) as well as the experimentally observed choice probabilities (dotted line) for choosing $M_1^{\sigma_2}$ in the second part of the study.

In cognitive science and perceptual learning hierarchical Bayesian inference over model classes and model parameters has been previously investigated in a number of studies (Tenenbaum et al., 2006; Körding et al., 2007; Sato et al., 2007; Holyoak, 2008; Kemp and Tenenbaum, 2008, 2009; Tenenbaum et al., 2011) In particular, (Körding et al., 2007) have studied integration vs. segregation of audio-visual stimuli in human subjects—which included inference over the two models M_1 and M_2: (M_1) there is only one source for both stimuli with a location parameter s and (M_2) there are two different sources for the two stimuli with location parameters s_{visual} and s_{audio}. To specifically look into the probabilities of model selection they modeled data from a similar previous experiment (Wallace et al., 2004), where subjects were asked to report their perception of unity. In contrast, our experimental paradigm allows for reporting model selection without the need of explicitly asking subjects verbally and without them being aware that one of the task dimensions represents a parameter variable and the other a model variable.

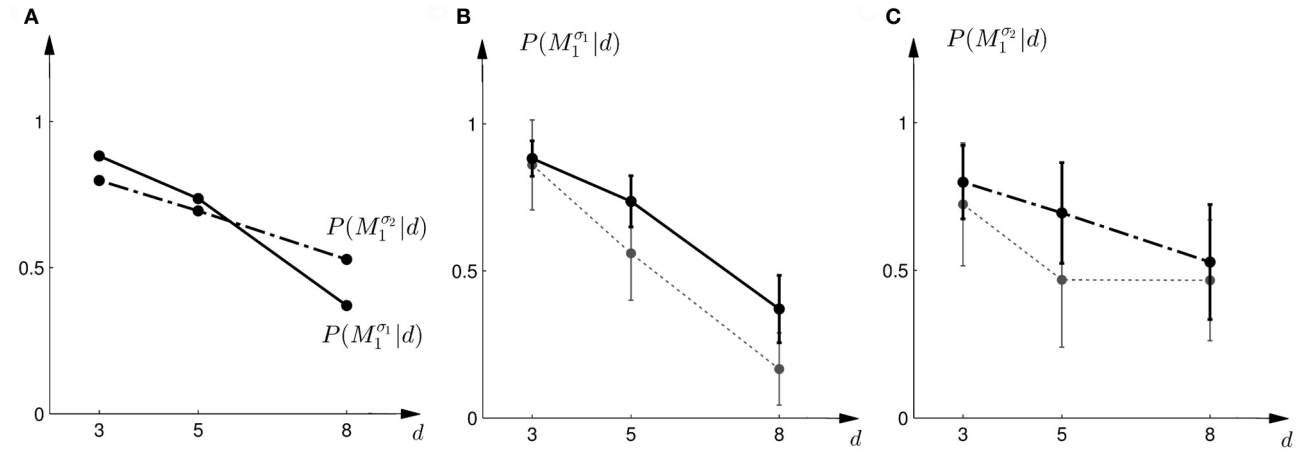

FIGURE 14 | Predictions and data: "halfway shift"-heuristic. (A) Shows the predicted choice probabilities for choosing M_1, assuming that subjects base their decision on a shift equal to half of the largest value that lies within the array of width d. The predictions are based on response functions that have been fitted to the standard trial data. The solid line corresponds to the predicted choice probabilities for the first part of the study, the dash-dot line shows the predicted choice probabilities for the second part of the study, where the variance of M_1 was increased. Since the response functions have been fitted individually per subject, the predictions show the mean across all subjects. The corresponding standard deviation error bars are shown in panels **(B,C)**. **(B)** Shows the predictions for the first part of the study (solid line) and the experimental results for three different widths d of the ambiguous feedback array. The dotted line shows the mean and standard deviation across subjects of the experimentally observed probabilities of choosing $M_1^{\sigma_1}$. **(C)** Shows the predictions (dash-dot line) as well as the experimentally observed choice probabilities (dotted line) for choosing $M_1^{\sigma_2}$ in the second part of the study.

We compared five different strategies that could explain subjects' model selection probabilities in probe trials. The two Bayesian explanations had the lowest negative likelihood and therefore explained the choice probabilities in the probe trials best. However, there are important differences between the two Bayesian explanations. The first explanation explicitly computes the marginal likelihoods and uses these likelihoods as a discriminative variable. This requires the probabilistic representation of the conditional priors $P(s|M)$, the prior over the models $P(M)$, and the likelihood model $P(d|s)$. The optimal strategy is then determined over the marginal likelihood that results from an integration of these distributions. In contrast, Bayesian policy inference results from a stochastic superposition of given policies, which in our case correspond to the model selection probabilities in standard trials when the visuomotor shift is known. In probe trials, the model selection probabilities can then be determined by an integral over these standard trial policies. If subjects' choice behavior was non-stochastic we could easily distinguish between these two possibilities, as decisions based on the marginal likelihood bear no intrinsic stochasticity—we imposed it here through the softmax-function—and decisions resulting from the probabilistic superposition of standard trial policies would always be stochastic. Given the error bars on our data and the fact that real decision-making processes are always somewhat noisy, it is hard to distinguish between the two processes, even though the Bayesian policy inference achieved the lowest negative likelihood—compare **Figures 7**, **10**.

We also examined three simple heuristics and tested in how far they might be able to explain the observed choice probabilities in probe trials. We investigated heuristics that did not consider any probabilistic representation of the task. In particular, we were investigating in how far standard trial policies could be harnessed to construct heuristics for the probe trials. A first heuristic assumed that subjects would always use the standard trial policy associated with a zero shift right in the middle of the ambiguous cursor array in the probe trial (the "average shift"-heuristic). A second heuristic assumed that subjects would always use the standard trial policy associated with the largest possible shift at the edge of the ambiguous cursor array in the probe trial (the "biggest shift"-heuristic). A third heuristic was a mixture between the two, always using the standard trial policy associated with a shift halfway between the middle and the edge of the cursor array. Especially, the first two heuristics provided very poor explanations of the choice behavior, because they either systematically under- or overestimated the probability of choosing one of the models. The "halfway shift"-heuristic achieved a negative log-likelihood value that was only slightly higher than model selection with Bayes factors, but the mismatch in the fits still seemed to be systematically biased—see **Figure 14**. More importantly, the question remains why any of these heuristics would be formed and applied. As subjects did not receive any performance feedback in the probe trials, they could not have learned the heuristics from trial and error.

Bayesian methods are typically used in two different ways in psychophysical studies. They can simply be used as techniques to analyze the data or they can be interpreted as processes that might take place in a "Bayesian brain" that tries to make sense of the world around it. Here we used different Bayesian and non-Bayesian explanations to describe subjects' choice behavior in a model selection task. This does not necessarily have any implications as to which precise algorithm the brain might use to

achieve this behavior. In fact, there are a number of methods that have been suggested for the problem of model selection: the Akaike Information Criterion (AIC) (Akaike, 1974), the Schwarz or Bayesian Information Criterion (BIC) (Schwarz, 1978), minimum description length (MDL) (Rissanen, 1978), Bayes factors (Kass and Raftery, 1993; MacKay, 2003), structural risk minimization (Vapnik, 1995), and regularization methods (Bishop, 2006). Model selection criteria like AIC and BIC can be considered as approximations to Bayesian model selection, but also MDL, regularization and complexity measures in statistical learning theory can be related to the consideration of prior probabilities in Baysian model selection (MacKay, 2003). Finally, how model selection is achieved by neurons in the brain is subject to neurophysiological investigation.

A key capability of biological organisms is to cope with an uncertain environment. Uncertainty has many sources. It can originate from noise in the nervous system (Faisal et al., 2008), but also from uncertainty that arises in the face of ambiguous stimuli. In dealing with uncertainty, Bayesian statistics have proven to be a powerful and unifying framework not only in cognitive sciences, but also in sensorimotor tasks (Körding and Wolpert, 2006) and neural computation (Knill and Pouget, 2004; Doya, 2007; Orban and Wolpert, 2011). In particular, hierarchical Bayesian models for inference and control might allow modeling a variety of learning processes on multiple levels of abstraction (Haruno et al., 2003). Our task design provides a means to study such hierarchical integration in the context of sensorimotor control.

MATERIALS AND METHODS

PARTICIPANTS

Three female and four male participants were recruited from the student population of the University of Tübingen. The study was approved by the local ethics committee and all participants were naive and gave informed consent. The local standard rate of eight Euros per hour was paid for participation in the study.

MATERIALS

We used a virtual reality setup consisting of a Sensable® Phantom® Premium 1.5 High Force manipulandum for tracking participants' hand movements in three dimensions and an NVIS® nVisor ST50 head-mounted display (HMD) for creating stereoscopic 3D virtual reality—see **Figure 15**. Movement position and velocity were recorded with a rate of 1 kHz. To prevent very fast movements, the manipulandum was operated with a weak isotropic viscous force field of $\vec{f} = \alpha \mathbb{I}_{3\times 3}\dot{\vec{x}}$, where $\alpha = 0.04\frac{Ns}{\text{cm}}$, $\mathbb{I}_{3\times 3}$ is the identity and $\dot{\vec{x}}$ is the three-dimensional velocity vector.

EXPERIMENTAL DESIGN

A model selection problem can be characterized by a bivariate distribution $P(s, M)$ over a continuous random variable s and a binary random variable M, where s plays the role of the model parameter and M plays the role of the model. To study model selection in a sensorimotor context, we designed a 3D visuomotor task where participants had to move a cursor from a start position to one of two targets, referred to as upper and lower target in the following—see **Figure 1**. During the movement, the horizontal position of the cursor was shifted, and the shift was generated by one of two possible statistical models $M \in \{M_1, M_2\}$ with 50:50 probability. Importantly, subjects were not informed about the 50:50 probability. Each target corresponded to a model M—the upper target corresponded to M_1 and the lower target corresponded to M_2. The correct target was the one whose corresponding model M actually generated the observed shift in any particular trial. The shifts were generated by first sampling a model M and then sampling a shift from the shift-distribution $P(s|M)$. Since shifts were generated probabilistically, any shift could in principle be generated by either model, however, with different probabilities. There was only brief sensory feedback of the shifted cursor during the movement. Participants had to use this feedback together with their knowledge of the previously learned statistical models $P(s|M)$ to not only infer the shift s, but also the model M. In these *standard trials*, participants reported their belief $P(s, M)$, where the shift s was indicated by a compensatory horizontal movement when hitting a target, and the belief about the model M was reported by choosing one of the two targets.

In order to test for model selection in case of feedback uncertainty, participants also experienced *probe trials*, in which they only reported their belief $P(M) = \int dsP(s, M)$ about the model M. This was achieved by increasing the width of the targets to cover the whole horizontal workspace, such that no horizontal compensatory movements were necessary in these trials. During the movement in probe trials, sensory feedback was briefly shown in shape of arrays consisting of uniformly and densely sampled rectangles that represented all the possible cursor locations—see **Figure 1B**. Each probe trial had one of three possible feedback array sizes (small, medium and large) that occurred equi-probably. Since the size of the feedback array constrained the uncertainty about the possible shifts s, Bayesian model selection required participants to "integrate out" different intervals of the

FIGURE 15 | Photograph of the experimental apparatus. The subject operates a Sensable® Phantom® Premium 1.5 High Force manipulandum and receives stereoscopic 3D visual feedback through an NVIS® nVisor ST50 head-mounted display.

parameter s when deciding on the model M by choosing either the upper or lower target.

TRIAL SETUP

Each participant performed two parts of the experiment consisting of 500 trials each. Before starting the experiment, participants were informed about the relation between observing the horizontal cursor shift and selecting one of the two targets. To start them off in the first part of the experiment, they were told that small shifts would be often associated with the upper target and larger shifts mostly with the lower target—but they were also instructed that they should use the first 100 training trials for learning this relationship precisely. Similarly, they were told for the second part of the experiment that small and very large shifts would be often associated with the upper target and medium large shifts mostly with the lower target. The initial 100 trials of each of the two sessions were standard trials only. To keep participants motivated, the hit ratio (in percent of the standard trials presented so far) was displayed. In the following 400 trials standard and probe trials were intermixed. For the probe trials, subjects were instructed that there would be a whole array of little cursors any of which could be the true cursor and that again they would have to decide which one was the correct target just like in the standard trials, but this time without knowing for sure which cursor was the correct one. The probability of presenting a probe trial was 0.45 if the previous trial was a standard trial and 0 if the previous trial was a probe trial. The second block of 500 trials was identical, only the probability distribution over shifts of model M_1 was broadened to investigate the effect on the model selection process.

EXPERIMENTAL PRIOR DISTRIBUTIONS

Each model induced a different prior probability density $P(s|M)$ over the horizontal shifts s. In part one of the study, the two models were a Gaussian and a bimodal mixture of Gaussians:

$$\begin{aligned} M_1: \quad & P(s|M_1^{\sigma_1}) = \mathcal{N}(0, 1\ \text{cm}^2) \\ M_2: \quad & P(s|M_2) = \frac{1}{2}\mathcal{N}(-2.5\ \text{cm}, 0.25\ \text{cm}^2) \\ & \quad + \frac{1}{2}\mathcal{N}(2.5\ \text{cm}, 0.25\ \text{cm}^2). \end{aligned}$$

In part two of the study, the same distributions were used, only the standard deviation of $P(s|M_1)$ was increased such that

$$\begin{aligned} M_1: \quad & P(s|M_1^{\sigma_2}) = \mathcal{N}(0, 16\ \text{cm}^2) \\ M_2: \quad & P(s|M_2) = \frac{1}{2}\mathcal{N}(-2.5\ \text{cm}, 0.25\ \text{cm}^2) \\ & \quad + \frac{1}{2}\mathcal{N}(2.5\ \text{cm}, 0.25\ \text{cm}^2). \end{aligned}$$

The prior probability of both models was always $P(M_1) = P(M_2) = \frac{1}{2}$. A plot of the prior distributions is shown in **Figure 3**.

EXPERIMENTAL PROCEDURE: STANDARD TRIALS

After hearing a beep, participants initiated a reaching movement by controlling a *cursor* (red sphere, radius 0.4 cm) from a start position (gray sphere, semi-transparent, radius 0.9 cm) to one of two target blocks (yellow cuboids, height 5 cm, width 2 cm)—see **Figure 1A**. One of the target blocks was in the upper half of the workspace, the other target block was in the lower half—with a distance of 2 cm in-between. Both target blocks were presented at a depth of 18.5 cm with respect to the start position. Once the cursor had left the start position, it was invisible and an additive random shift was applied to the cursor position. The shift was drawn from a distribution $P(s|M)$ once M had been sampled from $P(M)$. The correct target was determined by the sampled M, that is the upper target was correct if $M = M_1$ and the lower target was correct if $M = M_2$. While the cursor was invisible during the movement, after a movement depth of 5.5 cm visual feedback (red rectangle, width 0.8 cm, height 0.3 cm) of the shifted cursor position was displayed for 100 ms. When the movement exceeded a depth of 18.5 cm the trial ended. If the cursor was in-between the two targets without touching either of them, the trial continued until one of the targets was chosen. For hitting a target, the cursor had to at least touch the target block. When participants hit the correct target, a high-pitch beep was played. When participants hit the wrong target or missed the correct target, a low-pitch beep was played. In either case the incorrect target disappeared. At the end of the reach, the shifted cursor position was shown. If movement was still in progress after 2 s, the trial was aborted and had to be repeated.

EXPERIMENTAL PROCEDURE: PROBE TRIALS

In contrast to standard trials, the target width of the two targets was increased to 20 cm in probe trials, thereby covering the entire horizontal workspace. Crucially, this made any compensatory movements in the horizontal direction obsolete and reduced the task to a binary model selection problem that only required choosing either the upper or lower target. To further discourage horizontal compensatory movements in probe trials, we generated a "force tunnel" that did not allow left/right deviations from the middle of the workspace, but only up/down and forward/backward movements. Since sideward movements were not necessary in probe trials, the impact of the tunnel force was barely noticeable and most participants reported that they did not notice it at all when interviewed after the experiment.

Probe trials also started with a beep, after which participants initiated a reaching movement to one of the two targets (yellow cuboids, height 5 cm, width 20 cm)—see **Figure 1B**. At a movement depth of 5.5 cm visual feedback was displayed for 100 ms. However, in contrast to the standard trials, feedback was not shown as a little rectangle representing the shifted hand position, but as an array of multiple same-sized rectangles that were sampled simultaneously and uniformly from one of three possible horizontal intervals: $[-1.5, 1.5]$, $[-2.5, 2.5]$, and $[-4.0, 4.0]$ cm. The little rectangles had 0.8 cm width and there were 4, 7, and 10 little rectangles, respectively shown for small, medium and large bar size at any one time frame. The probability of showing one of the array sizes was one third. Participants were informed that this array indicated all possible cursor positions and that the true cursor was at one of the many possible positions seen in the array. Since sideward deviations were impossible due to the tunnel force and vertical deviations carried no information with

respect to possible shifts, the arrays were always centered in the workspace both horizontally and vertically. In order to make participants understand that the shift was sampled uniformly from the cursor array, at the end of the movement, after participants had chosen one of the two targets, a cursor position was drawn from the uniform distribution over the shown interval and displayed. However, no visual or auditive feedback was given to indicate whether the correct target was hit or not. The three possible widths for the interval (small: 3 cm, medium: 5 cm, and large: 8 cm) induced an increasing amount of uncertainty about possible shifts. In the probe trials we could therefore investigate how these different amounts of uncertainty with respect to the parameter s affected the selection of the model M, when the parameter s was not reported and therefore could be "integrated out."

MODELING

For the two Bayesian explanations of choice behavior in probe trials, we used the following observation model. Crucially, in our experiment observing an array of width $d \geq 0$ is the same for both models M_1 and M_2 and therefore the likelihood model only depends on the shift variable s, such that

$$P(d|s) = \begin{cases} \frac{1}{\frac{d_{\max}}{2} - |s|} & \text{if } d \geq 2|s| \text{ and } d \leq d_{\max} \\ 0 & \text{otherwise,} \end{cases}$$

where $d_{\max}$ represents the maximum possible array width. In our experiment $d_{\max} = 8$ cm. This observation model implies that for any given shift s, the array size cannot be smaller than s, since it must contain s, and the array size cannot exceed the maximum size $d_{\max}$. All array sizes in-between have equal probability. After observing array size d, Bayes' rule allows us to infer the posterior over both the shift and the model

$$P(s, M|d) = \frac{P(d|s)P(s|M)P(M)}{\sum_M \int ds\ P(d|s)P(s|M)P(M)}.$$

The Bayes factor can be derived from this posterior by realizing that $P(d|M) = P(M|d)$ and $P(M|d) = \int dsP(s, M|d)$. In case of the Bayesian policy inference the posterior $P(s|d)$ can be derived from the joint posterior by realizing that $P(s|d) = \sum_M P(s, M|d)$.

In line with Bayesian policy inference the posterior $P(s|d)$ is used for the probabilistic superposition of the standard trial policies $P(a = M_1|s)$ such that the choice probability in probe trials is given by $P(a = M_1|d) = \int ds\, P(s|d)P(a = M_1|s)$. The standard trial policies $P(a|s)$ were fitted to the data as follows. First, all standard trials were sorted into five equidistant bins depending on the magnitude of the shift in each trial. For the first part of the experiment the five bins were [0, 1], [1, 2], [2, 3], [3, 4], and [4, 5] cm. For the second part of the experiment the five bins were [0, 2], [2, 4], [4, 6], [6, 8], and [8, 10] cm. The relative frequencies of choosing model M_1 in these bins was fitted by a sigmoid psychometric function

$$P(a = M_1|s) = 1 - \frac{1}{1 + e^{\frac{-s+\zeta}{\theta}}}$$

for the first part of the experiment and a two-partite sigmoid function

$$P(a = M_1|s) = 1 - \frac{1}{1 + e^{\frac{-s+\gamma}{\delta}}} + \frac{1}{1 + e^{\frac{-s+\kappa}{\tau}}}$$

for the second part of the experiment. The free parameters $\zeta, \theta, \gamma, \delta, \kappa, \tau$ were fitted by minimizing square error. The fits can be seen in **Figure 9**.

ACKNOWLEDGMENTS

This study was supported by the DFG, Emmy Noether grant BR4164/1-1.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE T. Autom. Control* AC-19, 716–723.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning, Vol. 4*. New York, NY: Springer.

Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009a). Learning optimal adaptation strategies in unpredictable motor tasks. *J. Neurosci.* 29, 6472–6478.

Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009b). Motor task variation induces structural learning. *Curr. Biol.* 19, 352–357.

Chen-Harris, H., Joiner, W. M., Ethier, V., Zee, D. S., and Shadmehr, R. (2008). Adaptive control of saccades via internal feedback. *J. Neurosci.* 28, 2804–2813.

Diedrichsen, J. (2007). Optimal task-dependent changes of bimanual feedback control and adaptation. *Curr. Biol.* 17, 1675–1679.

Diedrichsen, J., and Dowling, N. (2009). Bimanual coordination as task-dependent linear control policies. *Hum. Mov. Sci.* 28, 334–347.

Doya, K. (ed.). (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Cambridge, MA, USA: MIT Press.

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433.

Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303.

Franklin, D. W., and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron* 72, 425–442.

Girshick, A. R., and Banks, M. S. (2009). Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *J. Vis.* 9, 8.1–820.

Haruno, M., Wolpert, D. M., and Kawato, M. (2003). Hierarchical mosaic for movement generation. *Int. Cong. Ser.* 1250, 575–590.

Holyoak, K. J. (2008). Induction as model selection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10637–10638.

Imamizu, H. (2010). Prediction of sensorimotor feedback from the efference copy of motor commands: a review of behavioral and functional neuroimaging studies. *Jpn. Psych. Res.* 52, 107–120.

Izawa, J., Rane, T., Donchin, O., and Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *J. Neurosci.* 28, 2883–2891.

Kass, R. E., and Raftery, A. E. (1993). Bayes factors and model uncertainty. *J. Am. Stat. Assoc.* 90, 466.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727.

Kemp, C., and Tenenbaum, J. B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10687–10692.

Kemp, C., and Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychol. Rev.* 116, 20–58.

Knill, D. C., and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE* 2:e943. doi:10.1371/journal.pone.0000943

Körding, K. P., and Wolpert, D. M. (2004). Bayesian integration in

sensorimotor learning. *Nature* 427, 244–247.

Körding, K. P., and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* 10, 319–326.

MacKay, D. J. C. (2003). *Information heory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

Nagengast, A. J., Braun, D. A., and Wolpert, D. M. (2009). Optimal control predicts human performance on objects with internal degrees of freedom. *PLoS Comput. Biol.* 5:e1000419. doi:10.1371/journal.pcbi.1000419

Newell, K. M., Liu, Y. T., and Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychol. Rev.* 108, 57–82.

Orban, G., and Wolpert, D. M. (2011). Representations of uncertainty in sensorimotor control. *Curr. Opin. Neurobiol.* 21, 1–7.

Ortega, P. A., and Braun, D. A. (2010). A minimum relative entropy principle for learning and acting. *J. Artif. Intell. Res.* 38, 475–511.

Poulet, J. F. A., and Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science* 311, 518–522.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica* 14, 465–471.

Sato, Y., Toyoizumi, T., and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* 19, 3335–3355.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.

Shadmehr, R., and Mussa-Ivaldi, F. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* 14, 3208–3224.

Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* 33, 89–108.

Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* 4:e179. doi:10.1371/journal.pbio.0040179

Tenenbaum, J. B., Griffiths, T. L., and Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* 10, 309–318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285.

Tin, C., and Poon, C.-S. (2005). Internal models in sensorimotor integration: perspectives from adaptive control theory. *J. Neural Eng.* 2, S147.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci.* 7, 907–915.

Todorov, E., and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235.

van Beers, R. J., Sittig, A. C., and Gon, J. J. (1999). Integration of proprioceptive and visual position-information: an experimentally supported model. *J. Neurophysiol.* 81, 1355–1364.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory, Vol. 8.* New York, NY: Springer.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., and Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Exp. Brain Res.* 158, 252–258.

Wolpert, D. M., and Flanagan, J. R. (2010). Motor learning. *Curr. Biol.* 20, R467–R472.

Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3(Suppl.), 1212–1217.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 August 2012; accepted: 03 October 2012; published online: 19 October 2012.

Citation: Genewein T and Braun DA (2012) A sensorimotor paradigm for Bayesian model selection. Front. Hum. Neurosci. ***6****:291. doi:* ***10.3389/fnhum.2012.00291***

5 Occam's Razor in sensorimotor learning

For a color-version of the plots in this chapter please see the digital version of this thesis or the original publication [Genewein and Braun, 2014].

PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Genewein T, Braun DA. 2014 Occam's Razor in sensorimotor learning. *Proc. R. Soc. B* **281**: 20132952. http://dx.doi.org/10.1098/rspb.2013.2952

Received: 11 November 2013
Accepted: 27 February 2014

Subject Areas:
neuroscience

Keywords:
Occam's Razor, sensorimotor control, structural learning, Bayesian model comparison, Gaussian processes

Author for correspondence:
Tim Genewein
e-mail: tim.genewein@tuebingen.mpg.de

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2013.2952 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Occam's Razor in sensorimotor learning

Tim Genewein[1,2,3] and Daniel A. Braun[1,2]

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[3]Graduate Training Centre of Neuroscience, Tübingen, Germany

A large number of recent studies suggest that the sensorimotor system uses probabilistic models to predict its environment and makes inferences about unobserved variables in line with Bayesian statistics. One of the important features of Bayesian statistics is Occam's Razor—an inbuilt preference for simpler models when comparing competing models that explain some observed data equally well. Here, we test directly for Occam's Razor in sensorimotor control. We designed a sensorimotor task in which participants had to draw lines through clouds of noisy samples of an unobserved curve generated by one of two possible probabilistic models—a simple model with a large length scale, leading to smooth curves, and a complex model with a short length scale, leading to more wiggly curves. In training trials, participants were informed about the model that generated the stimulus so that they could learn the statistics of each model. In probe trials, participants were then exposed to ambiguous stimuli. In probe trials where the ambiguous stimulus could be fitted equally well by both models, we found that participants showed a clear preference for the simpler model. Moreover, we found that participants' choice behaviour was quantitatively consistent with Bayesian Occam's Razor. We also show that participants' drawn trajectories were similar to samples from the Bayesian predictive distribution over trajectories and significantly different from two non-probabilistic heuristics. In two control experiments, we show that the preference of the simpler model cannot be simply explained by a difference in physical effort or by a preference for curve smoothness. Our results suggest that Occam's Razor is a general behavioural principle already present during sensorimotor processing.

1. Introduction

Prediction is a ubiquitous phenomenon in biological systems ranging from basic motor control in animals to scientific hypothesis formation in humans [1–6]. A fundamental problem of such predictive systems is how to choose between multiple competing hypotheses that explain observed data equally well, but make different predictions. A principled way to address this problem is *Occam's Razor*, suggesting that one should accept the simplest explanation requiring the fewest assumptions. Mathematically, Occam's Razor can be formalized within the framework of Bayesian inference—known as *Bayesian Occam's Razor* [7,8]. In Bayesian inference, the simplicity or complexity of a model can be illustrated by the distribution over different datasets the model can explain (figure 1*a*). A simple model predicts only a small number of specific datasets, whereas a more flexible complex model can explain a wider range of data. If a particular dataset can be explained by both models, the more complex model has to assign lower probability to this dataset than the simpler model, because it has to spread its probability mass over many different datasets. This naturally embodies Occam's Razor.

A generic example of Occam's Razor is depicted in figure 1*b*. The black dots represent a number of observed data points and the coloured curves represent potential underlying processes. This generic depiction could show the flight path of another animal obscured from vision or a scientist trying to find a law underlying a few noisy measurements. When trying to fit the observed data points with a curve, the difficulty lies in trading off the fitting error against the *complexity* of the hypothesis [9]. A linear model might be simple, for example, but will potentially incur large fitting errors. By contrast, a very flexible complex model will achieve low or even zero fitting error, but carries the danger of

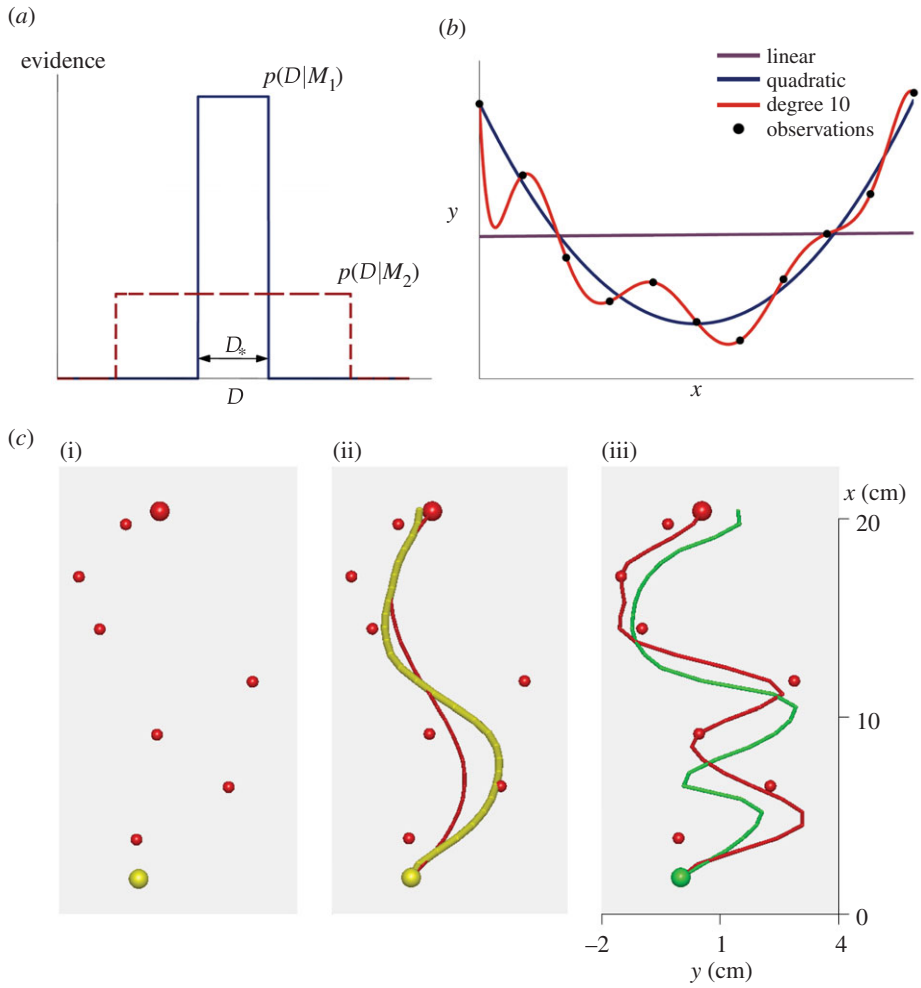

Figure 1. Bayesian Occam's Razor. (*a*) Schematic plot of evidence $p(D|M)$ for a simple model M_1 (blue, solid line) and a complex model M_2 (red, dashed line) for different data D, such as different random trajectories. Because both models have to spread unit probability mass over all compatible observations, the simpler model M_1 has a higher evidence in the overlapping region D_*. (*b*) Exemplary polynomials of different degrees fitted to noisy observations (black dots) by minimizing mean-squared error. The linear model (purple line) is not flexible enough to capture the underlying function and results in a large fitting error. The most complex model (red line) is too flexible and passes exactly through each data point, thus achieving a fitting error of zero. A reasonable fit should trade off the complexity of the model against the goodness of fit. In this example, this is achieved by the quadratic model (blue line). (*c*) Standard trials. (i) Noisy observations are shown to the participant as small red spheres. Start position (yellow sphere) and end position (larger red sphere) are noise-free. The yellow colour of the start sphere indicates that the underlying trajectory was drawn from M_1 with the long length scale. (ii) After completion of a standard trial, the underlying trajectory (red) is revealed and shown along with the participants trajectory (yellow). (iii) Example of a standard trial where the short-length-scale model M_2 was the generating model, as indicated by the green colour of the start position and participants trajectory. Note that the noisy observations are exactly the same in all three panels and the drawn trajectories were recorded from the same participant.

overfitting, which will ultimately lead to poor prediction and generalization. This trade-off between model fit and complexity is considered automatically by Bayesian inference, because determining the likelihood of a model requires consideration of not only the best-fitting parameter setting of each model—as this would virtually always lead to the preference of the more flexible model—but the *average* goodness of fit over all possible parameter settings of the model. A too complex model that implies many badly fitting parameter settings will therefore be disfavoured. Well-known approximations of this Bayesian complexity trade-off include counting the number of model parameters, as in the case of the Akaike or the Bayesian information criterion [10,11], but this simplification does not hold generally [8].

The goal of our study is to test whether Occam's Razor can be found in human sensorimotor control. This question is especially compelling, as recent studies have found evidence that the sensorimotor system integrates prior knowledge with new incoming information to make inferences about unobserved latent variables in a way that is consistent with Bayesian statistics [12–21]. We designed a sensorimotor task similar to the problem depicted in figure 1*b*, where participants were exposed to noisy observations similar to the black dots and had to guess and draw the curve they believed to be underlying the observations. Participants were trained on two different models: a simple model M_1 generating smooth trajectories and a complex model M_2 that could also explain more wiggly trajectories. They were then exposed to ambiguous stimuli to see whether they showed a preference for the simpler model.

2. Methods overview

In our virtual reality experiment, participants were shown noisy observation stimuli on a head-mounted display and could draw regression trajectories through these observation clouds with a manipulandum to indicate the presumed noiseless trajectory as shown in figure 1*c*. In standard training trials, participants were informed by a colour cue about the model (M_1 or M_2) that generated the observation stimulus so that they could learn the statistics

of each model. Figure 1*c*(ii) shows an example of a simple model M_1 trajectory, where the red curve shows the underlying function the participant is supposed to guess, and the yellow curve shows the participant's actually drawn trajectory. Figure 1*c*(iii) shows an example of a complex model M_2 trajectory that could explain the same observations. Participants were instructed to try and match the underlying functions as closely as possible. In standard training trials, the underlying curve (red) was revealed to participants after they had drawn their guessed trajectory. In probe trials, subjects were exposed to a stimulus of noisy observations sampled from one of the two models, but without giving them any additional cues or feedback about the model class or underlying curve. Participants were instructed that the observations were generated by one of the two models that they experienced during the standard training trials and that they should match the underlying curve as closely as possible, even though it was never revealed to them in probe trials.

To synthetically generate M_1 and M_2 trajectories, we used Gaussian Process (GP) models [22] that allow manipulating the trajectory smoothness with a single length-scale parameter. Simple M_1 trajectories could therefore be characterized by a long length scale λ_1, and complex M_2 trajectories could be characterized by a short length scale λ_2. To illustrate the complexity of the two models, we generated artificial observation datasets and ordered them according to the probability of the datasets under model M_1 (see [8] for a discussion on how to illustrate the hypothetical plot on real data). In the simulation results shown in electronic supplementary material, figure S1, it can be seen that the simpler model M_1 concentrates its probability mass on a smaller subset of data than model M_2, which corresponds in essence to the schematic plot in figure 1*a*. Another advantage of using GP models (with a Gaussian likelihood model) is that the log-probability of the model under uninformative prior can be expressed in closed form as

$$\log p(M_i|\mathbf{y}) \propto -\underbrace{\tfrac{1}{2}\mathbf{y}^{\mathrm{T}}\Sigma_{\lambda_i}^{-1}\mathbf{y}}_{\text{goodness of fit}} - \underbrace{\tfrac{1}{2}\log|\Sigma_{\lambda_i}|}_{\text{model complexity}}, \tag{2.1}$$

where Σ_{λ_i} is a covariance matrix associated with model i and $\mathbf{y}$ are the noisy observations. The first term is a data-driven goodness-of-fit error term. The second term is a data-independent complexity penalization and instantiates Occam's Razor. A more complex model, which is a model with shorter length scale λ and therefore more flexibility, is always associated with a higher complexity penalization. In physics, the complexity term is a log partition function that measures the effective number of possible states, which can also be interpreted in terms of the complexity of decision-making processes [23,24]. The complexity penalization term also has an information-theoretic interpretation within the framework of minimum description length [25] that specifies the minimum amount of information needed in a message that encodes the recorded data. The total message length $L = -\log p(D, M_i)$ can be decomposed into message length $L_{\mathrm{D}} = -\log p(D|M_i)$ needed to record the data under a given model M_i and the message length $L_{\mathrm{M}} = -\log p(M_i)$ required to specify the model itself. Complex models can encode data efficiently (low L_{D}), but require detailed model specifications (high L_{M}). Bayesian Occam's Razor consists in finding the model M_i with the shortest total message length $L = L_{\mathrm{D}} + L_{\mathrm{M}}$. It is important to note that the strength of the Razor crucially depends on the available class of models. For further details on the methods, see the electronic supplementary material.

3. Results

(a) Standard trials

Two example trajectories drawn by participants in standard trials can be seen in figure 1*c*. In both examples, the participant was shown the same noisy stimulus, but they drew different trajectories corresponding to the colour cue that informed them about the model (i.e. whether to expect a smooth or wiggly trajectory). To have a model-independent check of whether participants were able to distinguish the two model classes, we performed a Fourier analysis of their trajectories and found that in model M_2 trials trajectories had significantly increased higher-frequency components in their spectrum compared with model M_1 trials (cf. electronic supplementary material, figure S2). To further assess how well participants were able to learn the two model classes, we compared participants' trajectories with three different generation mechanisms: (i) samples from the posterior predictive GP with the correct length scale conditioned on the actual observations, (ii) a straight line connecting the first and last sample, and (iii) a connect-all trajectory that simply connects all the shown samples with straight lines. In particular, we computed the length scales of participants' trajectories by maximizing the marginal likelihood of the predictive distribution and compared these fitted length scales against the true model length scales. We found that the length scales inferred from participants' trajectories roughly match the true length scale, especially in the later part of the experiment, but, importantly, none of the non-probabilistic strategies—the straight-line (ii) or connect-all strategy (iii)—could generate length scales that were compatible with both trajectory types (cf. electronic supplementary material, figure S4). The fact that some of the estimated length scales were slightly lower than the length scales of the stimuli, especially in the early part of the experiment, does not necessarily imply that participants actually underestimated the length scale systematically, but could also be a consequence of a slight estimation bias towards lower length scales resulting from samples with mixed length scales (both too high and too low) (cf. electronic supplementary material, figure S5).

(b) Model choice in equal-error probe trials

To test for Occam's Razor directly, we introduced equal-error probe trials where the presented stimulus was ambiguous and could be explained by both models equally well. In these trials, the goodness-of-fit values in equation (2.1) were equal under both models. Importantly, in these trials, a decision-maker who does not care about model complexity would choose between the two models with 50 : 50 probability. We tested ambiguous stimuli with three different goodness-of-fit values (small, medium, large). Figure 2*a* shows subjects' actual choice behaviour in these trials. Subjects' choice probabilities were obtained by classifying their drawn trajectories into the two model classes. For all three error levels, we found that subjects significantly preferred the simpler model M_1 ($p < 0.001$, sign test against median of 0.5).

An alternative explanation for the preference for the simpler model could be the lower physical effort of drawing a smooth trajectory as opposed to a wiggly one. To control for this effect, we designed a control experiment in which participants did not draw the trajectory, but they clicked mouse buttons to indicate the model class. We randomly assigned

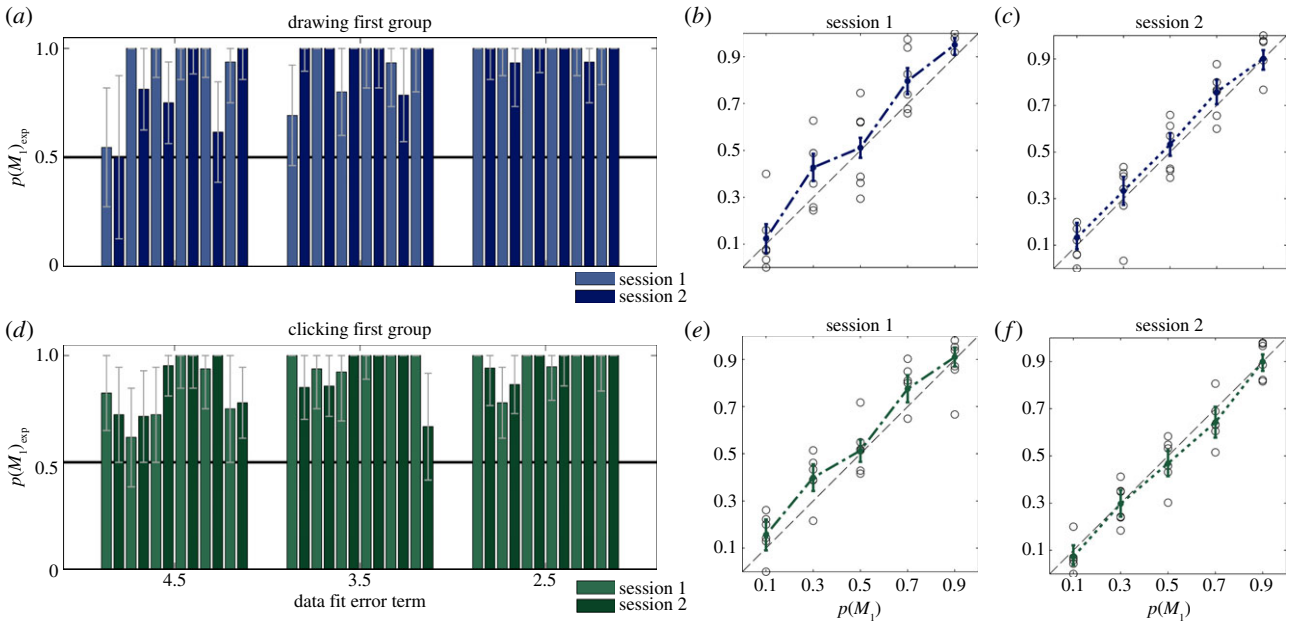

Figure 2. (*a*,*d*) Occam's Razor observed in the experiment. The plots show the choice probability for the simple model M_1 when presented with a stimulus that has equal goodness of fit for both models. Each bar corresponds to the choice behaviour of one participant in a particular error condition—error bars show the 95% CI. (*a*) Results for the group that performed the drawing session first and then the clicking session. (*d*) Results for the clicking first, then drawing group. Both groups show a clear bias towards the simpler model M_1. (*b*,*c*,*e*,*f*) Pooled group choice probabilities. Circles represent individual participants median choice probability, the thick line (dash-dotted in the first session and dotted in the second session) shows the median using pooled data of all participants along with 95% CIs. The dashed black line illustrates the ideal case, where theoretical choice probabilities and observed choice behaviour match exactly. (*b*,*c*) Results for the group that performed the drawing session first. (*e*,*f*) Results for the group that performed the clicking session first. The individually fitted choice probabilities are shown in the electronic supplementary material, figure S6.

subjects into two groups, one of them performing the *drawing* session before the *clicking* session—and the second one performing the two sessions in reverse order. In terms of the observed choice behaviour in ambiguous trials, we found no systematic difference between the two groups (figure 2*d*).

(c) Model choice in general probe trials

While ambiguous stimulus trials directly demonstrate Occam's Razor, the framework of Bayesian modelling allows for more general quantitative predictions. In particular, Bayesian models weight goodness of fit and model complexity for arbitrary stimuli, including the ambiguous ones just described. The comparison between the experimentally observed choice probabilities and the fitted theoretical choice probabilities for all probe trials is shown in figure 2*b*,*c*,*e*,*f* using the pooled data across all subjects both for drawing and clicking trials. Theoretical and empirical choice probabilities are in good agreement, as the 95% CIs mostly overlap with the identity line. The corresponding comparison of theoretical and empirical choice behaviour of individual participants can be seen in the electronic supplementary material, figure S6. To obtain the theoretical choice probabilities, we fitted psychometric sigmoid functions to subjects' model choice in probe trials, where we used the difference in model log-evidence from equation (2.1) as the discrimination variable [26]. The psychometric sigmoid function has a single parameter α to tune the sensitivity of the decision boundary. The α-values were fitted by maximizing the likelihood and are shown in the electronic supplementary material, table S1. The experimental choice probabilities for different stimuli were obtained by first discretizing the theoretical choice probabilities into five equidistant bins, and then determining the choice frequency of model M_1 for all five bins. To quantitatively assess the explanatory power of the individual fits, we performed linear regression on the individually observed choice patterns. For all participants and all conditions, the ideal slope of 1 lies within the 95% CIs of the fitted slope. The ideal intercept of 0 lies within the 95% CIs for all subjects, except for subject 2 in the first session (cf. electronic supplementary material, figure S7).

(d) Control experiment: spatial frequency

Finally, we tested for trajectory smoothness as a confounding variable. In the choice trials considered so far, the simple model was always the model with less spatial frequency. To test whether subjects really cared about model complexity rather than spatial frequency, we designed a control experiment where the simpler model had a higher spatial frequency than the more complex model. This unusual scenario can be created by modulating the noise and signal variability of the two models. If model M_2 generates wiggly trajectories that are highly reproducible across trials (i.e. with low variability), it constitutes a simple model, because it cannot explain many different datasets. If, by contrast, model M_1 generates smooth trajectories with low spatial frequency, but with high noise and signal variability, it might be able to explain more datasets than the wiggly model. To test whether the critical variable was indeed trajectory smoothness or model complexity, we therefore trained a group of participants on these two models and exposed them to probe trials in which the goodness-of-fit values were equal under both models (see the electronic supplementary material for details). If participants cared about smoothness, they should prefer the more complex model M_1 in these

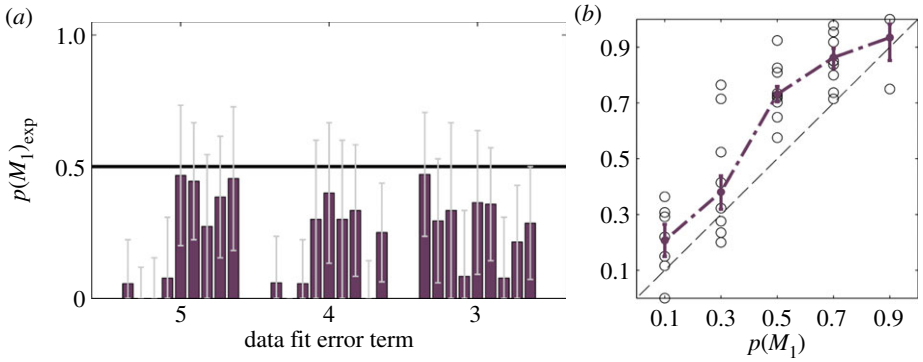

Figure 3. Control experiment where the simpler model M_2 had a higher spatial frequency. (*a*) Occam's Razor observed in the control experiment. The plots show the choice probability for the more complex model M_1 when presented with a stimulus with equal goodness of fit for both models. A bias towards the simpler but higher-spatial-frequency model M_2 can be seen. (*b*) Pooled group choice probabilities in the control experiment. Circles represent individual participants' median choice probability and the purple line shows the median using pooled data of all participants. The dashed black line illustrates the ideal case, where theoretical choice probabilities and observed choice behaviour match exactly. Error bars show the 95% CI. The individually fitted choice probabilities are shown in electronic supplementary material, figure S8.

trials. If, however, participants cared about model complexity as the critical variable, they should prefer the simpler, but more wiggly model M_2. Participants' choice probabilities can be seen in figure 3*a*. For all three error levels, we found that subjects significantly preferred the simpler model M_1 ($p < 0.001$, sign test against median of 0.5). Compared with figure 2*a*,*d*, it can be seen that the choice probabilities indicate a more stochastic choice behaviour, because the discrimination in the probe trials was more difficult owing to the different noise and signal variances of the two models. This is also reflected in more shallow psychometric curves in the probe trials and the associated lower α-values quantifying their reduced steepness (cf. electronic supplementary material, table S2). The choice probabilities in all probe trials are shown in figure 3*b*. To quantitatively assess the fitted choice probabilities, we performed linear regression on the individually observed choice patterns (cf. electronic supplementary material, figure S8). For all participants and all conditions, the ideal slope of 1 lies within the 95% CIs of the fitted slope. The ideal intercept of 0 lies within the 95% CIs for 5 out of 10 subjects (cf. electronic supplementary material, figure S9). While these results cannot completely rule out a weak smoothness bias, they clearly show that participants' choices are modulated by Bayesian model complexity, and that participants tend to prefer the simpler model and not the smoother model when both fit the observations equally well (cf. figure 3*a*).

4. Discussion

To test whether Occam's Razor plays a role in human sensorimotor learning, we designed a visuomotor experiment where participants had to produce a regression trajectory from noisy observations generated by one of two possible models with different complexity. Participants were trained on both generative models and were then presented with ambiguous stimuli where both models were able to explain the observed data equally well. We considered five different hypotheses: (1) subjects prefer the simpler model (Occam's Razor); (2) subjects are indifferent between the two models; (3) subjects ignore both models and either follow a straight-line or connect-all strategy; (4) subjects decide based on physical effort; and (5) subjects decide based on trajectory smoothness (spatial frequency). In accordance with Occam's Razor, we found that participants showed preference for the simpler model in ambiguous trials. Over all trials, we found that their behaviour was quantitatively consistent with Bayesian Occam's Razor trading off goodness of fit and model complexity. To control for the influence of physical effort required for drawing regression trajectories, we designed a control experiment where the indication of either model required the same physical effort. We found that subjects' preferences were essentially unaltered, suggesting that the difference in effort required for drawing the two different trajectory types was negligible and that their choices were indeed affected by the underlying trajectory complexity. We also designed a control experiment where the simpler model implied underlying curves with a high spatial frequency but low variability across trials, and found that participants' choices were mainly governed by model complexity and not trajectory smoothness.

In our study, we made a number of simplifying modelling assumptions that might limit the generalizability of our results: we assumed that trajectories can be well described by a Gaussian Process (GP) model, we assumed a squared exponential kernel for the GP, and we assumed a fixed observation noise set by the experimenter. First, the reasons for choosing a GP model in our experiment include their mathematical tractability, their clear distinction between model complexity and data complexity, they allow modelling very general smooth trajectories since GPs are non-parametric, and they have been previously shown to adequately capture human motion [27]. Second, we assumed that trajectories generated with a squared exponential kernel can be adequately mimicked by human motion. Our assumption is based on the close relationship between squared exponential kernels and radial basis function networks that have been previously suggested for modelling human sensorimotor processing [28]. To test the appropriateness of this modelling assumption, we also fitted a neural network kernel [22,29] to participants' motion trajectories and found that the squared exponential kernel provided a better explanation for all subjects in all conditions (cf. electronic supplementary material, figure S10). The reason for this could simply be that the synthetic trajectories were generated from the squared exponential kernel and that participants were able to learn this.

Third, in our fits we assumed that subjects learned the variance of the observation noise that enters as an additive constant on the diagonal of the covariance matrix Σ_{λ_i} in equation (2.1). This assumption was motivated by the fact that the observation noise stayed the same throughout the experiment and was set by the experimenter. The magnitude of this observation noise was on the order of centimetres, and thus far larger than perceptual noise owing to natural limitations of visual acuity. We also evaluated the predictive likelihood of a range of length scales and observation noise values for participants' movement trajectories, and found that the likelihoods were sharply peaked within the neighbourhoods of the experimentally induced values, suggesting that our assumption regarding the observation noise was reasonable (cf. electronic supplementary material, figure S11).

The problem of disambiguating competing explanatory hypotheses when faced with ambiguous stimuli has been previously investigated. In fact, the human sensorimotor system is frequently confronted with such decision-making situations, for instance when perceiving a visual or motion illusion [30,31]. Interestingly, the emergence of many illusions can be explained within the Bayesian framework by using priors that reflect environmental statistics [32–35], including priors for light-from-above illumination in object perception [36,37], priors for low number of categories in object categorization [38] and priors for lower speed in motion perception [30,39]. Deciding for a particular hypothesis can also be thought of as choosing the simpler explanation, which highlights the important role of the prior in defining what is simple or difficult. As the prior is subjective, it can encode information about the statistics of the stimulus, but also subject-specific features like cognitive difficulty.

The problem of model selection has been recently addressed in a number of studies. Körding *et al*. [40] investigated integration versus segregation of audio-visual stimuli in human subjects, which can be cast as selecting between two different models: in one model, there was only one source explaining both stimuli (auditory and visual), and in the other model there were two different sources at different locations explaining both stimuli. When subjects reported their perception of unity [41], they were effectively doing inference over the two different models. Another recent study [42] introduced a sensorimotor paradigm for Bayesian model selection. In their task, subjects had to point to one of two targets (representing two models) after observing a cursor shift (representing the model parameter) drawn from one of two possible distributions (the prior over the model parameters given either model). When facing ambiguous visual feedback of the shift parameter, participants had to 'integrate out' the compatible range of parameter values. It was found that their choice behaviour was consistent with Bayesian model selection. In motor control, selecting between different models is relevant in the context of structure learning [43–49], where abstract invariants form structures or abstract models that are applicable to a range of motor tasks.

In contrast to these previous studies on model selection, our current study explicitly investigates the trade-off between fitting error and model complexity. We therefore designed a sensorimotor regression task based on GP models that allowed for an analytic expression of the model complexity, which we could exploit for the design of ambiguous probe trials. Thus, we could perform a quantitative analysis of the preference for simpler models in probe trials and compare against human behaviour. The main contribution of this study is the illustration of Occam's Razor as it is depicted in figure 1*a* in the human sensorimotor system. Constructing tasks that allow this interpretation of model complexity is not trivial [8]. Our study thus adds a new angle to the growing number of psychophysical experiments where participants' behaviour was successfully modelled within a Bayesian framework.

The study was approved by the local ethics committee, and all participants were naive and gave informed consent.

Conflict of interest statement. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data accessibility. The primary data for the study are publicly available at Dryad doi:10.5061/dryad.94jt2.

Funding statement. This study was supported by the DFG, Emmy Noether grant no. BR4164/1-1.

References

1. Wolpert DM, Ghahramani Z. 2000 Computational principles of movement neuroscience. *Nat. Neurosci.* **3**(Suppl), 1212–1217. (doi:10.1038/81497)
2. Knill DC, Pouget A. 2004 The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719. (doi:10.1016/j.tins.2004.10.007)
3. Ma WJ, Beck JM, Latham PE, Pouget A. 2006 Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438. (doi:10.1038/nn1790)
4. Shadmehr R, Smith MA, Krakauer JW. 2010 Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108. (doi:10.1146/annurev-neuro-060909-153135)
5. Franklin DW, Wolpert DM. 2011 Computational mechanisms of sensorimotor control. *Neuron* **72**, 425–442. (doi:10.1016/j.neuron.2011.10.006)
6. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011 How to grow a mind: statistics, structure, and abstraction. *Science* **331**, 1279–1285. (doi:10.1126/science.1192788)
7. Mackay DJC. 2003 *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.
8. Murray I, Ghahramani Z. 2005 A note on the evidence and Bayesian Occam's razor. Technical report, Gatsby Unit. See http://mlg.eng.cam.ac.uk/zoubin/papers/05occam/occam.pdf.
9. Bishop CM. 2006 *Pattern recognition and machine learning*, vol. 4. Berlin, Germany: Springer.
10. Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
11. Schwarz G. 1978 Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464. (doi:10.1214/aos/1176344136)
12. van Beers RJ, Sittig AC, Gon JJ. 1999 Integration of proprioceptive and visual position information: an experimentally supported model. *J. Neurophysiol.* **81**, 1355–1364.
13. Ernst MO, Banks MS. 2002 Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433. (doi:10.1038/415429a)
14. Todorov E. 2004 Optimality principles in sensorimotor control. *Nat. Neurosci.* **7**, 907–915. (doi:10.1038/nn1309)
15. Körding KP, Wolpert DM. 2004 Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247. (doi:10.1038/nature02169)
16. Körding KP, Wolpert DM. 2006 Bayesian decision theory in sensorimotor control. *Trends Cog. Sci.* **10**, 319–326. (doi:10.1016/j.tics.2006.05.003)

17. Körding K. 2007 Decision theory: what should the nervous system do? *Science* **318**, 606–610. (doi:10.1126/science.1142998)
18. Wolpert DM. 2007 Probabilistic models in human sensorimotor control. *Hum. Mov. Sci* **26**, 511–524. (doi:10.1016/j.humov.2007.05.005)
19. Hudson TE, Maloney LT, Landy MS. 2007 Movement planning with probabilistic target information. *J. Neurophysiol.* **98**, 3034. (doi:10.1152/jn.00858.2007)
20. Girshick AR, Banks MS. 2009 Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *J. Vis.* **9**, 8.1–8.20. (doi:10.1167/9.9.8)
21. Turnham EJA, Braun DA, Wolpert DM. 2011 Inferring visuomotor priors for sensorimotor learning. *PLoS Comput. Biol.* **7**, e1001112. (doi:10.1371/journal.pcbi.1001112)
22. Rasmussen CE, Williams C. 2006 *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.
23. Ortega PA, Braun DA. 2013 Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A* **469**, 20120683. (doi:10.1098/rspa.2012.0683)
24. Ortega PA, Braun DA. 2011 Information, utility and bounded rationality. In *Proceedings of the 4th International Conference on Artificial General Intelligence, Mountain View, CA, August 2011* (eds J Schmidhuber, KR Thorisson, M Looks), pp. 269–274. Berlin, Germany: Springer.
25. Rissanen J. 1978 Modeling by shortest data description. *Automatica* **14**, 465–471. (doi:10.1016/0005-1098(78)90005-5)
26. Kass RE, Raftery AE. 1993 Bayes factors and model uncertainty. *J. Am. Stat. Assoc.* **90**, 466.
27. Wang JM, Fleet DJ, Hertzmann A. 2008 Gaussian process dynamical models for human motion. *IEEE Trans. PAMI* **30**, 283–298. (doi:10.1109/TPAMI.2007.1167)
28. Pouget A, Snyder LH. 2000 Computational approaches to sensorimotor transformations. *Nat. Neurosci.* **3**, 1192–1198. (doi:10.1038/81469)
29. Williams CK. 1998 Computation with infinite neural networks. *Neural Comput.* **10**, 1203–1216. (doi:10.1162/089976698300017412)
30. Weiss Y, Simoncelli EP, Adelson EH. 2002 Motion illusions as optimal percepts. *Nat. Neurosci.* **5**, 598–604. (doi:10.1038/nn0602-858)
31. Taylor JL, McCloskey DI. 1991 Illusions of head and visual target displacement induced by vibration of neck muscles. *Brain* **114**, 755–759. (doi:10.1093/brain/114.2.755)
32. Geisler WS, Kersten D. 2002 Illusions, perception and bayes. *Nat. Neurosci.* **5**, 508–510. (doi:10.1038/nn0602-508)
33. Kersten D, Mamassian P, Yuille A. 2004 Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304. (doi:10.1146/annurev.psych.55.090902.142005)
34. Sato Y, Toyoizumi T, Aihara K. 2007 Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* **19**, 3335–3355. (doi:10.1162/neco.2007.19.12.3335)
35. Brayanov JB, Smith MA. 2010 Bayesian and 'anti-bayesian' biases in sensory integration for action and perception in the size-weight illusion. *J. Neurophysiol.* **103**, 1518–1531. (doi:10.1152/jn.00814.2009)
36. Langer MS, Bülthoff HH. 2001 A prior for global convexity in local shape-from-shading. *Perception* **30**, 403–410. (doi:10.1068/p3178)
37. Adams WJ, Graf EW, Ernst MO. 2004 Experience can change the 'light-from-above' prior. *Nat. Neurosci.* **7**, 1057–1058. (doi:10.1038/nn1312)
38. Gershman SJ, Niv Y. 2013 Perceptual estimation obeys Occam's razor. *Front. Psychol.* **4**, 623. (doi:10.3389/fpsyg.2013.00623)
39. Stocker AA, Simoncelli EP. 2006 Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* **9**, 578–585. (doi:10.1038/nn1669)
40. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. 2007 Causalinference in multisensory perception. *PLoS ONE* **2**, e943. (doi:10.1371/journal.pone.0000943)
41. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. 2004 Unifying multisensory signals across time and space. *Exp. Brain Res.* **158**, 252–258. (doi:10.1007/s00221-004-1899-9)
42. Genewein T, Braun DA. 2012 A sensorimotor paradigm for Bayesian model selection. *Front. Hum. Neurosci.* **6**, 291. (doi:10.3389/fnhum.2012.00291)
43. Braun DA, Aertsen A, Wolpert DM, Mehring C. 2009 Motor task variation induces structural learning. *Curr. Biol.* **19**, 352–357. (doi:10.1016/j.cub.2009.01.036)
44. Braun DA, Aertsen A, Wolpert DM, Mehring C. 2009 Learning optimal adaptation strategies in unpredictable motor tasks. *J. Neurosci.* **29**, 6472–6478. (doi:10.1523/JNEUROSCI.3075-08.2009)
45. Narain D, Mamassian P, van Beers RJ, Smeets JBJ, Brenner E. 2013 How the statistics of sequential presentation influence the learning of structure. *PLoS ONE* **8**, e62276. (doi:10.1371/journal.pone.0062276)
46. Kemp C, Tenenbaum JB. 2008 The discovery of structural form. *Proc. Natl Acad. Sci. USA* **105**, 10 687–10 692. (doi:10.1073/pnas.0802631105)
47. Kemp C, Tenenbaum JB. 2009 Structured statistical models of inductive reasoning. *Psychol. Rev.* **116**, 20–58. (doi:10.1037/a0014282)
48. Haruno M, Wolpert DM, Kawato M. 2001 MOSAIC model for sensorimotor learning and control. *Neural Comput.* **13**, 2201–2220. (doi:10.1162/089976601750541778)
49. Braun DA, Waldert S, Aertsen A, Wolpert DM, Mehring C. 2010 Structure learning in a sensorimotor association task. *PLoS ONE* **5**, e8973. (doi:10.1371/journal.pone.0008973)

Electronic Supplementary Material
Occam's Razor in sensorimotor learning

Tim Genewein[1,2,3,*], Daniel A. Braun[1,2]
[1] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2] Max Planck Institute for Intelligent Systems, Tübingen, Germany
[3] Graduate Training Centre of Neuroscience, Tübingen, Germany
∗ **Correspondence:**
Tim Genewein, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany
E-mail: tim.genewein@tuebingen.mpg.de

Materials and Methods

Participants. Eight female and thirteen male participants were recruited from the student population of the University of Tübingen. The study was approved by the local ethics committee and all participants were naive and gave informed consent. The local standard rate of eight Euros per hour was paid for participation in the study. One subject was excluded from the study because the choice behavior was statistically indistinguishable from random behavior.

Materials. We used a virtual reality setup consisting of a $\mathrm{Sensable}^{\circledR}$ $\mathrm{Phantom}^{\circledR}$ Premium 1.5 High Force manipulandum for tracking participants' hand movements in three dimensions and an $\mathrm{NVIS}^{\circledR}$ nVisor ST50 head-mounted display (HMD) for creating stereoscopic 3D virtual reality. Movement position and velocity were recorded with a rate of 1kHz. The manipulandum was operated with a weak isotropic viscous force field of $\vec{f} = -\alpha\dot{\vec{x}}$, where $\alpha = 0.02\frac{Ns}{cm}$ and $\dot{\vec{x}}$ is the three-dimensional velocity vector. The sole purpose of the force field was to prevent very fast movements and could never push subjects into an unwanted movement direction. All movements with the manipulandum during the execution of a trial were projected into the horizontal plane. The workspace was $18.5cm$ in the forward-backward direction (the x-axis) and $20cm$ in the left-right direction (y-axis).

Occam's Razor. Occam's razor takes a particularly succinct form for Gaussian Process models. A Gaussian process can be thought of as a distribution over functions

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

fully specified by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$ such that

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \end{aligned}$$

In our models we assumed $m(\mathbf{x}) \equiv 0$ and a squared exponential covariance function

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2\lambda^2}(\mathbf{x} - \mathbf{x}')^2\right) \tag{1}$$

with length scale λ and signal variance σ_f^2. Example functions $f(\mathbf{x})$ with two different length scales can be seen in Figure 1C in the main article (red curves). Importantly, in our experiments subjects

did not directly observe the functions $f(\mathbf{x})$, but only a finite set of noisy samples $y_i = f(\mathbf{x}_i) + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$ and $i = 1, \ldots, N$. For notational convenience, the observation samples can be represented by a vector $\mathbf{y} = \{y_1, \ldots, y_N\}$ and the input locations as a matrix $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. The covariance function at the input locations can then be abbreviated as the matrix $K_\lambda = K_\lambda(X, X)$ for a particular choice of λ. In general, the inputs $\mathbf{x}_i$ can be vectorial (hence the bold face notation), but in our experiment the input is one-dimensional. In the text we refer to the drawn functions $f(\mathbf{x})$ also as *trajectories*, in particular when comparing them to subjects' movement trajectories, even though only the spatial coordinates of subjects' movements were used for analysis, and all temporal information was discarded.

In terms of Bayesian model comparison each length scale λ is a different *model* M_λ that induces a different probability distribution $P(\mathbf{f}|X, M_\lambda)$ over the function space. To obtain the posterior probability $P(M_\lambda|\mathbf{y}, X)$ of the model M_λ given the observations $\mathbf{y}$ and the input locations X, we can apply Bayes' rule $P(M_\lambda|\mathbf{y}, X) \propto P(\mathbf{y}|X, M_\lambda)P(M_\lambda)$, where $P(M_\lambda)$ is the prior over the different models and $P(\mathbf{y}|X, M_\lambda)$ is the marginal likelihood of the data for model M_λ. For flat prior probabilities $P(M_\lambda)$, the decisive factor for model comparison is the marginal likelihood, which for our zero-mean Gaussian process (with a Gaussian likelihood model) is given by

$$\log P(\mathbf{y}|X, M_\lambda) = \underbrace{-\frac{1}{2}\mathbf{y}^T(K_\lambda + \sigma_n^2 I_N)^{-1}\mathbf{y}}_{\text{data fit error term}} \underbrace{-\frac{1}{2}\log|K_\lambda + \sigma_n^2 I_N|}_{\text{complexity term}} - \frac{N}{2}\log 2\pi, \tag{2}$$

where I_N is the identity matrix of dimension N. Occam's razor can be seen clearly in the second term. For fixed input locations, the third term of the marginal likelihood is a negligible constant. The first term is a data-driven goodness-of-fit error term. The second term is a complexity penalization. A more complex model, that is a model with shorter length scale λ and therefore more flexibility, is always associated with a higher complexity penalization. A schematic plot of the Occam's Razor effect in Bayesian inference is shown in Figure 1A. To illustrate this effect with the two same GP models as used in our work, we generated $20,000$ random trajectories and computed their evidence according to Equation 2—the results are shown in Supplementary Figure 1B,C.

Experimental design. We designed a planar visuomotor task where participants had to draw a regression curve through a number of points that represented noisy observations $\mathbf{y}$ of a computer-generated noiseless trajectory $\mathbf{f}$—compare Figure 1C in the main article. The noiseless trajectory could be generated either by a simple model M_1 or by a complex model M_2. The simple model generated noiseless trajectories by means of a Gaussian Process with long-ranging correlations λ_1, and the complex model was a Gaussian Process with short-ranging correlations $\lambda_2 < \lambda_1$. Accordingly, complex trajectories of model M_2 were more flexible and wiggly than the simpler and straighter trajectories of model M_1. By designing trials with observations that can be explained equally well by both models, we could then study whether subjects preferred the simpler model, as suggested by Occam's razor.

Each subject performed two different sessions. In the *drawing session* subjects' goal was to match the underlying noiseless trajectory as closely as possible with their movement trajectory drawn by means of the manipulandum. Subjects' movement trajectory was automatically classified as an M_1- or an M_2-trajectory based on the marginal log-likelihood score. In the *clicking session* subjects decided whether the observed stimulus was generated by the simpler model M_1 or by the more complex model M_2 by pressing either the left or right mouse button respectively. The clicking session was introduced to ensure that subjects' model selection was not affected by the effort required for drawing.

Each session consisted of two kinds of trials: *standard trials* and *probe trials*. In *standard trials* subjects knew which model generated the observations. They could learn the statistics of the two models, as they received feedback of the underlying noiseless trajectory after each trial. In *probe trials* subjects did not know the generating model and had to decide which model generated the data—either by pressing a mouse button (clicking session) or by drawing a trajectory (drawing session). In probe trials subjects did not receive any feedback regarding the noiseless trajectory.

Stimulus generation. In all trials and sessions the stimulus generation process was very similar. In each trial, one of the two models was selected with $50:50$ probability and the noiseless trajectory $\mathbf{f}$ was sampled from the corresponding GP. In case of the simple model M_1 the length scale was $\lambda_1 = 8cm$, in the case of the complex model M_2 the length scale was $\lambda_2 = 2cm$. The signal variance was $\sigma_f^2 = 2.5^2 cm^2$ for both cases. The trajectory was sampled at 29 points equidistantly spaced along the x-axis of the workspace (the forward-backward direction). At the beginning of the trial this noiseless trajectory was never shown. Instead, only the first and last point of the trajectory were shown as well as 7 other points in the middle of the trajectory, that is points with the number $\{2, 6, 10, 14, 18, 22, 26\}$. The proximal locations of these points along the forward-backward axis correspond to X in Equation 2. The first and last point were shown veridically, the 7 middle points had zero-mean Gaussian noise added with $\sigma_n^2 = 1.25^2 cm^2$. This allowed us to always make the start position part of the trajectory. Together the lateral position of these 9 points made up the observation vector $\mathbf{y}$. The first point of the trajectory was always at the same location in the middle of the lateral workspace closest to the subject, which was ensured by conditioning the Gaussian Process on the start location when sampling the 29 points of the trajectory.

In the set of *standard trials* we selected 30 trials at random that were presented twice with exactly the same observation pattern, but the second time in a trial labeled to belong to the other model. This way we could inspect how the exact same stimulus pattern is interpreted depending on the presumed model—see Figure 1C in the main article. In the set of *probe trials* we also designed special "equal error probe trials" that occurred randomly with probability 0.2, that is roughly a fifth of the probe trials. In these trials the error-term of Equation 2 was (approximately) the same under both models. To allow for enough trials for averaging we restricted the possible errors to three different nat value clusters $(2.5 \pm 0.5, 3.5 \pm 0.5, 4.5 \pm 0.5)$ corresponding to small, medium and large fitting errors. We could then investigate whether subjects showed a clear bias in these trials towards the simpler model as predicted by Occam's razor.

Trial Setup. All participants performed two sessions consisting of 650 trials each with a break of at least one day in between the sessions. In the *drawing session*, the participants had to draw the trajectory in the virtual work space using the manipulandum. In the *clicking session*, the participants had to indicate their model choice by either pressing the left mouse button (corresponding to M_1) or the right mouse button (corresponding to M_2). Twelve participants were split into two groups, where Group I performed the drawing session first, followed by the clicking session, and Group II performed the clicking session first, followed by the drawing session.

Of the 650 trials per session, the first 50 trials were used to get the participants acquainted with the virtual reality setup, the operation of the manipulandum and the standard trials. The first 25 trials of each session consisted of M_1-standard trials only, followed by 25 standard trials exclusively of type M_2. These initial training trials were discarded from the evaluation of the experimental data.

In the remaining 600 trials of each session there were two trial types: *standard trials* and *probe trials*. Standard trials occurred with probability 0.65. In these trials subjects were indicated the underlying

model class and they received feedback about the underlying noiseless trajectory after the trial. Probe trials occurred with probability 0.35. In these trials subjects were not informed about the underlying model class and did not receive any feedback about the noiseless trajectory.

Standard Trials: Drawing Session. In standard trials of the *drawing session*, the start location was displayed as a blue sphere (radius $0.6cm$). Subjects had to move a cursor (blue sphere, radius $0.4cm$) representing their three-dimensional hand position into the start sphere and remain there for $0.1s$ to initiate the trial. A beep occurred to indicate the start of the trial and subjects had to initiate their movement within $2s$ after the beep. At the same time as the beep, noisy observations of the noiseless trajectory were shown (red spheres, radius $0.2cm$). The first observation data point (that coincides with the start-position) was shown as a sphere with radius $0.4cm$ either in yellow (if the true model was M_1) or green (if the true model was M_2). The cursor color changed to yellow or green as well. The final observation point was displayed as a red sphere with radius $0.5cm$. In total, these were the 9 observation points used for stimulus generation and labeled $\mathbf{y}$. The stimulus is depicted in Figure 1C in the main article.

To further indicate the correct generating model, a landscape-cue was shown in the background of the workspace sketching a mountain range with either smooth or jagged ridge corresponding to model M_1 and M_2 respectively. The ridge was sampled from a GP with the same length scale as the true model, but reduced signal variance ($\sigma_f^2 = 0.7^2 cm^2$). The landscape-cue had the same color as the cursor and the start position. Thus, subjects had both a color and shape cue indicating the model class, and therefore the smoothness of the underlying noiseless trajectory. Participants were instructed about the relation between the generating model, the smoothness of the underlying trajectory and the color of the first observation point, the cursor and the landscape cue, as well as the jaggedness of the landscape cue.

Subjects had a time-window of $7s$ to draw the trajectory they believed to match most closely the unobserved underlying noiseless trajectory $\mathbf{f}$. The trial ended when subjects crossed the distal boundary of the workspace at $18.5cm$. During movement, the trace of participants' movements was drawn as a trajectory and displayed either as a yellow or as a green line, depending on the correct model. As an additional cue, the green line was three times thicker than the yellow line. After completion of a standard trial, the noiseless underlying trajectory was shown as a red line along with the participants' completed trajectory. The participants' trajectory was classified as the model with larger marginal likelihood score. If the trajectory was assigned to the wrong model, an error-beep was played back, the screen was flashed red for $0.2s$ and the participants' trajectory was colored with the color corresponding to the wrong model. With this feedback, participants could learn to draw trajectories that could be clearly assigned to their intended model. We found that participants could do this easily after experiencing very few trials. If the timing conditions were not met, the trial had to be repeated.

Standard Trials: Clicking Session. In the *clicking session* the participant did not draw a trajectory, but indicated the model choice by using the mouse buttons. As in the standard trials of the drawing session, the participant observed 9 noisy observations. Participants had $2s$ to indicate their choice by clicking either the left or the right mouse button. If these timing-conditions were violated, the trial was aborted and had to be repeated. If the chosen model was the correct model, then the noiseless trajectory $\mathbf{f}$ was shown in yellow or green depending on the model, otherwise another trajectory was sampled from the wrong model conditioned on the displayed stimulus points (yellow or green color) and shown together with the noiseless trajectory (red color).

Probe Trials: Drawing Session. In *probe trials* the color of the start-sphere, cursor and landscape-cue was always white, and thus did not provide any information about the underlying model. Additionally, the landscape-cue was rectangular becoming more and more opaque towards the upper edge, such that the jaggedness of the landscape was not visible — resembling a hazy landscape devoid of any shape cue. The noisy observation points were drawn in red as in the standard trials. In probe trials, subjects were told that the noiseless trajectory was generated by one of the two possible models and that they should again draw a trajectory that they believed would match the noiseless underlying trajectory as closely as possible. Subjects movement trajectories were also drawn in white color. Importantly, participants did not receive any feedback about the noiseless underlying trajectory or whether their choice was correct or incorrect. This prevented participants from learning the statistics of the probe trials. The timing requirements were the same as in the standard trials.

Probe Trials: Clicking Session. In probe trials during the *clicking session*, participants saw the same stimulus as in the probe trials of the drawing session and indicated their model choice by pressing the left or right mouse button corresponding to choosing model M_1 and M_2. Then they observed a sampled trajectory from the chosen model conditioned on the observed data points. In contrast to the standard trials, this trajectory was colored white, regardless of the participants choice. No Feedback was provided and the timing requirements were the same as in the standard trials of the click session.

Theoretical Choice Probabilities. We predict the participants probability of choosing model M, given the noisy observations $\mathbf{y}$ expressed as a sigmoid-function

$$P(M_1|\mathbf{y}) = \frac{1}{1 + e^{-\alpha \log \mathcal{BF}}}, \tag{3}$$

where $\mathcal{BF}$ is the marginal likelihood ratio (Bayes Factor) that quantifies the evidence of model M_1 compared to the evidence of model M_2. In the framework of *Bayesian model selection* the Bayes Factor results from comparing the posterior probabilities of the two models

$$\frac{P(M_1|\mathbf{y})}{P(M_2|\mathbf{y})} = \frac{P(\mathbf{y}|M_1)P(M_1)}{P(\mathbf{y}|M_2)P(M_2)} = \frac{P(\mathbf{y}|\mathbf{x}_{\text{obs}}, M_1)}{P(\mathbf{y}|\mathbf{x}_{\text{obs}}, M_2)} = \mathcal{BF}, \tag{4}$$

which holds whenever $P(M_1) = P(M_2)$ as in our case. For Gaussian Processes the Bayes Factor can be computed by the ratio of marginal likelihoods given in Equation 2. If $\mathcal{BF}$ is greater than 1 then M_1 is the more likely explanation, and if $\mathcal{BF}$ is smaller than 1 then M_2 is more likely. Since $P(M_2|\mathbf{y}) = 1 - P(M_1|\mathbf{y})$ we can write

$$\log \mathcal{BF} = \log P(M_1|\mathbf{y}) - \log(1 - P(M_1|\mathbf{y})), \tag{5}$$

which leads precisely to Equation 3 for $\alpha = 1$. For $\alpha = 1$ subjects' policy corresponds to *probability matching*, for higher α the model predicts a more deterministic choice strategy. For evaluating the experimental data, this parameter α was fitted to each individual participants' probe trial data by maximizing the log likelihood of their actual choices.

Trajectory Classification. In order to classify trajectories drawn by subjects either as M_1- or M_2-trajectories, we first extracted the 29 points of the raw trajectory (sampled at $1kHz$) that were closest in x-position to the 29 points used for sampling the GP trajectories. We then computed the marginal likelihood of these 29 points for the two models according to Equation 2. Instead of the noise variance σ_n^2 that was used to generate the noisy observations, we used a variance of $\sigma_{\text{subj}}^2 = 0.8^2 cm^2$ that was determined in a pilot study to achieve the lowest number of misclassification in standard trials.

Control Experiment. In the control experiment the two generative models had different parameter settings. The experiment itself was conducted as a Drawing Session with standard trials and probe trials. In this case model M_1 was the more complex model with length scale $\lambda_1 = 6\text{cm}$, signal variance $\sigma^2_{1,f} = 0.75\text{cm}^2$, and noise variance $\sigma^2_{n,1} = 1.2\text{cm}^2$. Model M_1 had a zero-mean prior as in the previous experiment. Model M_2 was the simpler model with length scale $\lambda_2 = 3\text{cm}$, signal variance $\sigma^2_{f,2} = 0.25\text{cm}^2$, and noise variance $\sigma^2_{n,2} = 0.75\text{cm}^2$. Model M_2 had a prior non-zero mean function given by $m(\mathbf{x}) = A\sin(\omega\mathbf{x})$ where $A = 1\text{cm}$ and $\omega = 0.5\text{cm}^{-1}$.

Psychometric function The computation of the theoretical choice behavior is based on the evidence ratio $\mathcal{BF}$ between the two models as the decision variable, which in our case of equal prior probabilities transforms to the difference in log-space given by Equation 5. If subjects were to follow a simple strategy of *probability matching*, where their choice probability for model M_1 corresponds to the posterior belief probability $P(M_1|\mathbf{y})$, then the choice probability is given by fitting the psychometric function in Equation 3 with $\alpha = 1$. An optimal noise-free decision-maker would be modeled by $\alpha \to \infty$. However, for real-world decision-makers we expect finite α values. For $\alpha > 1$ the model predicts a strategy that is more deterministic than probability matching. Strategies with $\alpha < 1$ have even higher entropy than probability matching. The parameter α controls the slope of the psychometric sigmoid as shown in Equation 3. For evaluating the experimental data in the main text, we fitted α to each individual participants' probe trial data by maximizing the log likelihood of their actual choices. The results for a typical subject are shown in Supplementary Figure 3—the α-values for all subjects are presented in Supplementary Table 1. For most subjects the value moves closer to one in the second session, that means that the choice behavior gets closer to probability matching and suggests that participants were still learning the statistics during the first session. The results for the spatial frequency control experiment are shown in 2.

Data Analysis: Length Scale Comparison In order to find the best fitting length scale of trajectories, we allowed a trajectory noise σ_{traj} in addition to the observation-noise σ_n for the different trajectory types that was simply added to the diagonal elements of the posterior predictive covariance matrix. The value of this trajectory noise was fitted, using maximum marginal likelihood of the posterior predictive GP across all subjects and conditions simultaneously. The best fitting trajectory noise values were $\sigma^2_{traj} = 0.06cm^2$ for participants' trajectories and the connect-all trajectories and $\sigma^2_{traj} = 0.0001cm^2$ for samples from the posterior GP and the straight line trajectories. The reason for adding this noise was to obtain a more robust fitting procedure. We found that the best-fitting length scales for participants' trajectories were generally matched by the generating length scales, but were often slightly shorter—compare Supplementary Figure 4. Since this bias was particularly strong in the drawing first group, it is most likely a consequence of insufficient learning when subjects' approximate the Gaussian process trajectories with varying length scales. Such variations could lead to the observed bias towards lower length scales in our maximum predictive likelihood fit—see Supplementary Figure 5.

Data Analysis: Linear Regression To assess, how well the theoretical choice probabilities match the participants' choice behavior observed in the experiment, we linearly regressed subjects' observed choice probability against the theoretical choice probability for each participant and each condition individually. Ideally, the fitted line would have a slope of 1 and an intercept of 0—as illustrated by the dashed black line in Supplementary Figure 6 for instance. The line fit was performed on the mean choice probability for each bin, using the previously fitted α values. The results of the regression analysis are shown in Supplementary Figure 7 where the error bars represent 95% confidence intervals. As shown in the figure,

the ideal slope and intercept lie within the confidence intervals for all subjects in all conditions except for the intercept for Subject 2 in the first session. The results of the linear regression analysis for the control experiment, where the simpler model M_2 implies trajectories with a higher spatial frequency, are shown in Supplementary Figure 9 with the individual participants' choice propabilities shown in Supplementary Figure 8.

Choice of GP kernel In our experiment we used a squared exponential covariance function (see Equation 1) to generate synthetic trajectories and to model and analyze the recorded data. Perhaps the most obvious alternative to the squared exponential kernel for modeling human sensorimotor processing is the neural network kernel (Williams, 1998)[1]. This kernel can be interpreted as the covariance function of a neural network with a single hidden layer and an infinite number of hidden neurons with the error function $\mathrm{erf}(z) = 2/\sqrt{(\pi)} \int_0^z e^{-t^2} dt$ as activation function. The neural network kernel has the following mathematical form:

$$k_{\mathrm{NN}}(\mathbf{x}_p, \mathbf{x}'_p) = \sigma^2_{f,\mathrm{NN}} \sin^{-1}\left(\frac{\mathbf{x}^T \Sigma_{\mathrm{NN}} \mathbf{x}'}{\sqrt{(1+\mathbf{x}^T \Sigma_{\mathrm{NN}} \mathbf{x})(1+\mathbf{x}'^T \Sigma_{\mathrm{NN}} \mathbf{x}')}}\right), \tag{6}$$

where $\sigma^2_{f,\mathrm{NN}}$ is the signal variance and $\Sigma_{\mathrm{NN}} = \frac{1}{\lambda^2_{\mathrm{NN}}}\mathrm{I}$ with I being the identity matrix. The vectors $\mathbf{x}, \mathbf{x}'$ on the right hand side are the vectors $\mathbf{x}_p, \mathbf{x}'_p$ augmented with an extra bias entry with unit value. λ_{NN} plays the role of a length scale parameter because it governs the inverse variance of the prior over the weights from the input to the hidden layer—a large λ_{NN} implies small weights that lead in general to slowly varying outputs whereas short length scales λ_{NN} imply larger weights that lead to more quickly varying outputs—see (Rasmussen, 2006)[2] chapter 4 for details. In order to compare the neural network kernel with the squared exponential kernel we fitted the parameters $\sigma^2_{f,\mathrm{NN}}$ and λ_{NN} to participants' drawn trajectories by maximizing the median marginal likelihood of the predictive GP for both conditions of the standard trials (M_1- and M_2-trials), where the median was taken over all participants. We found for M_1-trials: $\sigma^2_{f,\mathrm{NN}} = 250cm^2, \lambda_{\mathrm{NN}} = 11cm$ and for M_2-trials $\sigma^2_{f,\mathrm{NN}} = 2000cm^2, \lambda_{\mathrm{NN}} = 4cm$. With these best-fit values the complexity penalization term in Equation 2 is lower for the M_1 best-fit parameters than for the M_2 best-fit parameters, which means that M_1 under the neural network kernel is the simpler model compared to M_2 under the neural network kernel—so the complexity relation between the two models is the same as in the squared exponential case. The results of the comparison between the neural network and the squared exponential kernel are shown in Supplementary Figure 10. For all participants in both trial types the squared exponential kernel yields a higher marginal predictive likelihood and is thus the preferred model for explaining the data. The reason for this could simply be that the synthetic trajectories were generated from the squared exponential kernel and that participants were able to learn this.

Fixed observation noise In our analysis we assumed that the observation noise σ^2_n was learned by participants. This assumption is motivated by the fact that the observation noise stayed constant throughout the whole experiment, also across both generative models M_1 and M_2. To verify the assumption we computed the marginal likelihood of the predictive GP for participants trajectories over a whole range of values for σ^2_n and λ_i (the length-scale parameter has to be taken into account as well since it might interact with the observation noise parameter). The results in Supplementary Figure 11 show that the

[1] Williams, C. K. (1998). Computation with infinite neural networks. Neural Computation, 10(5), 1203-1216.

[2] Rasmussen, C. E. (2006). Gaussian processes for machine learning.

marginal predictive likelihood is sharply peaked in the neighborhood of the true value of the observation noise, but the length scales are slightly biased towards shorter length scales—which is consistent with the findings in Supplementary Figure 4 and can be explained by the sensitivity of the fitting procedure to mixed-lengthscale trajectories, see Supplementary Figure 5.

Figures

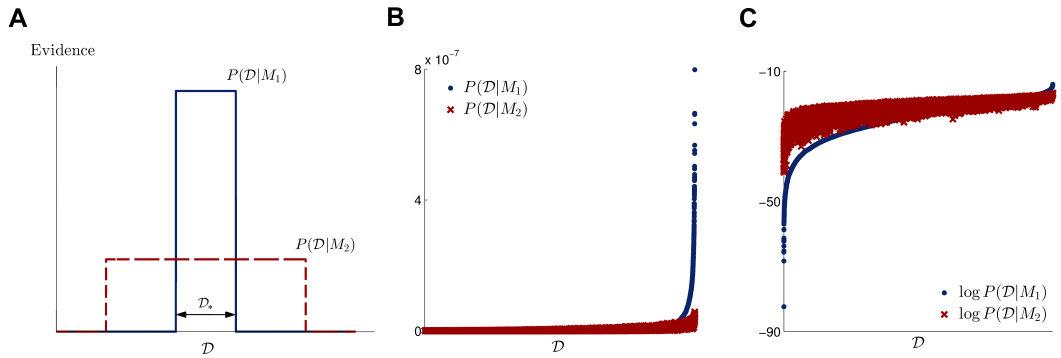

Figure S 1. Bayesian Occam's razor. **A** Schematic plot of the evidence $P(\mathcal{D}|M)$ for a simple model M_1 (blue, solid line) and a complex model M_2 (red, dashed line) for different data $\mathcal{D}$, for example different random trajectories. Because both models have to spread unit probability mass over all compatible observations, the simpler model M_1 has a higher evidence in the overlapping region $\mathcal{D}_*$. **B** Evidence plot for the Gaussian Process models used in the experiment. The plot shows the evidence for the two GP models with different length scales for $20,000$ simulated sets of random observations, where each set consisted of 9 uncorrelated normally distributed ($\mu = 0cm, \sigma^2 = 1.25cm^2$) data points. The observations have been ordered using the evidence for the simple model with long length scale (M_1, blue). **C** Same as in B on log scale. On the very right end of the data set axis, the simple model M_1 provides the better explanation for a small portion of the data sets, whereas on the left end of the data axis the complex model M_2 is the better predictor.

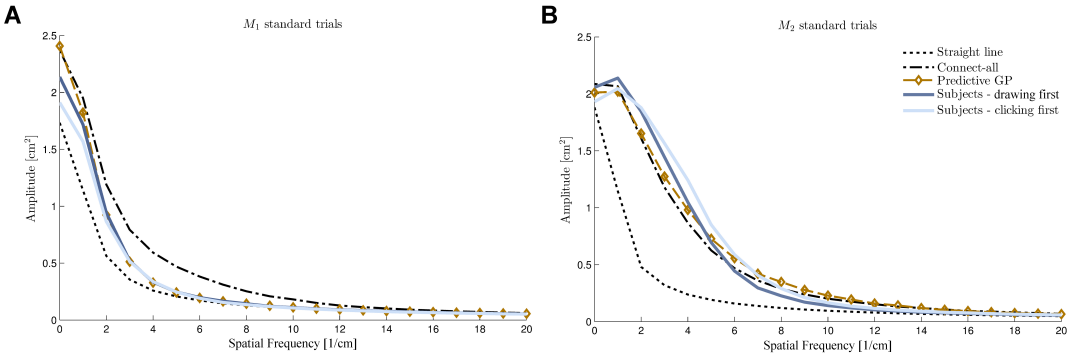

Figure S 2. Spectral analysis of trajectories. The amplitude spectra shown in both panels were obtained by performing a fast Fourier transform of participants' drawn trajectories—using full temporal resolution (1kHz) recordings. The figures show the average over participants' individual spectra. For comparison we also show spectra of a straight line from the first to the last observation, a trajectory connecting all observations with straight lines and samples from the predictive GP with the correct length scale. **A** Results for standard trials where M_1 was the generating model. Participants' spectra lie in-between the straight line and the connect-all trajectories. The spectrum of the predictive GP is matched quite well by participants' drawings. **B** Results for standard trials where M_2 was the generating model. All spectra except the straight line spectrum are quite similar. Only the spectrum of the predictive GP trajectories matches the spectra of participants' trajectories well in both standard trial conditions.

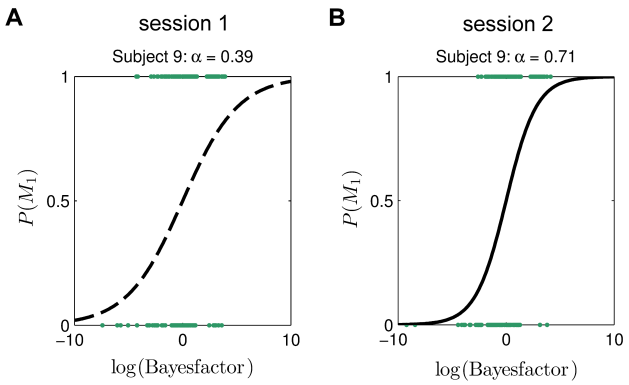

Figure S 3. Fitting the α-value of the psychometric function to the experimentally observed choices in the probe trials for subject 9 from the clicking first group. **A** First session of the experiment (clicking session in this case). Dots represent observed model choices, ordered by the log Bayes Factor of the data observed in the trial (see Equation 5). **B** Second session—drawing session in this case. The maximum likelihood α-value of 0.71 in the second session implies a "sharper" psychometric function.

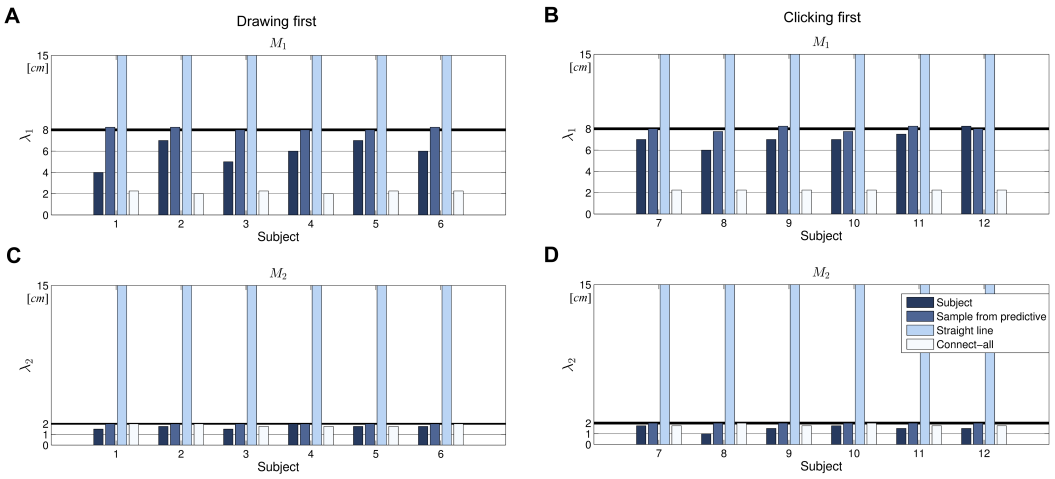

Figure S 4. Best-fitting GP model length scales for two conditions in two different sessions (**A**-**D**) and four different trajectory types (color map) including participants' drawn trajectories (dark blue), samples from the posterior predictive GP conditioned on the observations (medium dark blue), a straight line connecting the first and last observation (medium light blue) and a connect-all trajectory connecting all the observations with straight lines (light blue). The solid black line shows the length scales of the true generating GP models respectively and are well fitted by the inferred length scale of the posterior sample.

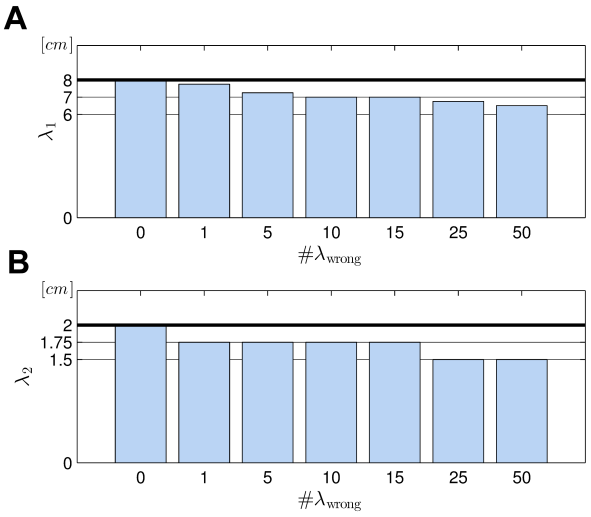

Figure S 5. Maximum likelihood inference of length scale when artificially generating 200 trajectory samples with precisely the same length scale, and when adding $2 \times \#\lambda_{\text{wrong}}$ samples of trajectories with $\pm 25\%$ shorter and longer length scales. Adding an increasing amount of trajectories with wrong length scale leads to an estimation bias towards lower length scales. In **A** the true length scale of the 200 trajectories was $8cm$, in **B** the true length scale was $2cm$. The reason for this bias is that the more flexible model with shorter length scale is a better explanation for a mix of trajectories.

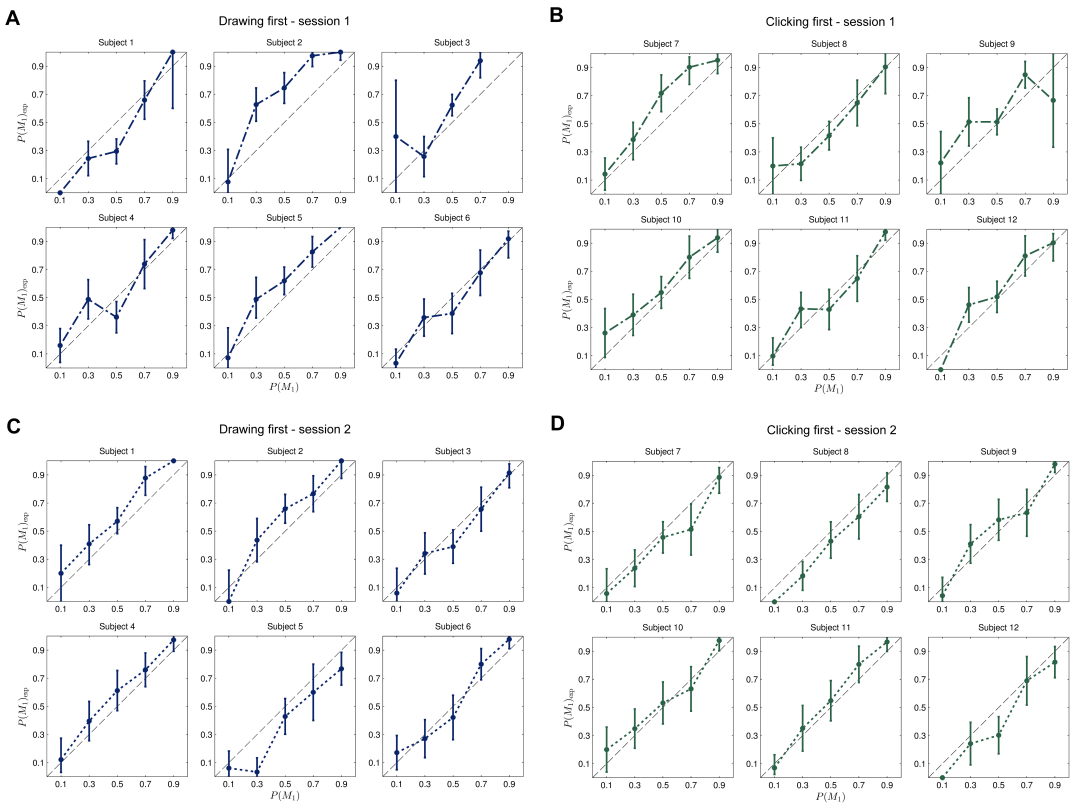

Figure S 6. Individual choice probabilities. **A**,**C** Results for the group that performed the drawing session first. The dash-dotted or dotted lines show the median over all trials along with 95% confidence intervals. The dashed black line illustrates the ideal case, where theoretical choice probabilities and observed behavior match exactly. **B**,**D** Results for the group that performed the clicking session first.

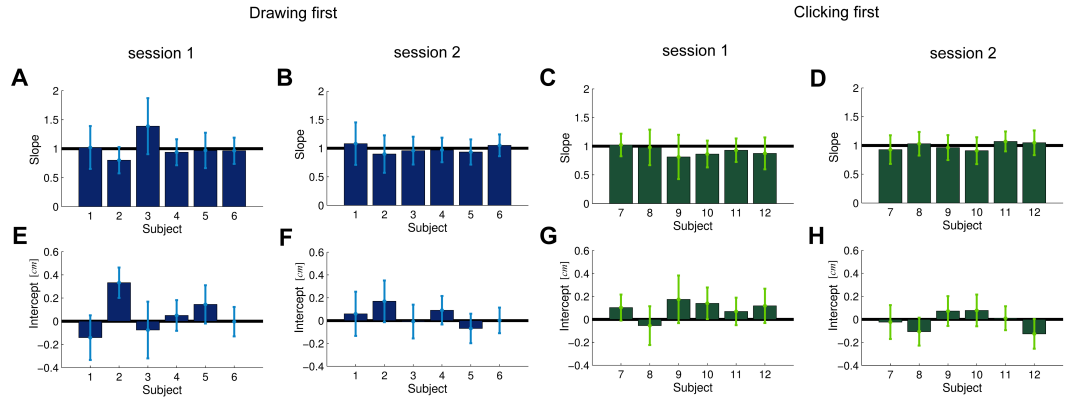

Figure S 7. Fitted slopes and intercepts after performing linear regression on the individual subjects' choice patterns. The error bars show 95% confidence intervals, the solid black lines show the ideal values. Panels **A**-**D** show the fitted slopes for both groups in both sessions. Panels **E**-**H** show the fitted intercepts correspondingly. The only case, where the ideal value does not lie within the confidence intervals is Subject 2 of the Drawing first group in the first session—see Panel **E**.

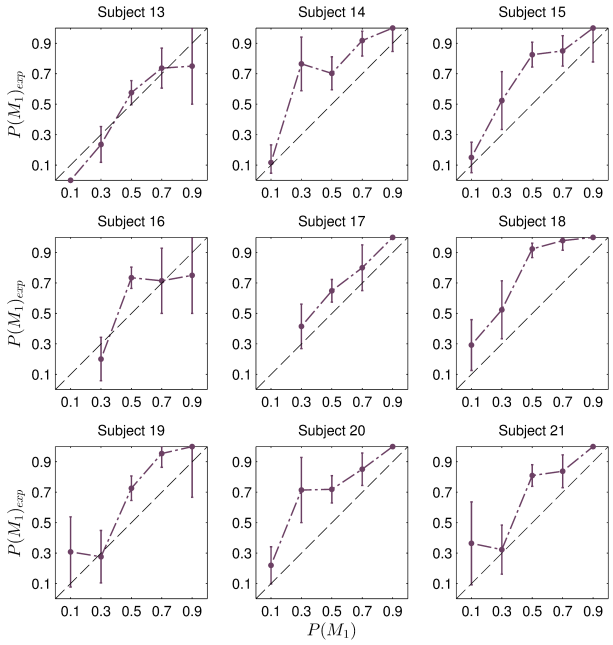

Figure S 8. Individual choice probabilities in the control experiment. The dash-dotted lines show the median over all trials along with 95% confidence intervals. The dashed black line illustrates the ideal case, where theoretical choice probabilities and observed behavior match exactly. Note that the number of data points in each bin depends on the fitted values for α, which for very low values of α (e.g. subjects 16 and 17) can result in empty bins for very low and very high probability predictions.

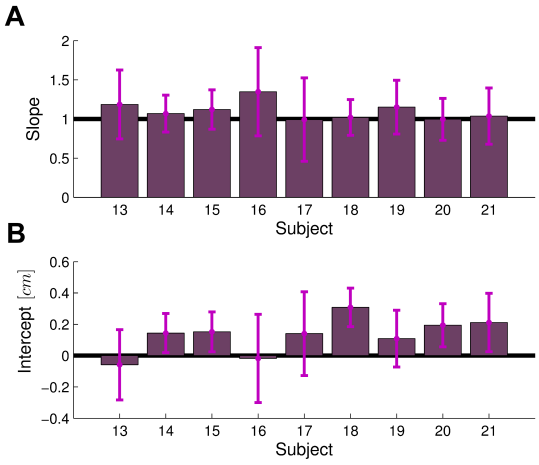

Figure S 9. Fitted slopes and intercepts after performing linear regression on the individual subjects' choice patterns in the control experiment. The error bars show 95% confidence intervals, the solid black lines show the ideal values.

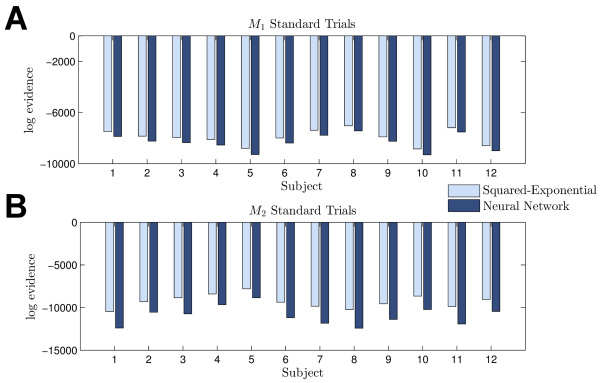

Figure S 10. Both panels show the log marginal likelihood of the predictive GP of drawn trajectories for each participant and for two different kernels: a squared exponential kernel and a neural network kernel. For the squared exponential kernel the parameters were set to the values of the generating models M_1 and M_2. The parameters of the neural network kernel were determined by maximizing the marginal predictive likelihood of participants' trajectories. **A** Results for M_1 standard trials. **B** Results for M_2 standard trials. The squared exponential kernel provides the better explanantion for the data of all participants in all conditions.

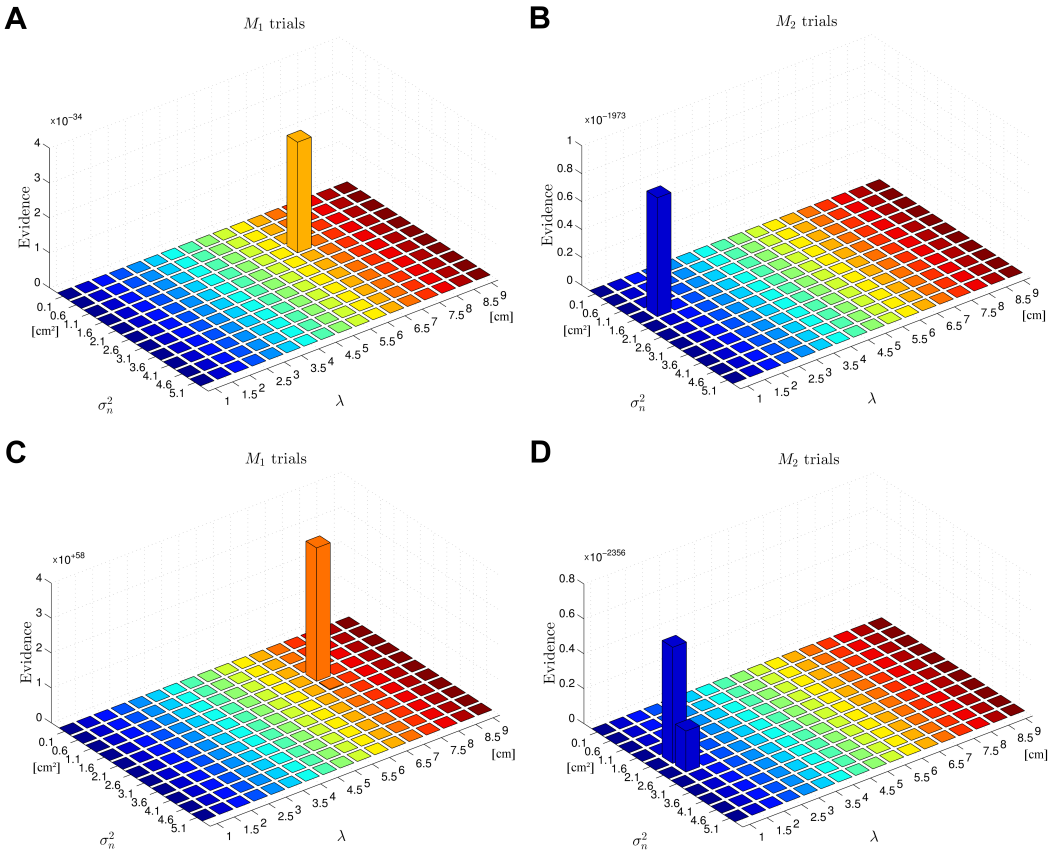

Figure S 11. Marginal likelihood of the predictive GP of participants' trajectories for a grid of parameter values for the squared exponential kernel (σ_n^2 and λ_i). Each panel shows the median over all participants of the same group (drawing first versus clicking first) in the same condition (M_1 standard trials or M_2 standard trials). Note that some of the results show very small numbers close to zero, however, exceeding the numerical precision during computation of the results has been prevented. **A**, **B** Results for the drawing first group. **C**, **D** Results for the clicking first group. The best fitting observation noise parameters are peaked in the neighborhood of the true value $\sigma_n = 1.5625\ cm^2$. The best fitting length scales are biased towards shorter length scales which is due to the sensitivity of the predictive likelihood to mixed length scale trajectories (see Supplementary Figure 5).

Tables

		α session 1	α session 2
Drawing first	Subject 1	0.39	0.39
	Subject 2	0.53	0.43
	Subject 3	0.27	0.69
	Subject 4	0.61	0.79
	Subject 5	0.43	0.63
	Subject 6	0.71	1.07
Clicking first	Subject 7	0.79	0.53
	Subject 8	0.43	0.73
	Subject 9	0.39	0.71
	Subject 10	0.61	0.75
	Subject 11	0.73	1.19
	Subject 12	0.53	0.69

Table S 1. Fitted α values in the experiment. Subjects 1 to 6 belong to the drawing first group that first performed the drawing session—subjects 7 to 12 belong to the clicking first group that started the experiment with the clicking session.

		α control session
Drawing	Subject 13	0.29
	Subject 14	0.59
	Subject 15	0.55
	Subject 16	0.23
	Subject 17	0.21
	Subject 18	0.35
	Subject 19	0.33
	Subject 20	0.43
	Subject 21	0.33

Table S 2. Fitted α values in the control experiment in which the simpler model had a higher spatial frequency.

6 Structure Learning in Bayesian Sensorimotor Integration

For a color-version of the plots in this chapter please see the digital version of this thesis or the original publication [Genewein et al., 2015a].

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Structure Learning in Bayesian Sensorimotor Integration

Tim Genewein[1,2,3]*, **Eduard Hez**[1,2,4], **Zeynab Razzaghpanah**[1,2,4], **Daniel A. Braun**[1,2]

1 Max Planck Institute for Biological Cybernetics, Tübingen, Germany, **2** Max Planck Institute for Intelligent Systems, Tübingen, Germany, **3** Graduate Training Centre of Neuroscience, Tübingen, Germany, **4** University of Tübingen, Tübingen, Germany

* tim.genewein@tuebingen.mpg.de

OPEN ACCESS

Citation: Genewein T, Hez E, Razzaghpanah Z, Braun DA (2015) Structure Learning in Bayesian Sensorimotor Integration. PLoS Comput Biol 11(8): e1004369. doi:10.1371/journal.pcbi.1004369

Editor: Max Berniker, University of Illinois at Chicago, UNITED STATES

Received: November 1, 2014

Accepted: June 1, 2015

Published: August 25, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files, in particular S1 Dataset.

Funding: This study was supported by the DFG, Emmy Noether grant BR4164/1-1. (http://dfg.de). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Previous studies have shown that sensorimotor processing can often be described by Bayesian learning, in particular the integration of prior and feedback information depending on its degree of reliability. Here we test the hypothesis that the integration process itself can be tuned to the statistical structure of the environment. We exposed human participants to a reaching task in a three-dimensional virtual reality environment where we could displace the visual feedback of their hand position in a two dimensional plane. When introducing statistical structure between the two dimensions of the displacement, we found that over the course of several days participants adapted their feedback integration process in order to exploit this structure for performance improvement. In control experiments we found that this adaptation process critically depended on performance feedback and could not be induced by verbal instructions. Our results suggest that structural learning is an important meta-learning component of Bayesian sensorimotor integration.

Author Summary

The human sensorimotor system has to process highly structured information that is affected by uncertainty and variability at all levels. Previously, it has been shown that sensorimotor processing is very efficient at extracting structure even in variable environments and it has also been shown how sensorimotor integration takes into account uncertainty when processing novel information. In particular, the latter integration process has been shown to be consistent with Bayesian theory. Here we show how the two processes of structure learning and sensorimotor integration work together in a single experiment. We find that when human participants learn a novel motor skill they not only successfully extract structural knowledge from variable data, but they also exploit this structural knowledge for near-optimal sensorimotor integration.

Introduction

The sensorimotor system continuously integrates incoming information with previous experience across different modalities. Previous studies have shown that such integration processes in environments with uncertainty are consistent with Bayesian learning [1–6], where previous experience—the prior—and sensory evidence are weighted according to their reliability [7–12]. Several sensory illusions could be modeled and explained by the Bayesian integration of prior information with sensory feedback [13–18]. The same mathematical formalism also adequately describes integration of information from different sensory modalities, for example visual and haptic information [19, 20].

When integrating sensory information, it is important to know the statistics of the environment. Previous studies have mostly investigated Gaussian statistics factorizing into one-dimensional random variables, but a number of other distributions have been tested as well [8, 11, 19–25]. Here we are interested in the effect of the structure of the distribution given by the dependencies between multiple hidden variables that can be learnt as higher order invariants. Structural learning has previously been investigated in sensorimotor learning tasks with randomly changing task parameters [26–37]. In these previous studies it has been suggested that the sensorimotor system is faced with two concurrent learning problems in such randomly changing tasks, that is adapting to the current environmental parameters and extracting structural knowledge that remains invariant over many variations of environmental parameters.

In the current study we investigate the role of structural learning in a Bayesian sensorimotor integration task with a two-dimensional hidden variable that determined the two-dimensional displacement of visual feedback of the hand position. As in previous tasks, the value of the hidden variable has to be inferred during the integration process in each trial by combining sensory feedback with previous experience. Determining this value can be regarded as an example of parameter adaptation. However, since the hidden variable has two dimensions, we can also introduce structural dependencies between the two dimensions that remain invariant across trials. The question in the current study is whether and how such structural invariants of hidden variables influence the sensorimotor integration process of sensory feedback with prior experience.

Results

Trial setup

Participants performed a reaching task in a 3D virtual reality setup in which their virtual hand position was represented by a small spherical cursor. The aim of the task was to steer the cursor into a target sphere that was always at the same position in front of them. Similarly, the starting position was fixed throughout the experiment. To initiate a trial, participants had to move the cursor into the starting position. After a beep, the cursor disappeared and participants started their movement towards the target without visual feedback of their virtual hand position. Each trial, a two-dimensional translational shift was randomly drawn from a Gaussian, as depicted in Fig 1 and applied to the virtual hand position such that it was shifted with respect to the actual hand position. This shift was constant throughout the trial, but changed from trial to trial. Importantly, the Gaussian distribution over the shift remained constant over the course of the whole experiment.

Halfway through participants' movement towards the target, feedback of the virtual hand position was briefly displayed for 150*ms*. This feedback was the only information participants could use to correct their shifted movement trajectory towards the target. There were four different feedback conditions that were chosen randomly in each trial with the following

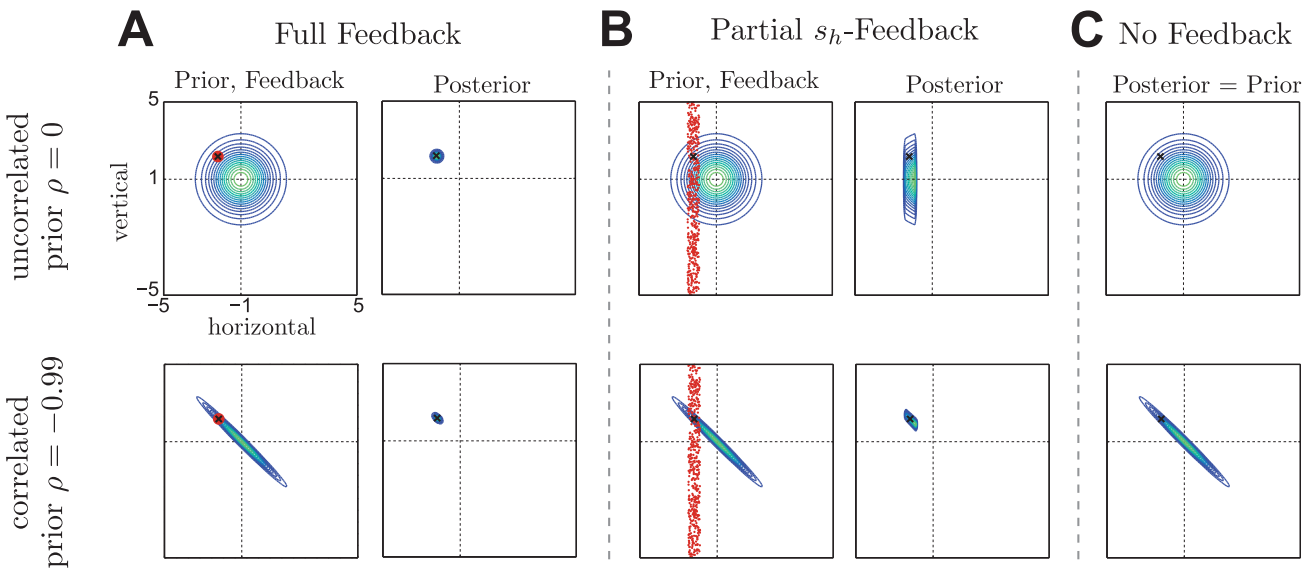

Fig 1. Bayesian integration for two-dimensional Gaussian priors under different feedback conditions A–C. The uncorrelated case is shown in the top row and the correlated case is shown in the bottom row. Prior and posterior are represented through iso-probability contours, the visual feedback is depicted in red and the true shift is marked as a black X. The black dotted lines indicate the prior mean. **A** Due to the very reliable feedback in the *full feedback* condition, the posterior is peaked very sharply—regardless of the correlation in the prior. **B** The *partial* s_h-feedback is reliable in the s_h dimension but provides no information about the shift in the s_v dimension. This leads to an important difference in the posterior between the correlated and uncorrelated group: knowing the correlation structure reduces uncertainty about the s_v dimension of the shift, leading to a more concentrated posterior. **C** In *no-feedback* trials, participants can only rely on their prior experience. This feedback condition allows to test for the prior beliefs directly.

doi:10.1371/journal.pcbi.1004369.g001

proportions: full feedback (1/2 of trials), partial s_h-feedback (1/6 of trials), partial s_v-feedback (1/6 of trials), and no-feedback (1/6 of trials). In the *full feedback* condition (Fig 1A) feedback was given by a small spherical cursor, which gave participants very precise information about the shift and allowed them to hit the target accurately. In the *partial* s_h*-feedback* condition (Fig 1B) feedback was given by an elongated Gaussian cloud of points with width 0.6*cm* in the horizontal direction and full workspace width in the vertical direction. The cloud consisted of 50 small circles (radius 0.1*cm*) and their exact position was re-sampled several times during the display of the feedback, thus creating the visual effect of a flickering vertical bar, very similar to the depiction in Fig 1B. This sensory feedback gave relatively precise information about the horizontal shift s_h, but no information about the vertical shift s_v. In the *partial* s_v*-feedback* condition this was reversed. The Gaussian cloud of points was elongated over the full workspace width in the horizontal direction and had a narrow vertical expansion of 0.6*cm*. Therefore, the sensory feedback provided relatively precise information about the vertical shift s_v, but no information about the horizontal shift s_h. In the *no-feedback* condition (Fig 1C) no feedback was provided, such that participants could only rely on their prior experience of the statistics in these trials. The critical feedback conditions are the partial feedback conditions. If the correlation structure between the two shifts is unknown, the feedback only provides information for one dimension. If, however, the correlation structure has been learnt over many trials, the partial feedback provides information for both dimensions of the shift.

Sessions and groups

Participants were recorded in this experiment over four different days—compare Fig 2A. The first six participants were assigned to the correlated group and the second six participants were assigned to the uncorrelated group. The correlated group was trained on shifts drawn from a

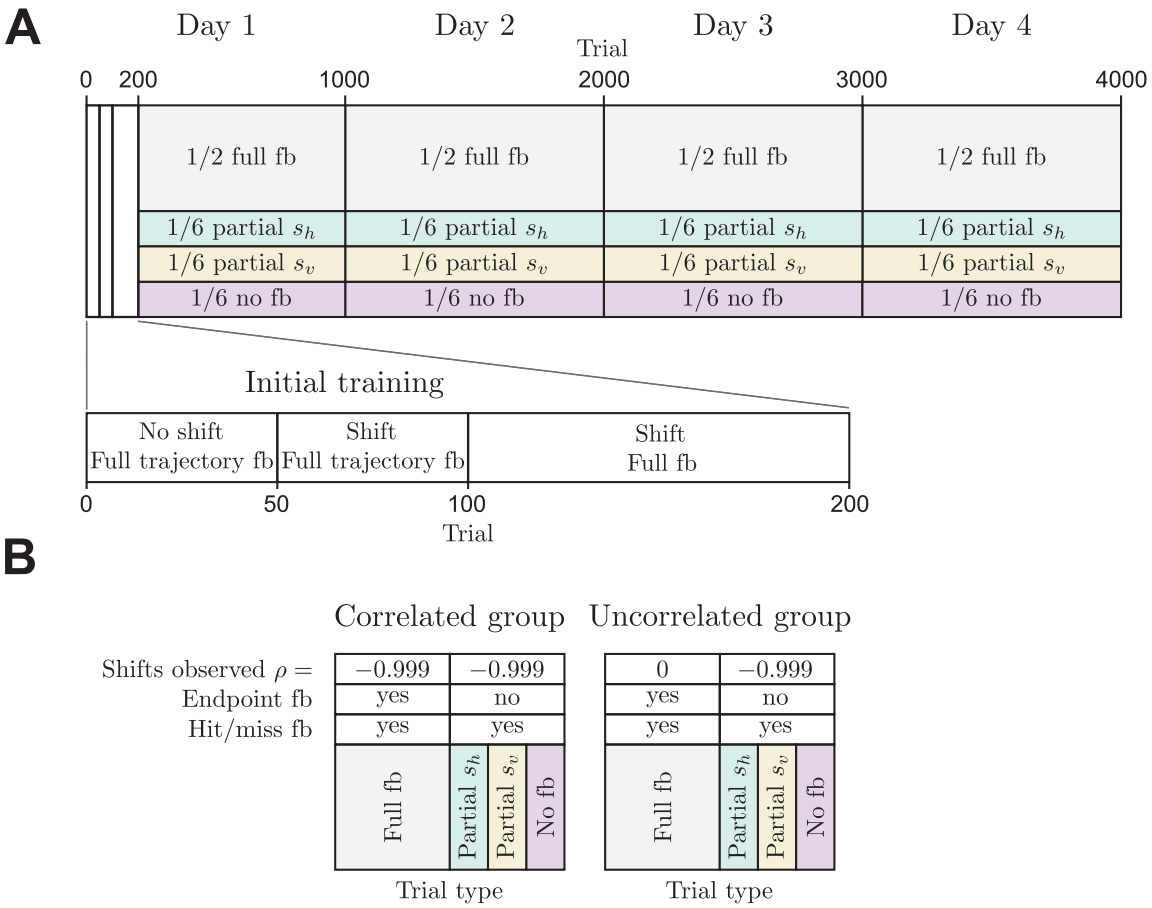

Fig 2. Schematic of the experimental design. A Participants were recorded over four sessions spread across four days. The first session included an additional training phase (first 200 trials) to allow participants to get used to the experimental setup. After this initial training phase, the different trial types were presented randomly according to the specified proportions. **B** Experimental conditions for the correlated and uncorrelated group.

doi:10.1371/journal.pcbi.1004369.g002

correlated Gaussian ($\rho = -0.999$), while the uncorrelated group was trained on shifts drawn from an uncorrelated Gaussian ($\rho = 0$)—see Fig 2B. This training was given in full feedback trials, that not only provided most information during the movement compared to other feedback conditions, but also gave terminal feedback of the cursor position at the end of the trial. In contrast, all other feedback conditions served as test trials without terminal visuomotor feedback. However, to keep participants motivated they were informed in all feedback conditions whether they had hit the target or not through auditory feedback. To test whether participants of the correlated group were able to extract the structural invariant of the correlation in the hidden variable during training, in test trials (that is partial- and no-feedback trials) we exposed both groups to correlated shifts ($\rho = -0.999$) under partial- or no-feedback. In particular, we would expect the correlated group to differ from the uncorrelated group in the processing of the uninformative feedback dimension in partial feedback trials, as knowing the correlation structure allows transferring feedback information from the informative to the uninformative feedback dimension. In principle, the uncorrelated group could have also learnt the correlation in test trials by exploiting the hit-or-miss feedback provided in these trials. However, we would expect such reinforcement-learning to be much slower since the

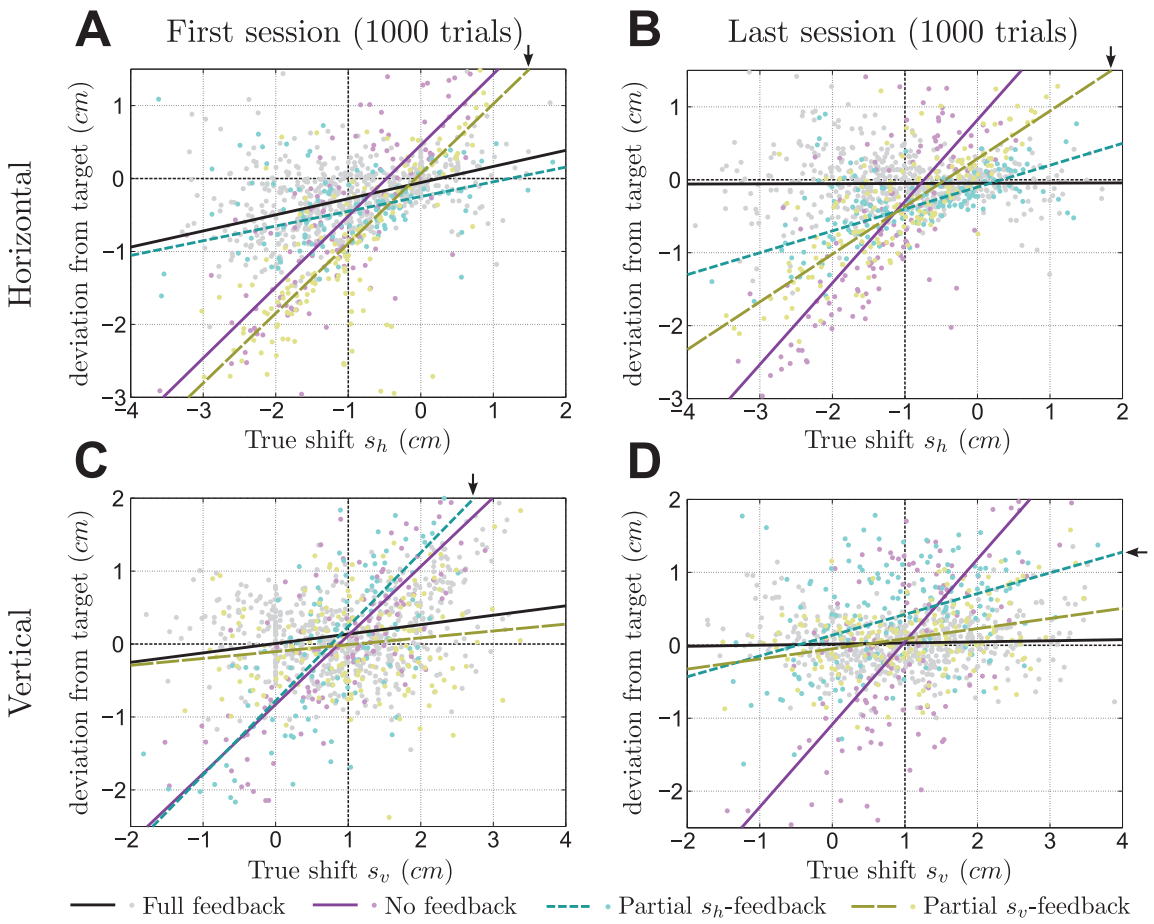

Fig 3. Data of a typical participant (no. 6, correlated group). The plots show the deviation of the final virtual hand position from the target as a function of the true shift. Dots represent individual trials and the lines show robust fits through the corresponding dots. Different colors indicate different feedback conditions. The crossing of the black dashed lines indicates the optimal pivot-point. Top row **A,B**: horizontal deviation as a function of the horizontal shift s_h. Bottom row **C,D**: vertical deviation as a function of the vertical shift s_v. The left column A,C shows results recorded in the first session of the experiment, the right column B,D shows results from the last session. In early trials, the participant's reaction to partial feedback trials in the noninformative dimension is very similar to behavior in no-feedback trials. Importantly, across sessions there is a significant reduction in slope in the noninformative dimension of the partial feedback trials, indicating learning of the correlation structure (compare changes in lines highlighted with arrows, that is the yellow dashed lines in panels A and B and cyan dashed lines in panels C and D).

doi:10.1371/journal.pcbi.1004369.g003

information of the reward feedback signal is poorer than the two-dimensional error signal observed by participants of the correlated group during training trials.

Typical participant

Fig 3 shows results of a typical participant from the correlated group performing in the task. The learning of the correlation structure can be seen when comparing the two panels on the left column showing performance in the first session to the two panels on the right column showing the last session. The plots show the deviation of the participant's shifted terminal hand position from the target as a function of the true shift. The top row shows the horizontal deviation depending on the horizontal shift component s_h, the bottom row shows the vertical deviation depending on the vertical shift s_v. The different colors indicate the different feedback conditions. Each dot shows a single trial and the lines were robustly fitted through the corresponding dots (using MATLAB's robustfit function with the default tuning constant of 4.685—

the function implements iteratively re-weighted least squares with a bi-square weighting function). Flat lines indicate low performance error due to reliable feedback information. Lines with a unit slope indicate performance of a learner that exclusively relies on prior information. Lines with slopes in between these two extremes indicate a (Bayesian) weighting of feedback and prior information [8]. The mathematical predictions of the perfect Bayesian actor, that knows the statistics of the task and exactly compensates the mean of the posterior belief, can be seen for our task in Eq (2) of the methods. For this participant, in the full feedback condition (black lines) the slope was close to zero in both dimensions, as participants could see their virtual hand position relatively clearly and could therefore compensate the error regardless of the magnitude of the true shift. In contrast, in the no-feedback condition (purple lines) participants had to completely rely on their learnt prior and would ideally compensate the most probable shift, that is the mean shift. Accordingly, in no-feedback trials the participant's deviation from the target as a function of the true shift is roughly described by a line with unit slope and intercept determined by the mean of the true shift—compare Fig 3.

Analysis of partial feedback trials in the last session

If the participant had not learnt the correlation structure, performance measured by the slope in the partial feedback condition should be similar to the slope in the no-feedback condition with respect to the uninformative feedback dimension. This is exactly what we see in the early session depicted in Fig 3A and 3C showing that the mean of the distribution over shifts has been roughly learnt, but the correlation between the two dimensions of the shift has not been learnt. In contrast, we found a significant reduction in the slope of the uninformative feedback dimension after extensive training in the last session—compare yellow dashed lines in Fig 3A and 3B and cyan dashed lines in Fig 3C and 3D. This indicates that the correlation structure has been learnt partially over the course of 4,000 trials. If the correlation had been fully learnt we would expect all partial feedback lines in panels B,D to be very similar to the full feedback condition, that is having a slope close to zero.

The results for all participants of the correlated and uncorrelated group are shown in Fig 4, where in the last session of the experiment the correlated group shows a significant difference

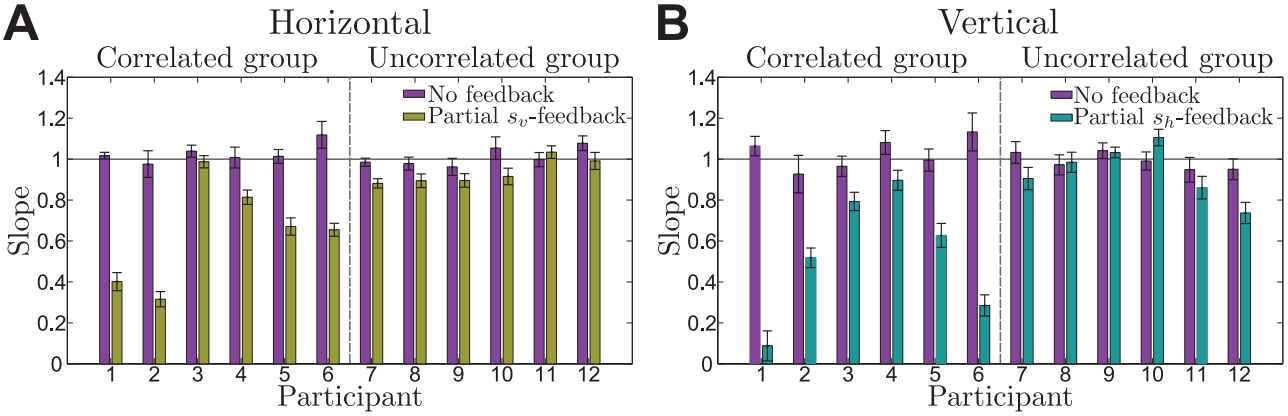

Fig 4. Performance of all participants in the last session of the experiment. The performance in the horizontal dimension is shown in panel **A**, performance in the vertical dimension in panel **B**. Performance is measured by slopes as in Fig 3 comparing no-feedback trials (purple) and partial feedback trials (yellow and cyan). Learning of the correlation structure is evident whenever the slope in the uninformative dimension of the partial feedback trials is significantly smaller than the slope in no-feedback trials (see also Fig 3). The perfect Bayesian response for no-feedback trials is characterized by a slope of one indicated by the thin black line, the Bayes-optimal slope for partial feedback trials would be zero—assuming that the Bayesian actor perfectly knows the statistics of the task. In both panels, error bars show standard errors of the robust fit.

doi:10.1371/journal.pcbi.1004369.g004

in slope in the uninformative feedback dimension of partial feedback trials compared to their performance in no-feedback trials ($p = 0.030$ signed-rank test for the horizontal slope and $p = 0.030$ signed-rank test for the vertical slope). The mean slope for the correlated group across participants in the last session was 0.640 ± 0.100 for the horizontal slope and 0.534 ± 0.125 for the vertical slope (mean ± standard error of the mean), which corresponds to the proportion of the perturbation that participants were not able to compensate. This suggests that information from the reliable dimension in partial feedback trials was successfully applied to the dimension providing no feedback which is only possible if the correlation structure has been learnt at least partially. In contrast, the uncorrelated group did not show a significant difference in slope in the uninformative feedback dimension of partial feedback trials compared to their performance in no-feedback trials ($p = 0.438$ signed-rank test for the horizontal slope and $p = 0.063$ signed-rank test for the vertical slope). The mean slope across participants of the uncorrelated group in the last session was 0.935 ± 0.025 for the horizontal slope and 0.937 ± 0.054 for the vertical slope (mean ± standard error of the mean). In principle, however, this group could have adapted their slope through reinforcement learning in partial and no-feedback trials, which might explain the close-to-significant p-value in the vertical dimension. More importantly, therefore, comparing the reduction in slope from the no-feedback to the partial-feedback trials between the correlated and uncorrelated group, we find a significant difference between both groups ($p = 0.041$ rank-sum test for the horizontal dimension and $p = 0.009$ rank-sum test for the vertical dimension, data from last session).

Analysis of no-feedback trials in the last session

In no-feedback trials, participants can only rely on their experience from previous trials, which allows to directly query their prior belief about the expected shift by investigating participants final hand positions. Fig 5A and 5B shows the mean of participants' final hand positions in no-feedback trials over the last session. To perfectly compensate the mean of the experimentally induced distribution over the shift, participants should on average reach to $[1,-1]cm$ in order to maximize their hitting probability. This holds for both the correlated and the uncorrelated group, since the mode of the distribution over the shift is unaffected by the correlation. As shown in Fig 5A and 5B we found that most participants learnt the mean shift with no significant difference between the correlated and the uncorrelated group ($p = 0.590$ for the horizontal dimension and $p = 0.065$ for the vertical dimension, rank-sum test).

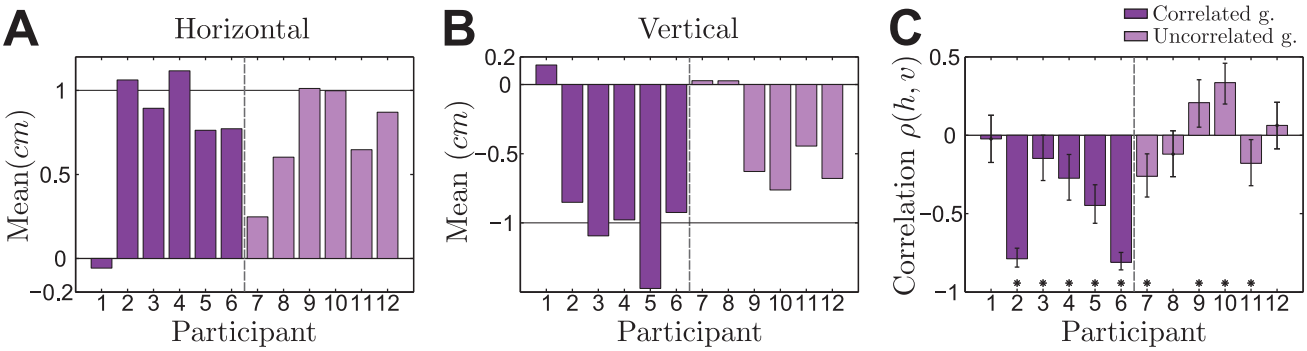

Fig 5. Means and correlation of the final hand position in no-feedback trials of the last session. A The mean of the final horizontal hand position for an ideal actor should be +1*cm* to fully compensate the mean shift. **B** The mean of the final vertical hand position for an ideal actor should be −1*cm* to fully compensate the mean shift. **C** Correlation coefficient between the vertical and horizontal components of the final hand position. Error bars indicate 95% confidence intervals—bars marked with a star show significant correlations (at a 5% level).

doi:10.1371/journal.pcbi.1004369.g005

Moreover, we computed the correlation coefficient between the vertical and horizontal components of participants' final hand position in no-feedback trials of the last session of the experiment. As shown in Fig 5C, we found a significant difference between both groups ($p = 0.041$, rank-sum test)—participants from the correlated group systematically showed a negative correlation in their final hand-position ($p = 0.030$, sign test), whereas participants from the uncorrelated group did not ($p > 0.990$, sign test). Importantly, a correlation in the two dimensions of the hand position cannot be explained by a perfect Bayesian actor model that exactly compensates the mean of the prior distribution over the shifts, even if some isotropic motor noise was added to this planned response. We consider two hypotheses that are not necessarily mutually exclusive. The first hypothesis is that the correlation could be a signature of a bounded rational actor that samples beliefs from its prior distribution over shifts and chooses its actions with regard to these samples. The second hypothesis is that the correlation simply reflects the correlation in the previous full feedback trial assuming a trial-by-trial adaptation process. We found evidence for both hypotheses. In particular, we found in accordance with the second hypothesis that participants' responses in no-feedback trials of both the correlated and uncorrelated group were significantly correlated with the shift of the previous full feedback trial (correlated group last session: correlation-strength -0.31 ± 0.26 horizontal and -0.34 ± 0.23 vertical (mean ± standard deviation)—uncorrelated group last session: correlation-strength -0.2 ± 0.15 horizontal and -0.23 ± 0.13 vertical). For the correlated group this correlation was significant for four out of six participants in the horizontal dimension and for five out of six participants in the vertical dimension—for the uncorrelated group we found a significant correlation for four out of six participants in both dimensions. If the correlation in the two dimensions of the hand position was entirely due to trial-by-trial adaptation, we would expect the correlation to be roughly stationary, as the xy-correlation in full feedback trials is already present in the earliest trials of the first session and changes only minimally across sessions. In contrast, we found that the correlated group started with a close-to-zero xy-correlation in no-feedback trials and showed learning-dependent improvement in the correlation over time (xy-correlation coefficient across participants in the first session -0.12 ± 0.17 versus the last session -0.42 ± 0.33, mean ± standard deviation), which would fit with the predictions of a bounded rational model of acting—compare Section: Model prediction.

Learning across sessions

To investigate behavior beyond the final session, we analyzed the dynamics of learning over the entire four sessions. We assess the evolution of participants' performance slopes in partial feedback trials and in no-feedback trials the evolution of participants' correlation between the two dimensions of their final hand position as well as the evolution of their mean responses. Fig 6 shows participants' evolution of performance slopes across the four sessions in partial feedback trials. The figure shows individual participants as thin colored lines and the median over participants as a thick black line. The results show a clear difference between the correlated and the uncorrelated group—the correlated group shows a steady decrease in slopes across sessions, whereas the uncorrelated group shows no such trend. This suggests that the correlated group gradually learnt to harness the informative feedback dimension to facilitate the sensorimotor integration process in the uninformative feedback dimension.
In contrast to the gradual learning of the correlation structure, we found no difference in learning of the mean of the distribution over shifts between the two groups—compare Fig 7. The results suggest that large parts of learning the mean shift already happened before the occurrence of the first no-feedback trials that we used to assess learning of the mean in the figure.

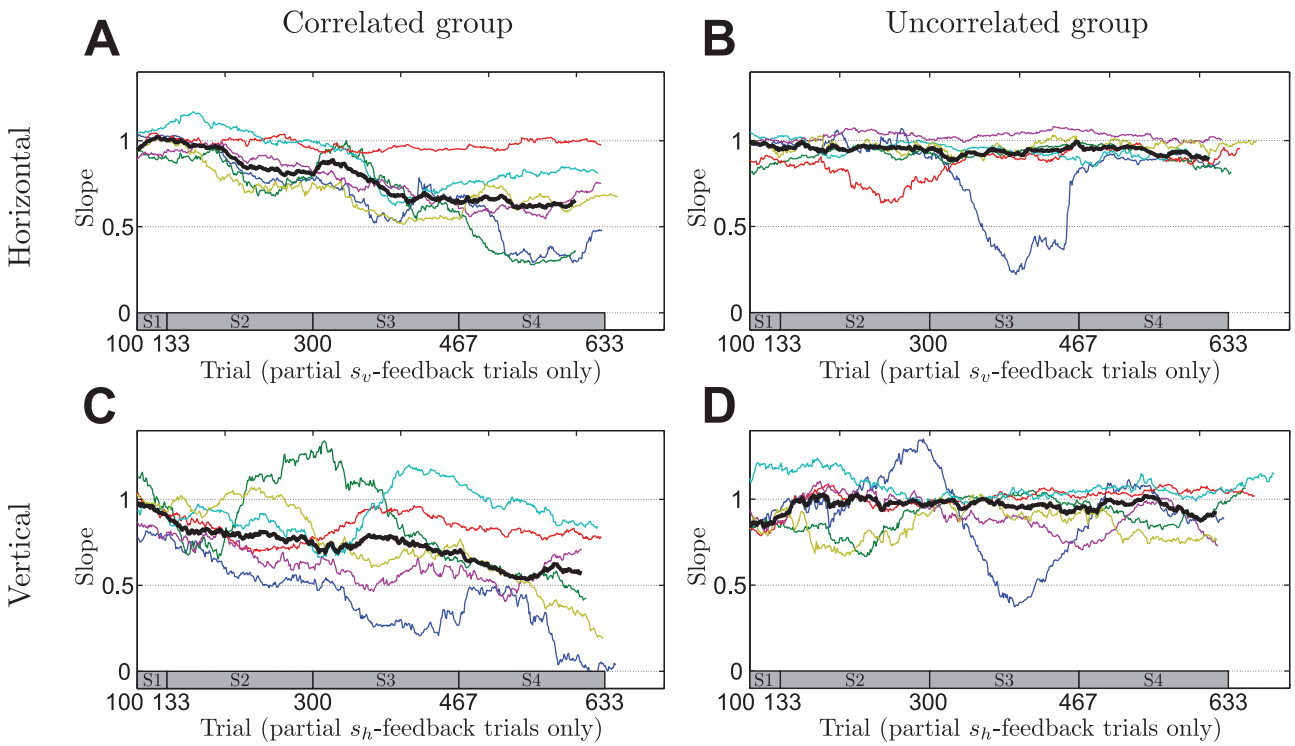

Fig 6. Changes in slope in partial feedback trials. The slope is a performance measure determined as in Fig 3 but using a sliding window of 100 trials. **A** Evolution of the horizontal slopes in partial s_v feedback trials of the correlated group where horizontal information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. **B** Same as A but showing data of the uncorrelated group. **C** Evolution of the vertical slopes in s_h feedback trials of the correlated group where vertical information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. **D** Same as C but showing data of the uncorrelated group. For the analysis only partial s_v- or partial s_h-feedback trials were taken out from the pooled data across all sessions. Thin colored lines indicate individual participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The thick black line shows the median over participants—taking only into account trials where data from all participants exists. The bar at the bottom of the figure indicates the corresponding session (on average).

doi:10.1371/journal.pcbi.1004369.g006

Finally, we investigated the evolution of participants' correlation between the two dimensions of their final hand position. In Fig 8A and 8B we show the evolution of the correlation coefficient between the horizontal and vertical component of participants' final hand position in no-feedback trials over the course of the whole experiment. Similar to the results in Fig 5C, we found that the correlated group shows an increasingly negative correlation across sessions, whereas the uncorrelated group does not show such a trend.

Control experiment: reinforcement learning vs. supervised learning

In our experimental design the correlated group could have learnt the correlation structure from two sources: first, from the error signal in full feedback trials allowing for some kind of supervised learning, and second, from the binary auditory performance feedback in partial and no-feedback trials allowing for some kind of reinforcement learning. As the uncorrelated group experienced the same statistics and binary performance feedback in partial and no-feedback trials, we can already exclude the possibility that the correlation structure in partial and no-feedback trials is learnt from binary feedback alone. However, it is unclear whether the binary feedback signal was crucial for the correlated group in learning the correlation structure.

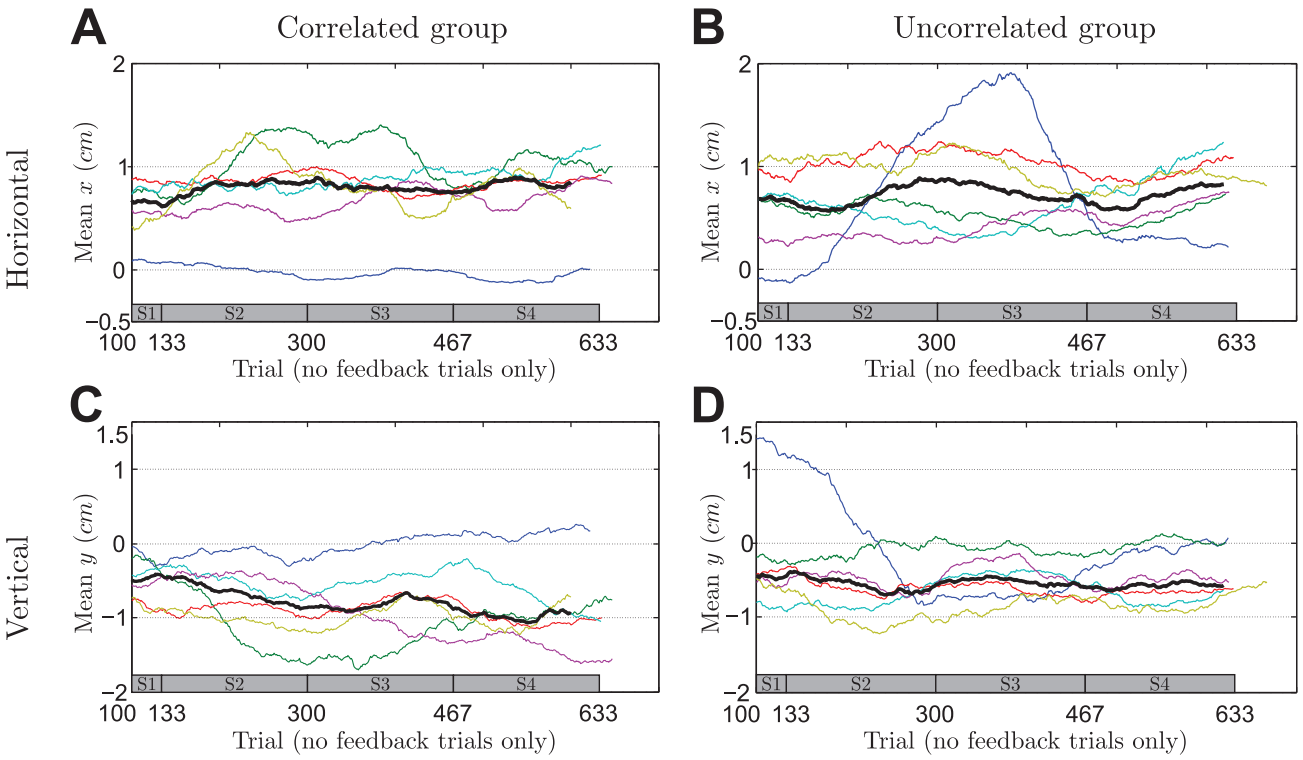

Fig 7. Learning of mean shift over all sessions revealed by performance in no-feedback trials averaged over a sliding window of 100 trials. A Learning of the mean in the horizontal dimension of the correlated group. **B** Same as A but showing data of the uncorrelated group. **C** Learning of the mean in the vertical dimension of the correlated group. **D** Same as C but showing data of the uncorrelated group. For the analysis only no-feedback trials were taken out from the pooled data across all sessions. Thin colored lines indicate individual participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The thick black line shows the median over participants—taking only into account trials where data from all participants exists. The bar at the bottom of the figure indicates the corresponding session (on average).

doi:10.1371/journal.pcbi.1004369.g007

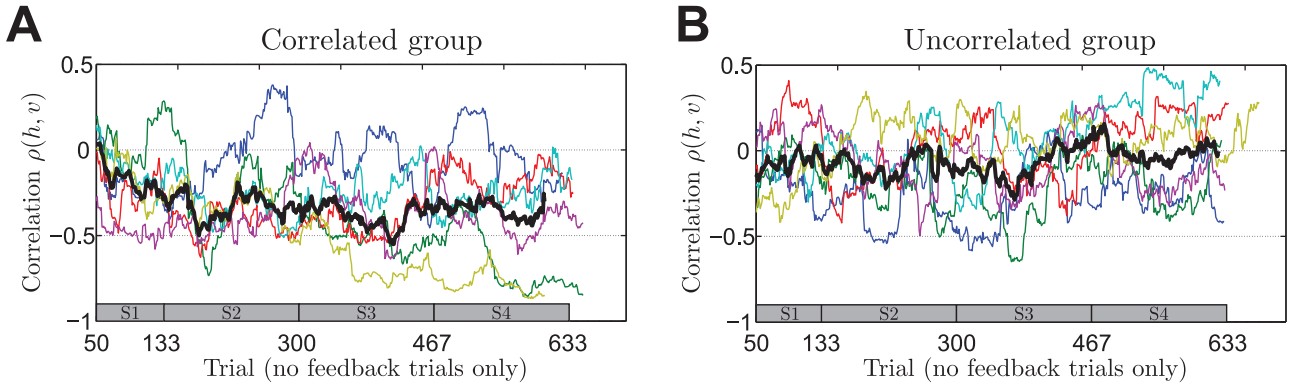

Fig 8. Adaptation of correlation between the vertical and horizontal terminal hand position measured in no-feedback trials. Correlation values are determined in a sliding window of 50 trials across all four sessions. **A** Correlated group. **B** Uncorrelated group. For the analysis only no-feedback trials were taken out from the pooled data across all sessions. Thin colored lines indicate individual participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The thick black line shows the median over participants—taking only into account trials where data from all participants exists. The bar at the bottom of the figure indicates the corresponding session (on average).

doi:10.1371/journal.pcbi.1004369.g008

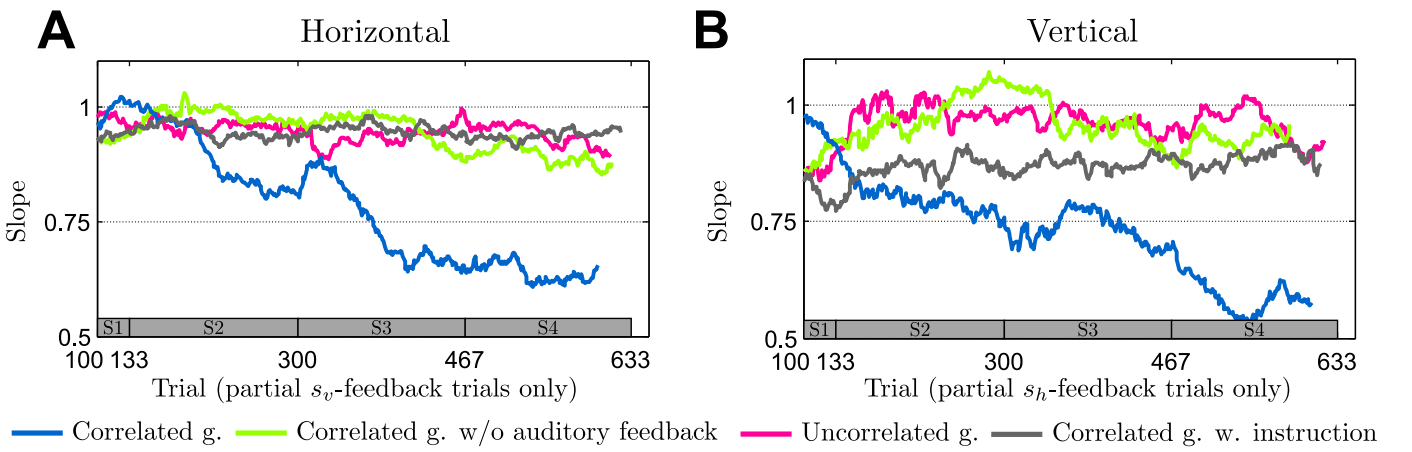

Fig 9. Changes in slope in partial feedback trials across groups. The slope is determined as in Fig 6. **A** Evolution of the horizontal slopes in partial s_v feedback trials where horizontal information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. **B** Evolution of the vertical slopes in s_h feedback trials where vertical information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. The correlated group shows a gradual and steady improvement across sessions whereas the other groups do not show such a trend. Different colored lines show the median over the different groups of participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The bar at the bottom of the figure indicates the corresponding session (on average).

doi:10.1371/journal.pcbi.1004369.g009

To control for this possible source of learning, we devised a control group that underwent the same experimental procedure as the correlated group with the important exception that this group did not receive any performance feedback in partial and no-feedback trials. We found that this group behaved similarly to the uncorrelated group in that they showed almost no reduction in slope in partial feedback trials ($p = 0.485$ horizontal and $p = 0.699$ vertical, ranksum test against uncorrelated group with data from the final session), in clear contrast to the correlated group that received binary performance feedback in partial and no-feedback trials ($p = 0.041$ horizontal and $p = 0.026$ vertical, ranksum test against correlated group with data from the final session). The same pattern is also visible in the evolution of slopes across sessions as shown in Fig 9 (evolution of slopes of individual participants is shown in Supplementary S1 Fig, evolution of means of individual participants is shown in Supplementary S2 Fig). This suggests that participants require both signals to learn, that is the immediate auditory feedback in partial- and no-feedback trials and the endpoint feedback reflecting the correlation structure in full-feedback trials.

Control experiment: cognitive strategies vs. motor learning

In our experimental design the optimal strategies in full and partial feedback conditions of the correlated group always required diagonal compensatory movements that were either directed left-up or right-down. This raises the question of whether participants could have learnt an explicit cognitive strategy instead of implicit sensorimotor integration. The hypothesis is that an explicit cognitive strategy can be verbally communicated and enable the participant more or less instantly to perform well. To control for this possibility, we devised a group of participants that was explicitly informed about the correlation structure, that is they were told that successful compensations would either be left-up or right-down. Crucially, if the correlated group simply learnt a cognitive strategy then the explicitly instructed group should be able to perform in their first session as well as the correlated group in their last session, assuming that the correlated group had figured out the cognitive strategy by the fourth session that the instructed group was given immediately. We found this not to be the case.

In the partial feedback trials, the correlated group performed significantly better at the end of the experiment than the instructed group in their first session (comparing the reduction of slope in the first session of the instructed group against the reduction of slope in the last session of the correlated group with a ranksum test: p = 0.026 horizontal dimension in partial s_v trials and p = 0.065 vertical dimension in partial s_h trials). The performance difference between the two groups is particularly obvious when comparing the evolution of the slope in partial feedback trials across sessions—compare Fig 9. The figure shows that the instructed group is not learning the correlation structure across sessions, as there is no statistical evidence for improvements of the slopes in partial feedback trials across sessions (comparing the reduction in slope with respect to the no-feedback slope between the first and the last session of the instructed group with a ranksum test: p = 0.937 horizontal and p = 0.937 vertical). The evolution of slopes of individual participants is shown in Supplementary S1 Fig and the evolution of means of individual participants is shown in Supplementary S2 Fig. There is, however, evidence that participants understood the instructions, as they showed a significant correlation between the horizontal and vertical dimension of their hand movements under partial feedback straight away: in the first session, the instructed group had a movement-correlation in horizontal and vertical partial feedback trials of −0.29 ± 0.15 and −0.27 ± 0.07 (mean ± standard deviation) respectively compared to the correlated group that showed no initial correlation of the horizontal and vertical dimension of their hand-movement in partial feedback trials (−0.06 ± 0.16 and −0.04 ± 0.17, mean ± standard deviation). This difference in correlation was significant (p = 0.026 horizontal and p = 0.026 vertical, ranksum test comparing the first session of the correlated group and the first session of the instructed group). The evolution of the xy-correlation in partial feedback trials across all sessions draws the same picture—compare Supplementary S3 Fig.

Surprisingly, the increased correlation in the hand movements in partial feedback trials of the instructed group did not produce a reduction in slope in these trials. In fact, the instructed group showed strongly increased slopes in the low-uncertainty dimension of the partial feedback trials in the first session of the experiment—compare Supplementary S3 Fig. In the low-uncertainty dimension of the partial feedback trials, an ideal actor should have a slope close to zero reflecting low uncertainty about the shift. The instructed group had an elevated slope of 0.725 ± 0.469 and 0.732 ± 0.455 (mean ± standard deviation) for horizontal and vertical partial feedback trials in the first session respectively, compared to the correlated group that had a slope of 0.342 ± 0.160 and 0.193 ± 0.154 (mean ± standard deviation) in their first session. This suggests that, while the instructions were clearly understood and followed, the explicit instructions actually impeded participants' ability to compensate the shifts in the low-uncertainty dimension of the partial feedback, particularly in the early sessions of the experiment. As performance in the low-uncertainty dimension does not require learning a statistical prior (and in fact in all the other groups there seems to be better performance and little performance improvement in the low uncertainty dimension—see Supplementary S3C and S3D Fig), this suggests that the deficient performance in the instructed group might be due to a shift in attentional focus, where subjects might pay more attention to following the instruction than to actual performance [38]. Further, the instructed group also shows impeded implicit learning—compare the evolution of the slope in partial feedback trials for the instructed group in Fig 9. While these results are not conclusive with respect to the origin of the deficient performance of the instructed group, they clearly demonstrate that explicit instructions did not instantly improve performance and therefore suggest that the correlated group were not following an explicit cognitive strategy.

Model predictions

As in the model described in [8], the ideal Bayesian actor optimally integrates prior knowledge about the shift with feedback information in each trial. For our experiment this mathematical prediction can be found in Eq (2) of the Methods. Importantly, this integration presumes that the prior is perfectly learnt to be consistent with the experimentally imposed prior. While this is the case in [8], in our study this is not the case, as can be seen for example in Fig 9, where the slopes in partial feedback trials never approach the Bayesian optimum of zero. This implies that the correlation in the prior is never fully learnt by participants. To model participants' behavior we therefore devised not only a Bayesian model of sensorimotor integration of prior and feedback, but also a Bayesian model of learning the prior and the corresponding correlation structure. In this model the actor has a belief about the prior over the shift s given by $p(s|\mu_0,\Sigma_0)$, where μ_0 and Σ_0 are hyper-parameters that the actor is learning over the course of many trials. If the initial belief over s is concentrated on uncorrelated shifts, as would be

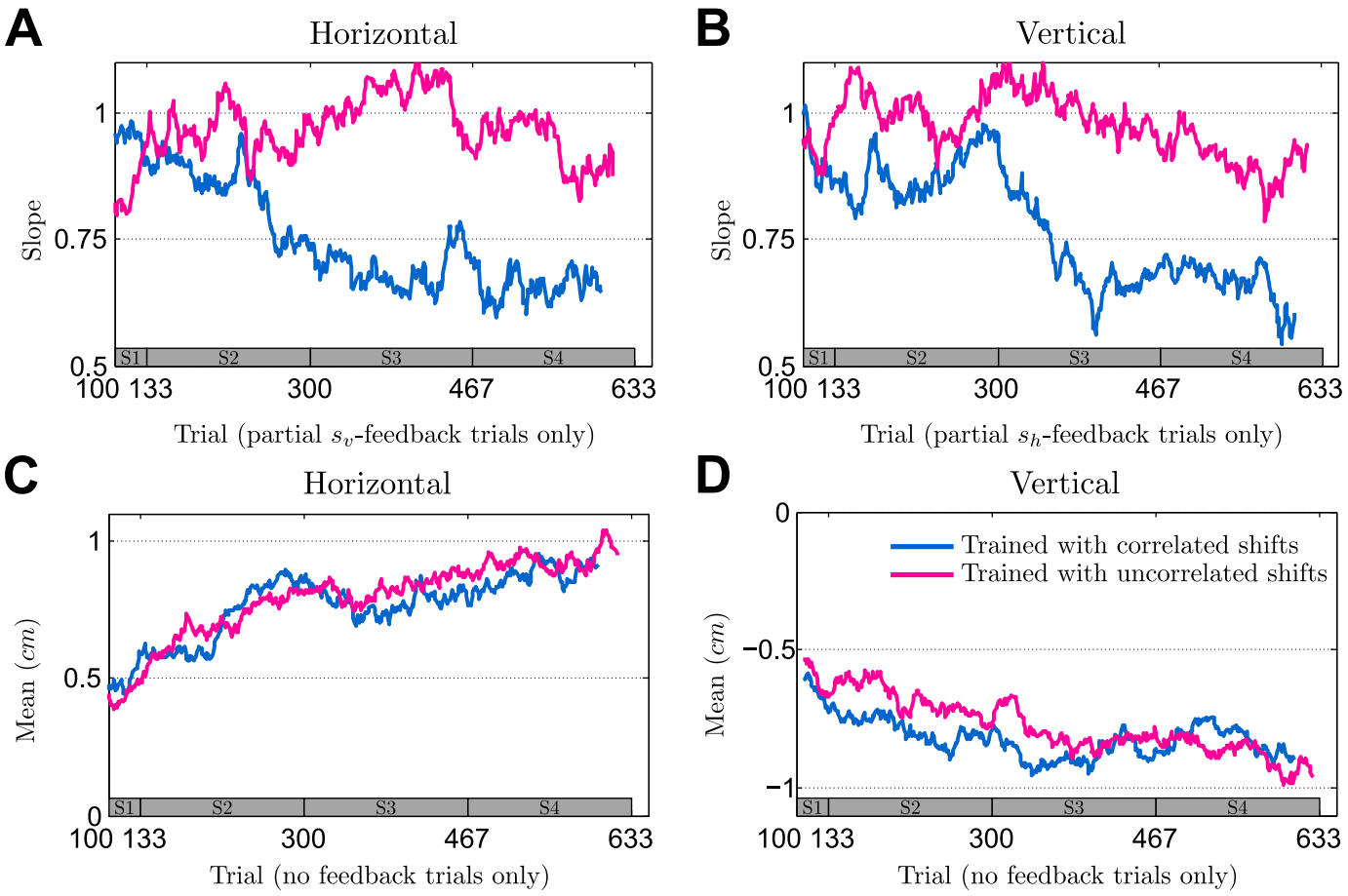

Fig 10. Simulation results. The plots show the simulation results as medians over six different simulation runs. Blue lines show the medians over six runs where the model was trained with correlated full feedback trials. Pink lines show the medians over six run where the model was trained with uncorrelated full feedback trials. **A** Median for evolution of horizontal slope in partial s_v-feedback trials—compare Fig 9A which shows the participants' results. **B** Median for evolution of vertical slope in partial s_h-feedback trials—compare Fig 9B which shows the participants' results. **C** Median for evolution of horizontal mean-response in no feedback trials—compare Fig 7 and Supplementary S2E Fig which shows the participants' results. **D** Median for evolution of vertical mean-response in no feedback trials—compare Fig 7 and Supplementary S2F Fig which shows the participants' results. The parameters of the model (strength of the initial belief over mean-shift and covariance matrix) were chosen in order to minimize the sum-of-squared-differences between the correlated simulation median and the median obtained from participants from the correlated group. The uncorrelated simulation run used the same set of parameters.

doi:10.1371/journal.pcbi.1004369.g010

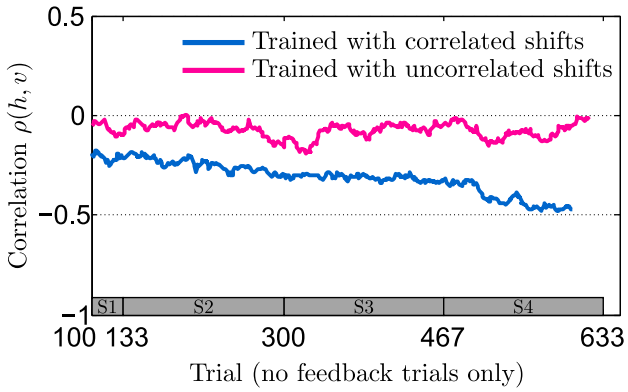

Fig 11. Simulation results. Median for evolution of xy-correlation in no feedback trials—compare Fig 8A and 8B which shows the participants' results. The plot shows the simulation results as medians over six different simulation runs. The blue line shows the median over six runs where the model was trained with correlated full feedback trials. The pink line shows the median over six runs where the model was trained with uncorrelated full feedback trials. The parameters of the model (strength of the initial belief over mean-shift and covariance matrix) were chosen in order to minimize the sum-of-squared-differences between the correlated simulation median and the median obtained from participants from the correlated group. The uncorrelated simulation run used the same set of parameters.

doi:10.1371/journal.pcbi.1004369.g011

plausible for everyday planar movements, the model can explain partial learning of the correlation structure—compare Fig 10.

Fig 10 shows the median of performance slopes and mean-responses under two different training conditions, each with six exemplary runs of the model. The blue curves show model predictions when trained on correlated trials, the pink curves show model predictions when trained on uncorrelated trials. Independent of the training regime the model predicts that both groups of participants should learn the mean shift equally well, which is in line with our experimental findings. In the case of partial feedback trials, the model predicts that the uncorrelated group shows no learning and should have a slope close to one. In contrast, when trained on correlated trials, the model predicts that the slope should decrease over time, indicating gradual learning of the correlation structure. Moreover, if actions are determined by samples from the distribution over shifts, the model predicts that the xy-correlation in no-feedback trials should gradually increase in magnitude when trained on correlated trials, whereas no such trend should be observed when trained on uncorrelated trials—simulation results are shown in Fig 11. Also this prediction fits with our experimental data shown in Fig 8.

Discussion

In this study, we designed a three-dimensional reaching task where we could displace the visual feedback of participants' hand positions with a two-dimensional translational shift. The statistics over the shift could be learnt by participants in training trials with precise visual feedback. We imposed a correlation between the two dimensions of the shift as a statistical structural invariant and found that participants gradually learnt this structural invariant during training. Participants exploited the structural knowledge to facilitate sensorimotor integration in test trials with partial feedback where the visual feedback was completely uninformative in one dimension. However, we only found this to be the case when participants had binary reward-feedback at the end of these trials. We also recorded a control group, where the correlation structure was absent during training but could have potentially been learnt through the binary reward-feedback in test trials. We found no statistically significant evidence that the correlation structure was learnt over the course of the experiment in the control group. We also found that

explicit instructions about the nature of the perturbation and the optimal compensatory response did not enhance participants' performance, which suggests that they were not following a cognitive strategy. In all groups, we used trials without any visual feedback to probe participants' prior beliefs over the shift and found that participants in all groups rapidly learnt the mean shift. Our results show that participants in our experiment were indeed able to extract structural invariants in order to enhance their performance in a Bayesian sensorimotor integration task.

In our experiment participants never learnt the correlation structure perfectly. A perfect Bayesian actor with full knowledge of the correlation should show the same behavior in partial feedback trials as in full feedback trials, as one fully visible dimension with correlation structure contains in principle the same amount of information as two fully visible dimensions. This raises the question whether learning of the correlation was still ongoing after four days of training or whether learning of the correlation is imperfect. In the latter case our results would hint at sub-optimal behavior in Bayesian integration tasks. The results in Fig 6A and 6C showing the learning progress over the course of the experiment suggest that learning had not yet flattened out by the end of the fourth session and participants would potentially continue to improve their performance in subsequent sessions of the experiment. Therefore, we cannot distinguish between these two possibilities in our data.

In control experiments we found that reward-feedback was crucial in order to improve the response in partial feedback trials. This raises the question why the group without binary reward-feedback would fail to show improvements under partial feedback despite undergoing training with correlated full feedback trials? There are at least three possibilities. First, this group might have lacked incentive in partial and no-feedback trials, as there was no performance feedback and therefore they might not have cared about their action. Second, this group might have not been able to transfer their skill from full feedback to partial feedback without additional reward cues, as the stimuli in the full- and partial-feedback conditions looked different. Third, this group might have failed to learn the correlation altogether, as in full feedback trials knowledge of the correlation is not necessary to perform well. The third hypothesis is unlikely as in previous studies of sensorimotor integration in a single dimension [8] participants were shown to learn the Bayesian prior despite the absence of any reward feedback in partial and no feedback conditions. While we cannot distinguish between the first two possibilities, our results seem to suggest that learning of the correlation in the full-feedback trials would narrow down the hypothesis space regarding the shifts sufficiently such that participants could exploit the reward-feedback either simply as an incentive (first possibility) or as a reinforcement learning signal for efficient adaptation (possibility two). In any case, the results of the uncorrelated group show that reinforcement learning with binary reward-feedback by itself is not sufficient to learn the correlation structure.

In no-feedback trials we found that participants of the correlated group showed a correlation between the vertical and horizontal dimension of their final hand position. This cannot be explained by a perfect Bayesian actor model that simply compensates the mean shift in no-feedback trials. The two possibilities we considered that could explain this finding are first, trial-by-trial adaptation of participants and second, a sampling strategy where participants sample beliefs from the prior distribution and act accordingly. In the first case, the correlation in no-feedback trials would simply show up in the correlated group as an aftereffect of the previous correlated full feedback trial. In the second case, participants would behave as bounded optimal Bayesian actors that actively sample from the learnt prior distribution rather than just picking the maximum [39–45]. Since the prior distribution exhibits the correlation structure, such a bounded optimal actor would also reflect the correlation structure in their actions in no-feedback trials. We found evidence for both hypotheses and, indeed, they are not mutually exclusive,

as a bounded rational actor could be implemented by a Monte-Carlo sampler that naturally introduces trial-by-trial correlations, because all changes in strategy are always stepwise [46, 47].

The sensorimotor integration of different sources of information has been studied previously, in particular the combination of information from different sensory modalities with different reliability and the combination of prior experience with feedback information [7–12, 19, 20]. Other studies have investigated how motor behavior adapts when perturbation statistics change dynamically across trials [21, 23, 24]. Our task belongs to the first category of studies, as there is no trial-by-trial dynamics of the perturbation, just samples from a stationary distribution. Most of these previous studies have reported a quantitative agreement between their data and Bayesian model predictions. Our task is an extension of [8], where the authors used a two-dimensional reaching task with a one-dimensional visuomotor shift to show that the human sensorimotor system optimally combines prior expectations of a hidden variable with noisy visual feedback. In their task, feedback of the virtual hand position was provided by isotropic Gaussian point clouds. In extension of this previous work, we investigate the role of higher-level statistical structure during Bayesian sensorimotor integration. The three-dimensional task setup allowed us to impose such higher-level structure in the space of the two-dimensional hidden variable, which was not possible in the planar task design in [8].

Structure learning has been proposed in the literature as an important meta-learning concept for extracting higher-level invariants in behavioral experiments, both in cognitive tasks [22, 48–52] as well as sensorimotor tasks [27, 28, 34, 35, 37, 53, 54]. In this study we investigate how structural invariants in a two-dimensional hidden variable influence sensorimotor integration, that is the combination of prior experience with uncertain feedback, where the feedback uncertainty could be manipulated experimentally. In contrast, previous studies on structure learning in sensorimotor control typically did not manipulate feedback reliability, and studies on Bayesian sensorimotor integration have typically not investigated multi-dimensional hidden variables with structured spaces. In particular, we designed visual feedback conditions where knowledge of the correlation structure allowed the integration of information across the two dimensions of the hidden shift variable. Therefore, in the current experiment the two structures (correlated vs. uncorrelated) can be subsumed by a single model with parameter ρ and the structure learning problem can be cast as learning the prior over ρ (or the covariance matrix). However, in general it need not always be the case that the models are nested. In the nomenclature of Bayesian networks structure learning refers in general to learning the dependencies between multiple (hidden) variables. These dependencies can be represented by multiple model classes M, such that structure learning implies learning a prior $p(M)$ over the model classes M. Upon arrival of new evidence, the sensorimotor system can then decide between the different models—see for example [37, 53, 54].

In our current paper, our results demonstrate that participants who were trained on a correlation structure could use their structural knowledge to guide their adaptation in test trials with binary reward-feedback. In contrast, participants in the control group that were not exposed to the correlation structure during training were unable to learn the structure in test trials from binary feedback. In summary, we find that structural invariants of hidden variables play an important role in the sensorimotor integration process of combining sensory feedback with prior experience. We find this process to be consistent with Bayesian inference.

Materials and Methods

Ethics statement

The study was approved by the ethics committee of the Max Planck Society (reference number: 0269/2010BO2). All participants gave written informed consent.

Participants

Sixteen female and eight male participants were recruited from the student population of the University of Tübingen. All participants were naive and the local standard rate of eight Euros per hour was paid for participation in the study.

Materials

We used a virtual reality setup consisting of a Sensable® Phantom® Premium 1.5 High Force manipulandum for tracking participants' hand movements in three dimensions and an NVIS® nVisor ST50 head-mounted display (HMD) for creating stereoscopic 3D virtual reality. Movement position and velocity were recorded with a rate of 1kHz.

Experimental design: overview

We designed a 3D-visuomotor task in virtual reality where participants had to perform reaching movements to a fixed target. The participants' hand position p_h was represented by a shifted hand position p_s. In each trial the virtual position p_s was translated in the vertical plane by adding a two-dimensional Gaussian random vector $s = [s_h, s_v]$, such that

$$p_s = p_h + \begin{bmatrix} s_h \\ s_v \\ 0 \end{bmatrix},$$

where the z-dimension corresponds to the veridical forward-backward movement direction and the vertical plane is spanned by h (right-left movement direction) and v (up-down movement direction). Crucially, the virtual position was only briefly displayed about halfway through the movement, which allowed inference of the unobserved shift depending on the preciseness of the display. As the hidden shift variable was bivariate and Gaussian with $s \sim \mathcal{N}(s; \mu, \Sigma)$, we could introduce statistical structure between the two dimensions of the shift by correlation, such that

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} -1 \\ +1 \end{bmatrix} cm \text{ and } \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

with $\sigma_1 = \sigma_2 = 1cm$ and the correlation coefficient ρ depending on the experimental condition. We choose a non-zero mean to be able to assess learning not only through correlation, but also through learning of the mean.

In particular, we trained the first six participants on a correlated 2D-Gaussian distribution over the shift (correlated group, $\rho = -0.999$) and the next six participants on an isotropic 2D-Gaussian distribution (uncorrelated group, $\rho = 0.0$). We refer to these training trials as full feedback trials, where the virtual position was displayed with very low uncertainty in both dimensions of the shift. We tested both groups of participants on a statistically identical set of test trials with either partial feedback or no feedback. Importantly, the shift in these test-trials was always drawn from the correlated 2D-Gaussian ($\rho = -0.999$), regardless of the group. Partial feedback trials were very reliable in one dimension but provided no information about the other dimension of the shift—only if the correlation structure had been learnt successfully, reliable feedback in one dimension allows to infer the shift in the other. No-feedback trials allowed us to test participants' learnt representation of the prior knowledge over the shift. The different feedback types are illustrated in Fig 1.

Experimental design: workspace

The workspace of the experiment was ±5*cm* in the left-right direction (*h*-axis), ±5*cm* in the up-down direction (*v*-axis) and 0–14*cm* in the forward-backward direction (*z*-axis). The *h*-*v* plane was tilted by 20° against the vertical direction of gravity to make it approximately perpendicular to participants' line of sight when looking down at the center of the workspace. The start-position was indicated by a white sphere (radius: 0.6*cm*) centered at $(h, v, z) = (0,0,0.5)cm$ and the target was indicated by a yellow sphere (radius: 0.5*cm*) centered at (0,0,14)*cm*. Before initiating the trial by moving into the start sphere, participants' virtual hand position was veridically displayed by a small cursor (blue sphere, radius: 0.3*cm*). To facilitate 3D-perception, we displayed a grid at the bottom and at the back of the workspace. We also showed a red rectangle moving along the grid to indicate the veridical depth of participants' virtual hand position.

Experimental design: trials

To start a trial, participants had to move the cursor representing their hand position into the start-sphere and remain steady for 0.1*s*. After that, a beep indicated the start of the trial. Simultaneously, the start-sphere and the cursor display vanished and the target-sphere was displayed. Participants had a maximum of 2*s* to complete their movement by passing through the target-plane located at 14*cm* in the z-direction—otherwise the trial was repeated. The average trial duration across all participants and trials was 1.041*s*.

After participants had moved 6*cm* into the forward direction towards the target, visual feedback was presented for 150*ms*. The feedback display was dynamic, that is tracking participants hand movements for the duration of the display. There were four different types of visual feedback. In *full feedback* trials (compare Fig 1A), the visual feedback consisted of a small red sphere (radius: 0.3*cm*), centered at the virtual hand position p_s. In *partial* s_h*-feedback* trials (compare Fig 1B), participants saw a vertically elongated rectangle centered on the horizontal component of p_s that consisted of 50 small red circles (radius: 0.1*cm*), each circle located randomly within the area spanned by the rectangle and re-sampled at 60*Hz*—compare Fig 1B. The bar stimulus had a width of 0.6*cm* in the horizontal direction and a height that covered the full vertical workspace, thus providing no information about the vertical component of p_s. In *partial* s_v*-feedback* trials, participants were shown the same kind of bar stimulus, but this time elongated in the horizontal direction with a height of 0.6*cm* in the vertical direction. The stimulus covered the full horizontal workspace, providing no information about the horizontal component of p_s. In *no-feedback* trials (compare Fig 1C), no visual feedback was shown to the participant. Accordingly, participants could only rely on their prior experience in these trials.

A trial was completed, once the participant crossed the vertical target-plane at $z = 14cm$ in the forward direction. This final hand-position in the vertical plane was analyzed in the Results. Regardless of the visual feedback type of the trial, the participant was informed of whether they had hit the target. A target hit was counted whenever the (potentially non-visible) final cursor position (sphere, radius: 0.3*cm*, corresponding to the final *shifted* hand-position) was intersecting with the target-sphere (radius: 0.5*cm*). To indicate a hit, the target sphere changed its color to green and a rewarding sound was played back. To indicate a miss, the target sphere changed its color to red and a deep-pitched buzzing sound was played back. In full feedback trials, the final cursor position was marked on the grid (blue sphere, radius: 0.3*cm*) in the target-plane (at $z = 14cm$).

In order to start a new trial, participants had to return their hand position to the start sphere. Since they did not see their shifted hand position represented by the cursor throughout the trial, they could use the highlighted rectangle on the grid to judge the cursor's depth. Once they moved their hand into the front half of the work space ($z \leq 7cm$), the target-sphere and

any additional final feedback (in case of full feedback trials) disappeared. Instead, the start-sphere and the veridical cursor were displayed. Participants were allowed to take breaks whenever they wanted in this inter-trial phase. Importantly, when participants returned to the start position after completion of a trial the cursor was faded out and no visual feedback of their hand position was shown. When getting close to the start position their hand position was shown veridically. Participants would thus not experience an abrupt jump in the cursor when returning to the start-position.

Experimental design: sessions

For each participant the experiment consisted of four sessions, spread over four days, with each session consisting of 1000 completed trials (see Fig 2). The first session included a three-staged training-phase (with full feedback trials only): for the first 50 trials there was no shift and the veridical cursor was displayed throughout the whole movement. In the subsequent 50 trials the shifted cursor was displayed throughout the whole trial and participants could see the jump from veridical to shifted cursor after movement onset. In the last training stage (the following 100 trials) only full feedback trials were presented, but no cursor was shown during the trial (except for the brief visual feedback of 150*ms* duration). After the training stage the different feedback types were presented in randomly interspersed order with the following probabilities: 1/2 for full feedback trials and 1/6 for partial s_h-feedback, partial s_v-feedback and no-feedback respectively. The second, third and fourth session did not include a training phase.

Experimental design: instructions

Participants were informed that their task was to hit the target with the virtual cursor and that the virtual cursor would "jump" immediately after movement onset (as they would experience in the second training stage). They were informed about the different feedback types and were told that in case of partial feedback the virtual cursor was somewhere behind the flickering bar and could not be outside the bar. In no-feedback trials they were instructed to guess where the cursor might have jumped to and try to blindly hit the target. As an additional incentive participants were shown their overall hit-ratio in partial- and no-feedback trials as a percentage above the workspace. Performance in full-feedback trials did not count towards this hit-rate display.

Experimental design: control experiments

We introduced two control groups (six participants each) to study the influence of explicit performance signals in partial and no-feedback trials and the potential impact of cognitive strategies. Like the correlated group, both control groups were exposed to correlated shifts in all feedback conditions. In the first control group, the *correlated group without auditory feedback*, participants did not receive any performance feedback about whether they had hit the target in partial- and no-feedback trials. This means that in these trials, the target color did not change according to whether the target was hit or not and a neutral sound was played back instead of the sounds indicating a hit or a miss. Additionally the hit-rate percentage in partial- and no-feedback trials was not shown to participants.

The second control group, the *correlated group with instruction* received additional instructions at the beginning of the experiment. In particular, they were informed about the correlation of the horizontal and vertical dimension of the shift. They were instructed as follows: "*If the cursor jumps to the left, it always jumps up as well and if it jumps to the right it always jumps down as well. This also means that if it jumps up it will also jump to the left and if it jumps down it will also jump to the right. This information is particularly useful for the trials with the bar-*

feedback". Participants were reminded of this instruction after the training phase ended in the first session and again before starting the second session. In order to test for trial-by-trial correlations between full feedback and no-feedback trials in this group, in approximately 8% of all trials an uncorrelated shift stimulus was presented in the full feedback trial just before a no-feedback trial. Uncorrelated full feedback trials never preceded a partial feedback trial, which is the trial type we used to evaluate learning of the correlation structure. Importantly, therefore, these uncorrelated trials do not affect the validity of the control experiment, because a cognitive strategy in partial feedback trials should not depend on the statistics of previous trials, especially if they do not directly precede.

Computational model: Bayesian sensorimotor integration

The visual feedback $d = [d_h, d_v]^{\mathrm{T}}$ is modeled using a Gaussian likelihood model: $p(d|s) = \mathcal{N}(d; s, \Sigma_{\mathrm{obs}})$. The off-diagonal entries of Σ_{obs} are zero, whereas the diagonal entries depend on the visual feedback type of the trial, that is

$$\Sigma_{\mathrm{obs}} = \begin{bmatrix} \sigma_h^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix},$$

In the full feedback trials both, the variance in h- and v-dimension are very low, in no-feedback trials the variance in both dimensions is infinite and in partial feedback trials the variance in one dimension is low whereas it is infinite in the other dimension. The posterior-belief over the shift s given the visual feedback d is obtained by combining prior knowledge over the shift with the likelihood model—leading to a Bayesian integration of both sources of information:

$$p(s|d) = \frac{p(d|s)p(s)}{\int p(d|s)p(s)ds}, \tag{1}$$

where the likelihood model is $p(d|s) = \mathcal{N}(d; s, \Sigma_{\mathrm{obs}})$ and the prior is given by $p(s) = \mathcal{N}(s; \mu, \Sigma)$ as described in Experimental design: overview.

If both the prior and the likelihood are Gaussian, the posterior can also be expressed as a Gaussian distribution $p(s|d) = \mathcal{N}(s; \mu_p, \Sigma_p)$

$$\mu_p = \Sigma_p(\Sigma_{\mathrm{obs}}^{-1}d + \Sigma^{-1}\mu) \tag{2}$$

$$\Sigma_p = (\Sigma^{-1} + \Sigma_{\mathrm{obs}}^{-1})^{-1} \tag{3}$$

with mean μ_p and covariance Σ_p. The parameters μ and Σ denote the mean and covariance-matrix of the prior and correspond to the parameters of the true distribution over the shift

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

In the four feedback conditions of our experiment, Eq (2) simplifies further to

- Full feedback condition ($\sigma_h \to 0$ and $\sigma_v \to 0$)

$$\mu_p = \begin{bmatrix} d_h \\ d_v \end{bmatrix} = d$$

- Partial s_h-feedback condition ($\sigma_v \to \infty$)

$$\mu_p = \begin{bmatrix} \frac{\sigma_1^2}{\sigma_h^2+\sigma_1^2} & 0 \\ \rho\frac{\sigma_1\sigma_2}{\sigma_h^2+\sigma_1^2} & 0 \end{bmatrix} \begin{bmatrix} d_h \\ d_v \end{bmatrix} + \begin{bmatrix} \frac{\sigma_h^2}{\sigma_h^2+\sigma_1^2} & 0 \\ -\rho\frac{\sigma_1\sigma_2}{\sigma_h^2+\sigma_1^2} & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- Partial s_v-feedback condition ($\sigma_h \to \infty$)

$$\mu_p = \begin{bmatrix} 0 & \rho\frac{\sigma_1\sigma_2}{\sigma_v^2+\sigma_2^2} \\ 0 & \frac{\sigma_2^2}{\sigma_v^2+\sigma_2^2} \end{bmatrix} \begin{bmatrix} d_h \\ d_v \end{bmatrix} + \begin{bmatrix} 1 & -\rho\frac{\sigma_1\sigma_2}{\sigma_v^2+\sigma_2^2} \\ 0 & \frac{\sigma_v^2}{\sigma_v^2+\sigma_2^2} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- no-feedback condition ($\sigma_h \to \infty$ and $\sigma_v \to \infty$)

$$\mu_p = \begin{bmatrix} \mu_h \\ \mu_v \end{bmatrix} = \mu$$

If participants maximized their hitting chances by following the maximum of the posterior given by μ_p, the only difference between the correlated and uncorrelated group occurs in the partial feedback conditions. In the uncorrelated group with $\rho = 0$, participants would integrate the informative feedback dimension with their prior information about this dimension, and they would solely rely on the prior in the uninformative feedback dimension. In the correlated group with $\rho = -0.999$ participants would differ from the uncorrelated group in how they process the uninformative feedback dimension by generating an estimate of the uninformative feedback dimension that relies on the informative feedback dimension and prior expectations on both dimensions.

Computational Model: hierarchical learning of correlation structure

In the previous section, the Bayesian integration of visual feedback information and prior knowledge about the shift requires knowledge about the parameters μ, Σ of the prior over the shift $p(s) = \mathcal{N}(s; \mu, \Sigma)$. In our experiment however, these parameters must be learnt by participants over the course of the experiment. In the Bayesian framework this learning process can be modeled by assuming a prior distribution over these parameters—the so-called hyper-prior—and updating the hyper-prior distribution in light of new observations in a Bayesian fashion. In our case the hyper-prior is again a parametric distribution (a normal inverse-Wishart distribution), which allows for a sequential Bayesian update of the parameters of this distribution, sometimes referred to as hyper-parameters. In the following model, the hyper-parameters are updated through the observed shifts in training-trials, that is in full feedback trials, only.

We denote the previously observed shifts in full feedback trials by $\mathcal{D} = \{d_1, \ldots, d_N\}$. Ultimately, we seek the belief over the shift s in the current trial after observing the visual feedback d and after having observed the previous training trials $\mathcal{D}$. This belief is formalized as the distribution $p(s|d,\mathcal{D})$. While an optimal Bayesian actor would respond with an action that corresponds to the (negative) mode of this belief, a bounded-rational Bayesian actor would sample beliefs from the distribution $p(s|d,\mathcal{D})$ and base its movement response on these samples. In our

case, we draw a single sample $\tilde{s} \sim p(s \mid d, \mathcal{D})$ and respond with $\tilde{r} = -\tilde{s}$. The distribution $p(s|d,\mathcal{D})$ is given by Bayes' rule

$$p(s|d,\mathcal{D}) = \frac{p(d|s)p(s|\mathcal{D})}{p(d|\mathcal{D})}.$$

The Gaussian likelihood model $p(d|s) = \mathcal{N}(d; s, \Sigma_{\text{obs}})$ remains the same as in the previous section. Additionally, we have introduced a data-dependent prior $p(s|\mathcal{D})$ that models the prior belief about the shift s after having observed the training data $\mathcal{D}$. The prior over the shift $p(s|\mathcal{D})$ depends on the update of the hyper-parameters μ_0, Σ_0 that specify the distribution $p(s|\mu_0, \Sigma_0)$. The update of the hyper-parameters is modeled probabilistically through $p(\mu_0, \Sigma_0|\mathcal{D}, \Sigma_{\text{obs}})$. This allows us to specify a model for Bayesian integration of prior beliefs and feedback information, where the prior beliefs are data-dependent:

$$p(s|d,\mathcal{D}) = \frac{p(d|s)p(s|\mathcal{D})}{p(d|\mathcal{D})} = \frac{p(d|s)\int d\mu_0 d\Sigma_0\; p(s|\mu_0,\Sigma_0)p(\mu_0,\Sigma_0|\mathcal{D},\Sigma_{\text{obs}})}{p(d|\mathcal{D})}. \tag{4}$$

where the update equation for the hyper-parameters μ_0, Σ_0 is given by

$$p(\mu_0,\Sigma_0|\mathcal{D},\Sigma_{\text{obs}}) = \frac{p(\mathcal{D}|\mu_0,\Sigma_0,\Sigma_{\text{obs}})p(\mu_0,\Sigma_0)}{p(\mathcal{D}|\Sigma_{\text{obs}})}, \tag{5}$$

with

$$\begin{aligned} p(\mathcal{D}|\mu_0,\Sigma_0,\Sigma_{\text{obs}}) &= \prod_{i=1}^{N} p(d_i|\mu_0,\Sigma_0,\Sigma_{\text{obs}}) \\ &= \prod_{i=1}^{N} \int ds\; \underbrace{p(d_i|s,\Sigma_{\text{obs}})}_{\mathcal{N}(d_i;s,\Sigma_{\text{obs}})}\; \underbrace{p(s|\mu_0,\Sigma_0)}_{\mathcal{N}(s;\mu_0,\Sigma_0)} \\ &= \prod_{i=1}^{N} \mathcal{N}(d_i;\mu_0,\Sigma_0+\Sigma_{\text{obs}}) = \prod_{i=1}^{N} \mathcal{N}(d_i;\psi,\Theta). \end{aligned} \tag{6}$$

It is crucial to note that the likelihood of a previously observed data point d_i has a Gaussian form $\mathcal{N}(d_i; \mu_0, \Sigma_0+\Sigma_{\text{obs}}) = \mathcal{N}(d_i; \psi, \Theta)$—see the standard textbooks [55, 56] by Bishop (2.115) or Murphy (4.126). If we replace μ_0, Σ_0 with ψ, Θ in Eq (5) and subsume Σ_{obs}, we can use a *normal inverse-Wishart* distribution as a prior distribution $p(\psi,\Theta) = \text{NIW}(\psi,\Theta)$, which is the conjugate prior for a Gaussian with unknown mean and covariance matrix. Conveniently, this leads to closed-form sequential update equations for the posterior parameters of the normal inverse-Wishart distribution after having observed N data-points.

$$p(\psi,\Theta|\mathcal{D}_N) = \frac{p(\mathcal{D}_N|\psi,\Theta)p(\psi,\Theta)}{p(\mathcal{D}_N)} = \text{NIW}(\psi,\Theta|m_N,\kappa_N,\nu_N,S_N) \tag{7}$$

$$m_N = \frac{\kappa_0 m_0 + N\bar{D}}{\kappa_N} \tag{8}$$

$$\kappa_N = \kappa_0 + N \tag{9}$$

$$\nu_N = \nu_0 + N \tag{10}$$

$$S_N = S_0 + S_{\bar{D}} + \frac{\kappa_0 N}{\kappa_0 + N}(\bar{D} - m_0)(\bar{D} - m_0)^{\mathsf{T}} \tag{11}$$

with $D = \frac{1}{N}\sum_{i=1}^{N} d_i$ being the empirical mean-shift and $S\bar{D} = \sum_{i=1}^{N}(d_i - \bar{D})(d_i - \bar{D})^{T}$ (see [56] Murphy section 4.6.3.3).

Putting it all together (and correcting for the subsumed Σ_{obs} in the hyper-prior) we get the following rejection-sampling scheme to simulate a participant:

1. Sample from $p(\mu_0,\Sigma_0|\mathcal{D},\Sigma_{\text{obs}})$ as given by Eq (5)
   a. Draw a sample from the normal inverse-Wishart $\tilde{\psi},\tilde{\Theta} \sim p(\psi, \Theta \mid \mathcal{D}_N)$
   b. $\tilde{\mu}_0 = \tilde{\psi}$ (follows from the last equality in Eq (6))
   c. $\tilde{\Sigma}_0 = \tilde{\Theta} - \Sigma_{\text{obs}}$ (follows from the last equality in Eq (6), always use the full feedback Σ_{obs} in this particular step as the model is trained on full feedback trials only)
   d. If $\tilde{\Sigma}_0$ is not positive semi-definite (that is if it has eigenvalues ≤ 0), discard samples and re-start at the first step, otherwise continue.
2. For a given $\tilde{\mu}_0, \tilde{\Sigma}_0$, draw a sample from $\tilde{s} \sim p(s \mid \tilde{\mu}_0,\tilde{\Sigma}_0)$.
3. Perform a rejection-acceptance step with the likelihood of the observed feedback given the sampled shift $p(d \mid \tilde{s})$. To do so evaluate if $u \leq p(d \mid \tilde{s})/l_{\max}$ for $u \sim \mathcal{U}[0; 1]$ and accept the sample if the inequality holds or reject otherwise. $l_{\max}$ is the maximum value of the likelihood given by $l_{\max} = 1\Big/\sqrt{\mid \Sigma_{\text{obs}} \mid (2\pi)^2}$, where $\mid \cdot \mid$ denotes the determinant.
4. If the sample was accepted, respond to the stimulus with a response $\tilde{r} = -\tilde{s}$. If the sample was rejected, restart at the first step.
5. In case of a full feedback trial, update the parameters of the normal inverse-Wishart with the sequential update rules for the parameters following Eqs (8)–(11).

For our simulation we used the following parameters. The initial belief about the mean-shift was chosen as $m_0 = [0, 0]^{\mathrm{T}}$ with an initial weight of $\kappa_0 = 300$. The initial belief about the covariance matrix was set to a diagonal matrix (no correlation between the horizontal and vertical dimension) with a variance of one for both dimensions with an initial weight of $\nu_0 = 3000$. For the inverse-Wishart prior, S_0 must then be specified in the following way:

$$S_0 = \begin{bmatrix} \nu_0 & 0 \\ 0 & \nu_0 \end{bmatrix}.$$

The weights of the initial beliefs κ_0 and ν_0 were determined by averaging over 30 simulation runs and then comparing the resulting medians of the quantities shown in Fig 10 and Fig 11 to the medians obtained from the participants of the correlated group (that is the median slopes in horizontal and vertical dimension, the median means in both dimensions as well as the median correlation in no feedback trials). In particular, we performed a grid-search over a range of parameter-values such that the sum-of-squared-errors between the time course of simulated medians and the participants' median was minimized. We found that the weights on the initial beliefs directly govern the learning-rates (as expected), which allows to reproduce a broad range of learning-behavior. The results obtained are not particularly sensitive to small changes in the parameters.

The results shown in Fig 10 and Fig 11 were obtained by taking the median over six virtual participants with the best-fit parameters. In the figure we compare two different runs—one run where the model was trained on correlated shifts (identical to the shifts experienced by the six participants in the correlated group) and another run where the model was trained on uncorrelated shifts (identical to the shifts experienced by the six participants of the uncorrelated group) without changing the model parameters.

The covariance matrix of the observation noise Σ_{obs} was dependent on the trial type, but was always a diagonal matrix (no correlation in the observation noise). For full feedback trials, both diagonal entries were set to $0.2cm^2$ reflecting reliable feedback. For partial-s_h feedback trials the entry for the horizontal dimension was $0.2cm^2$ and the entry for the vertical dimensions was set to $40cm^2$ as the feedback provided reliable information in the horizontal dimension and no information in the vertical dimension. For the partial s_v feedback trials the entries were reversed—the horizontal dimension was set to $40cm^2$ and the entry for the vertical dimensions was $0.2cm^2$. For the no feedback trials both diagonal entries were set to $40cm^2$ as the feedback provided no information about the shift in either dimension.

Supporting Information

S1 Fig. Evolution of slopes in partial feedback trials—individual participants and group medians for correlated group without auditory feedback and correlated group with instruction. Changes in slope in partial feedback trials. The slope is a performance measure determined as in Fig 3 in the main manuscript but using a sliding window of 100 trials. For the analysis only partial s_v- or partial s_h-feedback trials were taken out from the pooled data across all sessions. Thin colored lines indicate individual participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The thick black line shows the median over participants—taking only into account trials where data from all participants exists. The marked ticks on the x-axis at the bottom of the figure indicate the end of the corresponding session (on average). **A** Evolution of the horizontal slopes in partial s_v feedback trials of the correlated group without auditory feedback. Horizontal information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. **B** Same as A but showing data of the instructed group. **C** Evolution of the vertical slopes in s_h feedback trials of the correlated group without auditory feedback. Vertical information is not given by the feedback, but can only be obtained through knowledge of the correlation structure. **D** Same as C but showing data of the instructed group.
(EPS)

S2 Fig. Evolution of means in no feedback trials—individual participants and group medians for correlated group without auditory feedback and correlated group with instruction. Learning of mean shift over all sessions revealed by performance in no-feedback trials averaged over a sliding window of 100 trials. For the analysis only no-feedback trials were taken out from the pooled data across all sessions. Thin colored lines indicate individual participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. Thick black lines show the median over participants—taking only into account trials where data from all participants exists. The bar at the bottom of the figure indicates the corresponding session (on average). **A** Learning of the mean in the horizontal dimension of the correlated group without auditory feedback. **B** Same as A but showing data of the instructed group. **C** Learning of the mean in the vertical dimension of the correlated group without auditory feedback. **D** Same as C but showing data of the instructed group. **E** Learning of the mean in the horizontal dimension—showing the medians of all groups of participants. **F**

Learning of the mean in the vertical dimension—showing the medians of all groups of participants.
(EPS)

S3 Fig. Evolution of xy-correlation in partial feedback trials and low-uncertainty-dimension slopes in partial feedback trials—medians for all groups. Changes in correlation and low-uncertainty-slope in partial feedback trials using a sliding window of 100 trials. For the analysis only partial s_v- or partial s_h-feedback trials were taken out from the pooled data across all sessions. Different colored lines show the median over the different groups of participants and can vary in length since the exact number of relevant trials could fluctuate due to the probabilistic generation of trials. The marked ticks on the x-axis at the bottom of the figure indicate the end of the corresponding session (on average) **A** Adaptation of correlation between the vertical and horizontal terminal hand position measured in partial s_h-feedback trials. Large magnitudes of the correlation indicate a more "diagonal" movement, which is required by the optimal response in these trials. **B** Adaptation of correlation between the vertical and horizontal terminal hand position measured in partial s_v-feedback trials. Large magnitudes of the correlation indicate a more "diagonal" movement, which is required by the optimal response in these trials. **C** Evolution of the horizontal slopes in partial s_h feedback trials where horizontal information is given by the feedback with low uncertainty. Ideally, the value of this slope would be close to zero. **D** Evolution of the vertical slopes in partial s_v feedback trials where vertical information is given by the feedback with low uncertainty. Ideally, the value of this slope would be close to zero. In the upper panels it can be seen that the instructed group initially shows an increased magnitude in movement correlation in partial-feedback trials which indicates that they understood and followed the instruction. However in Fig 9 in the main manuscript it can be seen that the instructed group does not have a decreased slope in these trials. In contrast, their slope in the low-uncertainty dimension in these trials was increased compared to the other groups (shown in lower panels of this figure). This suggests that the instruction was not helpful but rather impeded their shift-compensation in the low-uncertainty dimension.
(EPS)

S1 Dataset. Data recorded from the experiment. All data required for reproducing the results and figures presented in the paper.
(ZIP)

Author Contributions

Conceived and designed the experiments: DAB TG. Performed the experiments: TG EH ZR. Analyzed the data: TG EH ZR. Wrote the paper: DAB TG.

References

1. Rao RP. An optimal estimation approach to visual perception and learning. Vision Res. 1999; 39(11): 1963–1989. doi: 10.1016/S0042-6989(98)00279-X PMID: 10343783
2. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 2004; 27(12):712–719. doi: 10.1016/j.tins.2004.10.007 PMID: 15541511
3. Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. Trends Cogn Sci. 2006; 10(7):319–326. doi: 10.1016/j.tics.2006.05.003 PMID: 16807063
4. Wolpert DM. Probabilistic models in human sensorimotor control. Hum Mov Sci. 2007; 26(4):511–524. doi: 10.1016/j.humov.2007.05.005 PMID: 17628731
5. Körding K. Decision Theory: What Should the Nervous System Do? Science. 2007; 318(5850):606–610. PMID: 17962554

6. Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. Nat Neurosci. 2013; 16(9):1170–1178. doi: 10.1038/nn.3495 PMID: 23955561
7. Jacobs RA. Optimal integration of texture and motion cues to depth. Vision Res. 1999; 39(21): 3621–3629. doi: 10.1016/S0042-6989(99)00088-7 PMID: 10746132
8. Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. Nature. 2004; 427(6971): 244–247. doi: 10.1038/nature02169 PMID: 14724638
9. Hudson TE, Maloney LT, Landy MS. Movement planning with probabilistic target information. J Neurophysiol. 2007; 98(5):3034–3046. doi: 10.1152/jn.00858.2007 PMID: 17898140
10. Girshick AR, Banks MS. Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. J Vis. 2009; 9(9):8. doi: 10.1167/9.9.8 PMID: 19761341
11. Turnham EJ, Braun DA, Wolpert DM. Inferring visuomotor priors for sensorimotor learning. PLoS computational biology. 2011; 7(3):e1001112. doi: 10.1371/journal.pcbi.1001112 PMID: 21483475
12. Grau-Moya J, Ortega PA, Braun DA. Risk-sensitivity in bayesian sensorimotor integration. PLoS Comput Biol. 2012; 8(9):e1002698. doi: 10.1371/journal.pcbi.1002698 PMID: 23028294
13. Langer MS, Bülthoff HH. A prior for global convexity in local shape-from-shading. Perception. 2001; 30(4):403–410. doi: 10.1068/p3178 PMID: 11383189
14. Weiss Y, Simoncelli EP, Adelson EH. Motion illusions as optimal percepts. Nat Neurosci. 2002; 5(6): 598–604. doi: 10.1038/nn0602-858 PMID: 12021763
15. Geisler WS, Kersten D. Illusions, perception and Bayes. Nat Neurosci. 2002; 5(6):508–510. doi: 10.1038/nn0602-508 PMID: 12037517
16. Adams WJ, Graf EW, Ernst MO. Experience can change the 'light-from-above' prior. Nat Neurosci. 2004; 7(10):1057–1058. doi: 10.1038/nn1312 PMID: 15361877
17. Stocker AA, Simoncelli EP. Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci. 2006; 9(4):578–585. doi: 10.1038/nn1669 PMID: 16547513
18. Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput. 2007; 19(12):3335–3355. doi: 10.1162/neco.2007.19.12.3335 PMID: 17970656
19. van Beers RJ, Sittig AC, Gon JJ. Integration of proprioceptive and visual position-information: An experimentally supported model. J Neurophysiol. 1999; 81(3):1355–1364. PMID: 10085361
20. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002; 415(6870):429–433. doi: 10.1038/415429a PMID: 11807554
21. Baddeley R, Ingram H, Miall R. System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. J Neurosci. 2003; 23(7):3066–3075. PMID: 12684493
22. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. PLoS one. 2007; 2(9):e943. doi: 10.1371/journal.pone.0000943 PMID: 17895984
23. Burge J, Ernst MO, Banks MS. The statistical determinants of adaptation rate in human reaching. J Vis. 2008; 8(4):20. doi: 10.1167/8.4.20 PMID: 18484859
24. Landy MS, Trommershäuser J, Daw ND. Dynamic estimation of task-relevant variance in movement under risk. J Neurosci. 2012; 32(37):12702–12711. doi: 10.1523/JNEUROSCI.6160-11.2012 PMID: 22972994
25. Acerbi L, Vijayakumar S, Wolpert DM. On the origins of suboptimality in human probabilistic inference. PLoS Comput Biol. 2014; 10(6):e1003661. doi: 10.1371/journal.pcbi.1003661 PMID: 24945142
26. Diedrichsen J. Optimal task-dependent changes of bimanual feedback control and adaptation. Curr Biol. 2007; 17(19):1675–1679. doi: 10.1016/j.cub.2007.08.051 PMID: 17900901
27. Braun DA, Aertsen A, Wolpert DM, Mehring C. Motor task variation induces structural learning. Curr Biol. 2009; 19(4):352–357. doi: 10.1016/j.cub.2009.01.036 PMID: 19217296
28. Braun DA, Aertsen A, Wolpert DM, Mehring C. Learning optimal adaptation strategies in unpredictable motor tasks. J Neurosci. 2009; 29(20):6472–6478. doi: 10.1523/JNEUROSCI.3075-08.2009 PMID: 19458218
29. Wolpert DM, Flanagan JR. Motor learning. Current Biology. 2010; 20(11):R467–R472. doi: 10.1016/j.cub.2010.04.035 PMID: 20541489
30. Diedrichsen J, Shadmehr R, Ivry RB. The coordination of movement: optimal feedback control and beyond. Trends Cogn Sci. 2010; 14(1):31–39. doi: 10.1016/j.tics.2009.11.004 PMID: 20005767
31. Johnson RL, Culmer PR, Burke MR, Mon-Williams M, Wilkie RM. Exploring structural learning in handwriting. Exp Brain Res. 2010; 207(3–4):291–295. doi: 10.1007/s00221-010-2438-5 PMID: 20972778
32. Braun DA, Waldert S, Aertsen A, Wolpert DM, Mehring C. Structure learning in a sensorimotor association task. PloS one. 2010; 5(1):e8973. doi: 10.1371/journal.pone.0008973 PMID: 20126409

33. Yousif N, Diedrichsen J. Structural learning in feedforward and feedback control. J Neurophysiol. 2012; 108(9):2373–2382. doi: 10.1152/jn.00315.2012 PMID: 22896725

34. Kobak D, Mehring C. Adaptation paths to novel motor tasks are shaped by prior structure learning. J Neurosci. 2012; 32(29):9898–9908. doi: 10.1523/JNEUROSCI.0958-12.2012 PMID: 22815505

35. Narain D, Mamassian P, van Beers RJ, Smeets JBJ, Brenner E. How the Statistics of Sequential Presentation Influence the Learning of Structure. PLoS ONE. 2013 04; 8(4):e62276. doi: 10.1371/journal.pone.0062276 PMID: 23638022

36. Ranganathan R, Wieser J, Mosier KM, Mussa-Ivaldi FA, Scheidt RA. Learning redundant motor tasks with and without overlapping dimensions: facilitation and interference effects. J Neurosci. 2014; 34(24): 8289–8299. doi: 10.1523/JNEUROSCI.4455-13.2014 PMID: 24920632

37. Narain D, Smeets JBJ, Mamassian P, Brenner E, van Beers RJ. Structure learning and the Occam's razor principle: A new view of human function acquisition. Frontiers in Computational Neuroscience. 2014; 8(121). doi: 10.3389/fncom.2014.00121 PMID: 25324770

38. Shea CH, Wulf G. Enhancing motor learning through external-focus instructions and feedback. Hum Mov Sci. 1999; 18(4):553–571. doi: 10.1016/S0167-9457(99)00031-7

39. Vulkan N. An economist's perspective on probability matching. Journal of economic surveys. 2000; 14(1):101–118. doi: 10.1111/1467-6419.00106

40. Mattsson LG, Weibull JW. Probabilistic choice and procedurally bounded rationality. Games and Economic Behavior. 2002; 41(1):61–78. doi: 10.1016/S0899-8256(02)00014-3

41. Ortega PA, Braun DA. A minimum relative entropy principle for learning and acting. Journal of Artificial Intelligence Research. 2010; 38(1):475–511.

42. Lieder F, Griffiths T, Goodman N. Burn-in, bias, and the rationality of anchoring. In: Advances in neural information processing systems; 2012. p. 2690–2798.

43. May BC, Korda N, Lee A, Leslie DS. Optimistic Bayesian sampling in contextual-bandit problems. The Journal of Machine Learning Research. 2012; 13(1):2069–2106.

44. Bonawitz E, Denison S, Griffiths TL, Gopnik A. Probabilistic models, learning algorithms, and response variability: sampling in cognitive development. Trends Cogn Sci. 2014; 18(10):497–500. doi: 10.1016/j.tics.2014.06.006 PMID: 25001609

45. Vul E, Goodman N, Griffiths TL, Tenenbaum JB. One and done? optimal decisions from very few samples. Cognitive science. 2014; 38(4):599–637. doi: 10.1111/cogs.12101 PMID: 24467492

46. Ortega PA, Braun DA, Tishby N. Monte Carlo methods for exact & efficient solution of the generalized optimality equations. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE; 2014. p. 4322–4327.

47. Haith A, Krakauer J. Motor Learning by Sequential Sampling of Actions. In: Translational and Computational Motor Control (TCMC) 2014. American Society of Neurorehabilitation; 2014. p. 2.

48. Tenenbaum JB, Griffiths TL, Kemp C. Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn Sci. 2006; 10(7):309–318. doi: 10.1016/j.tics.2006.05.009 PMID: 16797219

49. Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput. 2007; 19(12):3335–3355. doi: 10.1162/neco.2007.19.12.3335 PMID: 17970656

50. Kemp C, Tenenbaum JB. The discovery of structural form. Proc Natl Acad Sci U S A. 2008; 105(31): 10687–10692. doi: 10.1073/pnas.0802631105 PMID: 18669663

51. Kemp C, Tenenbaum JB. Structured statistical models of inductive reasoning. Psychol Rev. 2009 Jan; 116(1):20–58. doi: 10.1037/a0014282 PMID: 19159147

52. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: statistics, structure, and abstraction. Science. 2011; 331(6022):1279–1285. doi: 10.1126/science.1192788 PMID: 21393536

53. Genewein T, Braun DA. A sensorimotor paradigm for Bayesian model selection. Frontiers in Human Neuroscience. 2012; 6(291). doi: 10.3389/fnhum.2012.00291 PMID: 23125827

54. Genewein T, Braun DA. Occam's Razor in sensorimotor learning. Proceedings of the Royal Society of London B: Biological Sciences. 2014; 281 (1783). doi: 10.1098/rspb.2013.2952

55. Bishop CM. Pattern recognition and machine learning. Springer New York; 2006.

56. Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012.

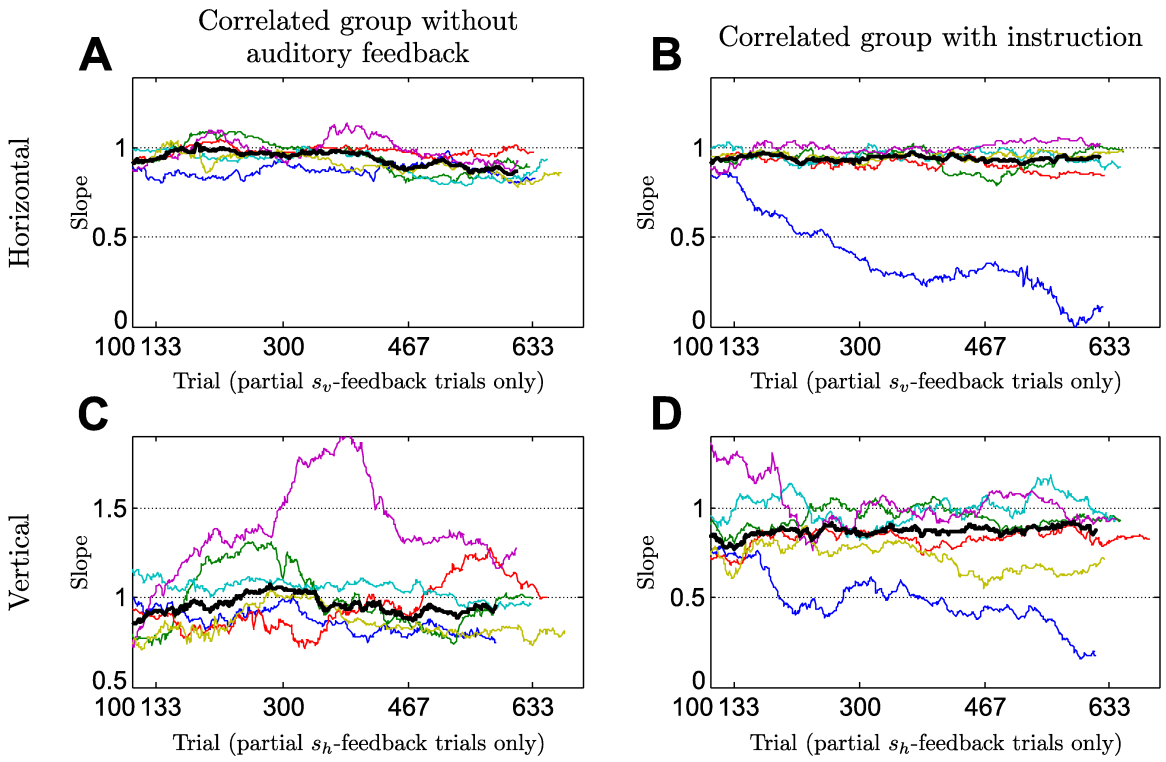

Figure 6.1: Supplementary Figure S1

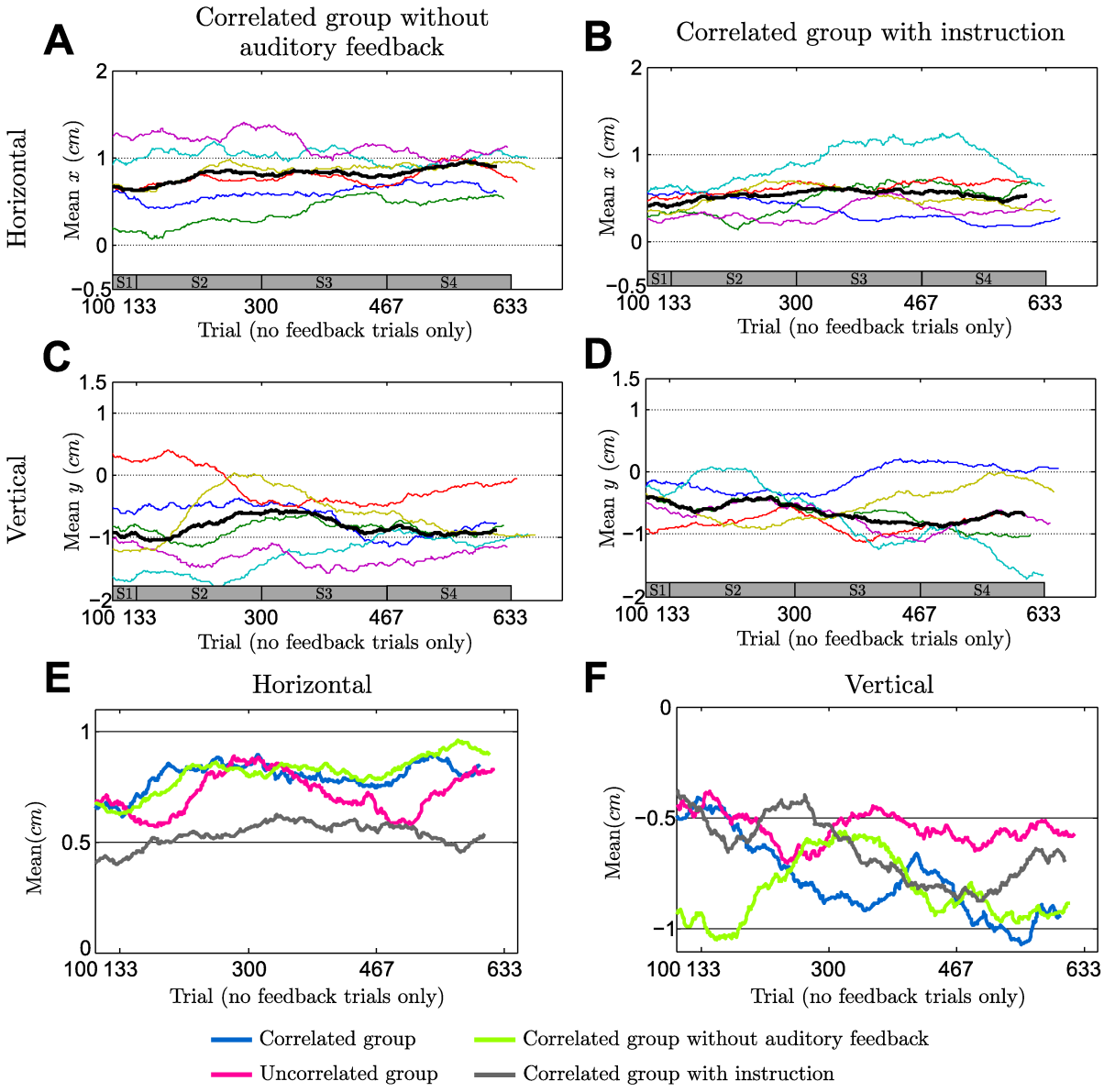

Figure 6.2: Supplementary Figure S2

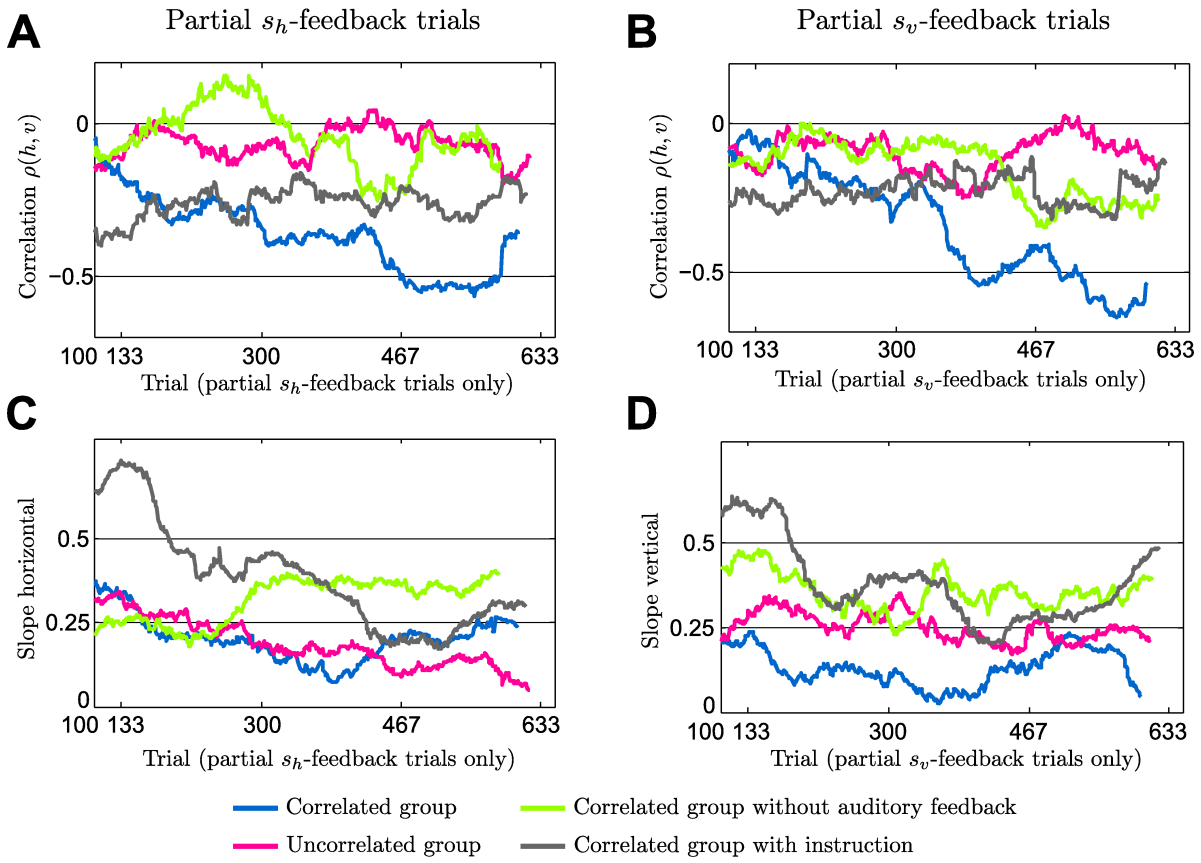

Figure 6.3: Supplementary Figure S3

7 Bounded Rationality, Abstraction, and Hierarchical Decision-Making: An Information-Theoretic Optimality Principle

For a color-version of the plots in this chapter please see the digital version of this thesis or the original publication [Genewein et al., 2015b].

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 11 November 2015
doi: 10.3389/frobt.2015.00027

Bounded Rationality, Abstraction, and Hierarchical Decision-Making: An Information-Theoretic Optimality Principle

Tim Genewein[1,2,3]*, *Felix Leibfried*[1,2,3], *Jordi Grau-Moya*[1,2,3] and *Daniel Alexander Braun*[1,2]

[1] Max Planck Institute for Intelligent Systems, Tübingen, Germany, [2] Max Planck Institute for Biological Cybernetics, Tübingen, Germany, [3] Graduate Training Centre of Neuroscience, Tübingen, Germany

Abstraction and hierarchical information processing are hallmarks of human and animal intelligence underlying the unrivaled flexibility of behavior in biological systems. Achieving such flexibility in artificial systems is challenging, even with more and more computational power. Here, we investigate the hypothesis that abstraction and hierarchical information processing might in fact be the consequence of limitations in information-processing power. In particular, we study an information-theoretic framework of bounded rational decision-making that trades off utility maximization against information-processing costs. We apply the basic principle of this framework to perception-action systems with multiple information-processing nodes and derive bounded-optimal solutions. We show how the formation of abstractions and decision-making hierarchies depends on information-processing costs. We illustrate the theoretical ideas with example simulations and conclude by formalizing a mathematically unifying optimization principle that could potentially be extended to more complex systems.

Keywords: information theory, bounded rationality, computational rationality, rate-distortion, decision-making, hierarchical architecture, perception-action system, lossy compression

OPEN ACCESS

Edited by:
Joschka Boedecker, University of Freiburg, Germany

Reviewed by:
Dimitrije Markovic, Dresden University of Technology, Germany
Sam Neymotin, State University of New York Downstate Medical Center, USA

***Correspondence:**
Tim Genewein
tim.genewein@tuebingen.mpg.de

Specialty section:
This article was submitted to Computational Intelligence, a section of the journal Frontiers in Robotics and AI

Received: *31 August 2015*
Accepted: *23 October 2015*
Published: *11 November 2015*

Citation:
Genewein T, Leibfried F, Grau-Moya J and Braun DA (2015) Bounded Rationality, Abstraction, and Hierarchical Decision-Making: An Information-Theoretic Optimality Principle. Front. Robot. AI 2:27. doi: 10.3389/frobt.2015.00027

1. INTRODUCTION

A key characteristic of intelligent systems, both biological and artificial, is the ability to flexibly adapt behavior in order to interact with the environment in a way that is beneficial to the system. In biological systems, the ability to adapt affects the fitness of an organism and becomes key to survival not only of individual organisms but species as a whole. Both in the theoretical study of biological systems and in the design of artificial intelligent systems, the central goal is to understand adaptive behavior formally. A formal framework for tackling the problem of general adaptive systems is decision-theory, where behavior is conceptualized as a series of optimal decisions or actions that a system performs in order to respond to changes to the input of the system. An important idea, originating from the foundations of decision-theory, is the maximum expected utility (MEU) principle (Ramsey, 1931; Von Neumann and Morgenstern, 1944; Savage, 1954). Following MEU, an intelligent system is formalized as a decision-maker that chooses actions in order to maximize the desirability of the expected outcome of the action, where the desirability of an outcome is quantified by a utility function.

A fundamental problem of MEU is that the computation of an optimal action can easily exceed the computational capacity of a system. It is for example in general prohibitive trying to compute

an optimal chess move due to the large number of possibilities. One way to deal with such problems is to study optimal decision-making with information-processing constraints. Following the pioneering work of Simon (1955, 1972) on bounded rationality, decision-making with limited information-processing resources has been studied extensively in psychology (Gigerenzer and Todd, 1999; Camerer, 2003; Gigerenzer and Brighton, 2009), economics (McKelvey and Palfrey, 1995; Rubinstein, 1998; Kahneman, 2003; Parkes and Wellman, 2015), political science (Jones, 2003), industrial organization (Spiegler, 2011), cognitive science (Howes et al., 2009; Janssen et al., 2011), computer science, and artificial intelligence research (Horvitz, 1988; Lipman, 1995; Russell, 1995; Russell and Subramanian, 1995; Russell and Norvig, 2002; Lewis et al., 2014). Conceptually, the approaches differ widely ranging from heuristics (Tversky and Kahneman, 1974; Gigerenzer and Todd, 1999; Gigerenzer and Brighton, 2009; Burns et al., 2013) to approximate statistical inference schemes (Levy et al., 2009; Vul et al., 2009, 2014; Sanborn et al., 2010; Tenenbaum et al., 2011; Fox and Roberts, 2012; Lieder et al., 2012).

In this study, we use an information-theoretic model of bounded rational decision-making (Braun et al., 2011; Ortega and Braun, 2012, 2013; Braun and Ortega, 2014; Ortega and Braun, 2014; Ortega et al., 2014) that has precursors in the economic literature (McKelvey and Palfrey, 1995; Mattsson and Weibull, 2002; Sims, 2003, 2005, 2006, 2010; Wolpert, 2006) and that is closely related to recent advances in the information theory of perception-action systems (Todorov, 2007, 2009; Still, 2009; Friston, 2010; Peters et al., 2010; Tishby and Polani, 2011; Daniel et al., 2012, 2013; Kappen et al., 2012; Rawlik et al., 2012; Rubin et al., 2012; Neymotin et al., 2013; Tkačik and Bialek, 2014; Palmer et al., 2015). The basis of this approach is formalized by a free energy principle that trades off expected utility, and the cost of computation that is required to adapt the system accordingly in order to achieve high utility. Here, we consider an extension of this framework to systems with multiple information-processing nodes and in particular discuss the formation of information-processing hierarchies, where different levels in the hierarchy represent different levels of abstraction. The basic intuition is that information-processing nodes with little computational resources can adapt only a little for different inputs and are therefore forced to treat different inputs in the same or a similar way, that is the system has to abstract (Genewein and Braun, 2013). Importantly, abstractions arising in decision-making hierarchies are a core feature of intelligence (Kemp et al., 2007; Braun et al., 2010a,b; Gershman and Niv, 2010; Tenenbaum et al., 2011) and constitute the basis for flexible behavior.

The paper is structured as follows. In Section 2, we recapitulate the information-theoretic framework for decision-making and show its fundamental connection to a well-known trade-off in information theory (the rate-distortion problem for lossy compression). In Section 3, we show how the extension of the basic trade-off principle leads to a theoretically grounded design principle that describes how perception is shaped by action. In Section 4, we apply the basic trade-off between expected utility and computational cost to a two-level hierarchy and show how this leads to emergent, bounded-optimal hierarchical decision-making systems. In Section 5, we present a mathematically unifying formulation that provides a starting point for generalizing the principles presented in this paper to more complex architectures.

2. BOUNDED RATIONAL DECISION-MAKING

2.1. A Free Energy Principle for Bounded Rationality

In a decision-making task with context, an actor or agent is presented with a world-state w and is then faced with finding an optimal action a_w^* out of a set of actions $\mathcal{A}$ in order to maximize the utility $U(w, a)$:

$$a_w^* = \arg\max_a U(w, a). \tag{1}$$

If the cardinality of the action-set is large, the search for the single best action can become computationally very costly. For an agent with limited computational resources that has to react within a certain time-limit, the search problem can potentially become infeasible. In contrast, biological agents, such as animals and humans, are constantly confronted with picking an action out of a very large set of possible actions. For instance, when planning a movement trajectory for grasping a certain object with a biological arm with many degrees of freedom, the number of possible trajectories is infinite. Yet, humans are able to quickly find a trajectory that is not necessarily optimal but good enough. The paradigm of picking a good enough solution that is actually computable has been termed *bounded rational* acting (Simon, 1955, 1972; Horvitz, 1988; Horvitz et al., 1989; Horvitz and Zilberstein, 2001). Note that bounded rational policies are in general stochastic and thus expressed as a probability distribution over actions given a world-state $p(a|w)$.

We follow the work of Ortega and Braun (2013), where the authors present a mathematical framework for bounded rational decision-making that takes into account computational limitations. Formally, an agent's initial behavior (or search strategy through action-space) is described by a prior distribution $p_0(a)$. The agent transforms its behavior to a posterior $p(a|w)$ in order to maximize expected utility $\Sigma_a p(a|w)U(w, a)$ under this posterior policy. The computational cost of this transformation is measured by the KL-divergence between prior and posterior and is upper-bounded in case of a bounded rational actor. Decision-making with limited computational resources can then be formalized with the following constrained optimization problem:

$$\begin{aligned} p^*(a|w) &= \arg\max_{p(a|w)} \sum_a p(a|w)U(w, a) \\ &\text{s.t. } D_{\mathrm{KL}}(p(a|w)||p_0(a)) \leq K. \end{aligned} \tag{2}$$

This principle models bounded rational actors that initially follow a prior policy $p_0(a)$ and then use information about the world-state w to adapt their behavior to $p(a|w)$ in a way that optimally trades off the expected gain in utility against the transformation costs for adapting from $p_0(a)$ to $p(a|w)$. The constrained optimization problem in equation (2) can be rewritten as an unconstrained

variational problem using the method of Lagrange multipliers:

$$p^*(a|w) = \arg\max_{p(a|w)} \underbrace{\sum_a p(a|w)U(w,a)}_{\mathbf{E}_{p(a|w)}[U(w,a)]} - \frac{1}{\beta}\underbrace{\sum_a p(a|w)\log\frac{p(a|w)}{p_0(a)}}_{D_{\mathrm{KL}}(p(a|w)||p_0(a))}, \tag{3}$$

where β is known as the *inverse temperature*. The inverse temperature acts as a conversion-factor, translating the amount of information imposed by the transformation (usually measured in nats or bits) into a cost with the same units as the expected utility (utils). The distribution $p^*(a|w)$ that maximizes the variational principle is given by

$$p^*(a|w) = \frac{1}{Z(w)}p_0(a)e^{\beta U(w,a)}, \tag{4}$$

with *partition sum* $Z(w) = \Sigma_a\, p_0(a)\, e^{\beta U(w,a)}$. Evaluating equation (3) with the maximizing distribution $p^*(a|w)$ yields the *free energy difference*

$$\begin{aligned}\Delta F(w) &= \max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta}D_{\mathrm{KL}}(p(a|w)||p_0(a)) \\ &= \frac{1}{\beta}\log Z(w),\end{aligned} \tag{5}$$

which is well known in thermodynamics and quantifies the energy of a system that can be converted to work. $\Delta F(w)$ is composed of the expected utility under the posterior policy $p^*(a|w)$ minus information processing cost that is required for computing the posterior policy measured as the Kullback-Leibler (KL) divergence between the posterior $p^*(a|w)$ and the prior $p_0(a)$.

The inverse temperature β governs the influence of the transformation cost and thus the boundedness of the actor which determines the maximally allowed deviation of the final behavior $p^*(a|w)$ from the initial behavior $p_0(a)$. A perfectly rational actor that maximizes its utility can be recovered as the limit case $\beta \to \infty$ where transformation cost is ignored. This case is identical to equation (1) and simply reflects maximum utility action selection, which is the foundation of most modern decision-making frameworks. Note that the optimal policy $p^*(a|w)$ in this case collapses to a delta over the best action $p^*(a|w) = \delta_{aa^*_w}$. In contrast, $\beta \to 0$ corresponds to an actor that has infinite transformation cost or no computational resources and thus sticks with its prior policy $p_0(a)$. An illustrative example is given in **Figure 1**.

Interestingly, the free energy principle for bounded rational acting can also be used for inference problems. In particular if the utility is chosen as a log-likelihood function $U(w, a) = \log q(w|a)$ and the inverse temperature β is set to one, Bayes' rule is recovered as the optimal bounded rational solution [by plugging into equation (4)]:

$$p^*(a|w) = \frac{p_0(a)q(w|a)}{\sum_a p_0(a)q(w|a)}.$$

Importantly, the inverse temperature β can also be interpreted in terms of computational or sample complexity (Braun and Ortega, 2014; Ortega and Braun, 2014; Ortega et al., 2014).

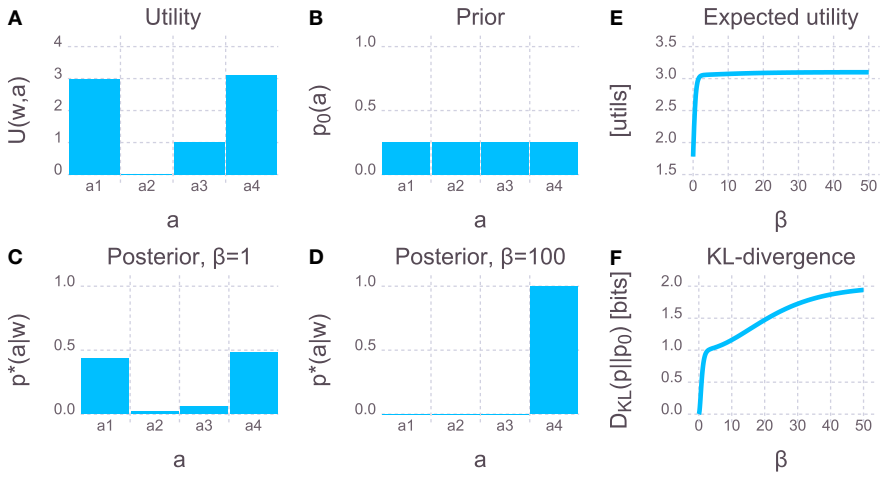

FIGURE 1 | Bounded rationality and the free energy principle. Imagine an actor that has to grasp a particular cup w from a table. There are four options $a1$ to $a4$ to perform the movement and the utility $U(w, a)$ shown in **(A)** measures the performance of each option. There are two actions $a1$ and $a4$ that lead to a successful grasp without spilling, and $a4$ is minimally better. Action $a3$ leads to a successful grasp but spills half the cup, and $a2$ represents an unsuccessful grasp. **(B)** Prior distribution over actions $p_0(a)$: no preference for a particular action. **(C)** Posterior $p^*(a|w)$ [equation (4)] for an actor with limited computational capacity. Due to the computational limits, the posterior cannot deviate from the prior arbitrarily far, otherwise the KL-divergence constraint would be violated. The computational resources are mostly spent on increasing the chance of picking one of the two successful options and decreasing the chance of picking $a2$ or $a3$. The agent is almost indifferent between the two successful options $a1$ and $a4$. **(D)** Posterior for an actor with large computational resources. Even though $a4$ is only slightly better than $a1$, the agent is almost unbounded and can deviate a lot from the prior. This solution is already close to the deterministic maximum expected utility solution and incurs a large KL-divergence from prior to posterior. **(E)** Expected utility $\mathbf{E}_{p*(a|w)}[U(w, a)]$ as a function of the inverse temperature β. Initially, allowing for more computational resources leads to a rapid increase in expected utility. However, this trend quickly flattens out into a regime where small increases in expected utility imply large increases in β. **(F)** KL-divergence D_{KL} $(p^*(a|w)||p_0(a))$ as a function of the inverse temperature β. In order to achieve an expected utility of $\approx$95% of the maximum utility roughly 1 bit suffices [leading to a posterior similar to **(C)**]. Further increasing the performance by 5% requires twice the computational capacity of 2 bits [leading to a posterior similar to **(D)**]. A bounded rational agent that performs reasonably well could thus be designed at half the cost (in terms of computational capacity) compared to a fully rational maximum expected utility agent. An interactive version of this plot where β can be freely changed is provided in the Supplementary Jupyter Notebook "1-FreeEnergyForBoundedRationalDecisionMaking."

The basic idea is that in order to make a decision, the bounded rational decision-maker needs to generate a sample from the posterior $p^*(a|w)$. Assuming that the decision-maker can draw samples from the prior $p_0(a)$, samples from the posterior $p^*(a|w)$ can be generated by rejecting any samples from $p_0(a)$ until one sample is accepted as a sample of $p^*(a|w)$ according to the acceptance rule $u \leq \exp(\beta(U(w,a) - T(w)))$, where u is drawn from the uniform distribution over the unit interval [0;1] and $T(w)$ is the aspiration level or acceptance target value with $T(w) \geq \max_a U(w,a)$. This is known as rejection sampling (Neal, 2003; Bishop, 2006). The efficiency of the rejection sampling process depends on how many samples are needed on average from $p_0(a)$ to obtain one sample from $p^*(a|w)$. This average number of samples $\overline{\#Samples}(w)$ is given by the mean of a geometric distribution

$$\begin{aligned}\overline{\#Samples}(w) &= \frac{1}{\sum_a p_0(a)\exp(\beta(U(w,a) - T(w)))} \\ &= \frac{\exp(\beta T(w))}{Z(w)},\end{aligned} \tag{6}$$

where the partition sum $Z(w)$ is defined as in equation (4). The average number of samples increases exponentially with increasing resource parameter β when $T(w) > \max_a U(w,a)$. It is also noteworthy that the exponential of the Kullback-Leibler divergence provides a lower bound for the required number of samples that is $\overline{\#Samples}(w) \geq \exp\left(D_{\mathrm{KL}}(p^*(a|w)||p_0(a))\right)$ (see Section 6 in the Supplementary Methods for a derivation). Accordingly, a decision-maker with high β can manage high sampling complexity, whereas a decision-maker with low β can only process a few samples.

2.2. From Free Energy to Rate-Distortion: The Optimal Prior

In the free energy principle in equation (3), the prior $p_0(a)$ is assumed to be given. A very interesting question is which prior distribution $p_0(a)$ maximizes the free energy difference $\Delta F(w)$ for all world-states w *on average* (assuming that $p(w)$ is given). To formalize this question, we extend the variational principle in equation (3) by taking the expectation over w and the arg max over $p_0(a)$

$$\begin{aligned}&\arg\max_{p_0(a)} \sum_w p(w) \\ &\quad \times \left[\arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w)\,\middle\|\,p_0(a)\right)\right].\end{aligned}$$

The inner arg max-operator over $p(a|w)$ and the expectation over w can be swapped because the variation is not over $p(w)$. With the KL-term expanded this leads to

$$\begin{aligned}&\arg\max_{p_0(a),p(a|w)} \sum_{w,a} p(w,a)U(w,a) \\ &\quad - \frac{1}{\beta}\sum_w p(w)\sum_a p(a|w)\log\frac{p(a|w)}{p_0(a)}.\end{aligned}$$

The solution to the arg max over $p_0(a)$ is given by $p_0^*(a) = \sum_w p(w)p(a|w) = p(a)$. [see Section 2.1.1 in Tishby et al. (1999) or Csiszár and Tusnády (1984)]. Plugging in the marginal $p(a)$ as the optimal prior $p_0^*(a)$ yields the following variational principle for bounded rational decision-making

$$\begin{aligned}&\arg\max_{p(a|w)} \underbrace{\sum_{w,a} p(w,a)U(w,a)}_{\mathbf{E}_{p(a|w)}[U(w,a)]} - \frac{1}{\beta}\underbrace{\sum_w p(w) D_{\mathrm{KL}}(p(a|w)\,\|p(a))}_{I(W;A)} \\ &\quad = \arg\max_{p(a|w)} J_{\mathrm{RD}}(p(a|w)),\end{aligned} \tag{7}$$

where $I(W;A)$ is the *mutual information* between actions A and world-states W. The mutual information $I(W;A)$ is a measure of the reduction in uncertainty about the action a after having observed w or vice versa since the mutual information is symmetric

$$I(W;A) = H(W) - H(W|A) = H(A) - H(A|W) = I(A;W),$$

where $H(L) = -\Sigma_l p(l)\log p(l)$ is the Shannon entropy of random variable L.

The exact same variational problem can also be obtained as the Langragian for maximizing expected utility with an upper bound on the mutual information

$$p^*(a|w) = \arg\max_{p(a|w)} \sum_{w,a} p(w,a)U(w,a) \quad \text{s.t. } I(W;A) \leq R \tag{8}$$

or in the dual point of view, as minimizing the mutual information between actions and world-states with a lower bound on the expected utility. Thus, the problem in equation (7) is equivalent to the problem formulation in rate-distortion theory (Shannon, 1948; Cover and Thomas, 1991; Tishby et al., 1999; Yeung, 2008), the information-theoretic framework for lossy compression. It deals with the problem that a stream of information must be transmitted over a channel that does not have sufficient capacity to transmit all incoming information – therefore some of the incoming information must be discarded. In rate-distortion theory, the distortion $d(w,a)$ quantifies the recovery error of the output symbol a with respect to the input symbol w. Distortion corresponds to a negative utility which thus leads to an arg min instead of an arg max and a positive sign for the mutual information term in the optimization problem. In this case, a maximum expected utility decision-maker would minimize the expected distortion which is typically achieved by a one-to-one mapping between w and a, which implies that the compression is not lossy. From this, it becomes obvious why MEU decision-making might be problematic: if the MEU decision-maker requires a rate of information processing that is above channel capacity, it simply cannot be realized with the given system.

The solution that extremizes the variational problem of equation (7) is given by the self-consistent equations [see Tishby et al. (1999)]

$$p^*(a|w) = \frac{1}{Z(w)} p(a) e^{\beta U(w,a)}, \tag{9}$$

$$p(a) = \sum_w p(w)p^*(a|w), \tag{10}$$

with *partition sum* $Z(w) = \Sigma_a p(a) e^{\beta U(w,a)}$.

In the limit case $\beta \to \infty$ where transformation costs are ignored, $p^*(a|w) = \delta_{aa^*_w}$ is the perfectly rational policy for each value of w *independent* of any of the other policies and $p(a)$ becomes a mixture of these solutions. Importantly, due to the low price of information processing $\frac{1}{\beta}$, high values of the mutual information term in equation (7) will not lead to a penalization, which means that actions a can be very informative about the world-state w. The behavior of an actor with infinite computational resources will thus in general be very world-state-specific.

In the case where $\beta \to 0$ the mutual information between actions and world-states is minimized to $I(W; A) = 0$, leading to $p^*(a|w) = p(a)\ \forall w$, the maximal abstraction where all w elicit the same response. Within this limitation, the actor will, however, emit actions that maximize the expected utility $\Sigma_{w,a} p(w)p(a)U(w, a)$ using the same policy for all world-states.

For values of the rationality parameter β in between these limit cases, that is $0 < \beta < \infty$, the bounded rational actor trades off *world-state-specific* actions that lead to a higher expected utility for particular world-states (at the cost of an increased information processing rate), against more robust or *abstract* actions that yield a "good" expected utility for many world-states (which allows for a decreased information processing rate).

Note that the solution for the conditional distribution $p^*(a|w)$ in the rate-distortion problem [equation (9)] is the same as the solution in the free energy case of the previous section [equation (4)], except that the prior $p_0(a)$ is now defined as the marginal distribution $p_0(a) = p(a)$ [see equation (10)]. This particular prior distribution minimizes the average relative entropy between $p(a|w)$ and $p(a)$ which is the mutual information between actions and world-states $I(W; A)$.

An alternative interpretation is that the decision-maker is a channel that transmits information from w to a according to $p(a|w)$. The channel has a limited capacity, which could arise from the agent not having a "brain" that is powerful enough, but a limited channel capacity could also arise from noise that is induced into the channel, i.e., an agent with noisy sensors or actuators. For a large capacity, the transmission is not severely influenced and the best action for a particular world-state can be chosen. For smaller capacities, however, some information must be discarded and robust (or abstract) actions that are "good" under a number of world-states must be chosen. This is possible by lowering β until the required rate $I(W; A)$ does no longer exceed the channel capacity. The notion that a decision-maker can be considered as an information processing channel is not new and goes back to the cybernetics movement (Ashby, 1956; Wiener, 1961). Other recent applications of rate-distortion theory to decision-making problems can be found for example in Sims (2003, 2006) and Tishby and Polani (2011).

2.3. Computing the Self-Consistent Solution

The self-consistent solutions that maximize the variational principle in equation (7) can be computed by starting with an initial distribution $p_{\text{init}}(a)$ and then iterating equations (9) and (10) in an alternating fashion. This procedure is well known in the rate-distortion framework as a Blahut-Arimoto-type algorithm (Arimoto, 1972; Blahut, 1972; Yeung, 2008). The iteration is guaranteed to converge to a unique maximum [see Section 2.1.1 in Tishby et al. (1999) and Csiszár and Tusnády (1984) and Cover and Thomas (1991)]. Note that $p_{\text{init}}(a)$ has to have the same support as $p(a)$. Implemented in a straightforward manner, the Blahut-Arimoto iterations can become computationally costly since the iterations involve evaluating the utility function for every action-world-state-pair (w, a) and computing the normalization constant $Z(w)$. In case of continuous-valued random variables, closed-form analytic solutions exist only for special cases. Extending the sampling approach presented at the end of Section 2.1 could be one potential alleviation. A proof-of-concept implementation of the extended sampling scheme is provided in the Supplementary Jupyter Notebook "S1-SampleBasedBlahutArimoto."

2.4. Emergence of Abstractions

The rate-distortion objective for decision-making [equation (7)] penalizes high information processing demand measured in terms of the mutual information between actions and world-states $I(W; A)$. A large mutual information arises when actions are very informative about the world-state which is the case when a particular action is mostly chosen under a particular world-state and is rarely chosen otherwise. Policies $p(a|w)$ with many world-state-specific actions are thus more demanding in terms of informational cost and might not be affordable by an agent with limited computational capacity. In order to keep informational costs low while at the same time optimizing expected utility, actions that yield a "good" expected utility for *many* different world-states must be favored. This leads to abstractions in the sense that the agent does not discriminate between different world-states out of a subset of all world-states, but rather responds with the same policy for the entire subset. Importantly, these abstractions are driven by the agent-environment structure encoded through the utility function $U(w, a)$. Limits in computational resources thus lead to abstractions where different world-states are treated as if they were the same.

To illustrate the influence of different degrees of computational limits and the resulting emergence of abstractions we constructed the following example. The goal is to design a recommender system that observes an item bought w and then recommends another item a. In this example the system can either recommend another concrete item or the best-selling item of a certain category or the best-selling item of a super-category which subsumes several categories (see **Table 1**). An illustration of the example is shown in **Figure 2A**. The possible items bought are shown on the x-axis and possible recommendations are shown on the y-axis. The super-categories and categories as well as the corresponding bought items can be seen in **Table 1** where each bought item also indicates the corresponding concrete item that scores highest when recommended.

The utility of each (w, a)-pair is color-coded in blue in **Figure 2A**. For each possible world-state there is one concrete item that can be recommended that will (deterministically) yield the highest possible utility of 3 utils. Further, each bought item belongs to a category and recommending the best-selling item of the corresponding category leads to a utility of 2.2 utils. Finally, recommending the best-selling item of the corresponding super-category yields a utility of 1.6 utils. For each world-state there is

TABLE 1 | Recommender system example.

Super-category	Category	Bought item	Best recommended item
Electric devices and electronics	Computers	Laptop	Laptop sleeve
		Monitor	Monitor cable
		Game pad	Video game
	Small appliances	Coffee machine	Coffee capsules
		Vacuum cleaner	Vacuum cleaner bags
		Electric toothbrush	Brush heads
Food and cooking	Fruit	Grapes	Cheese
		Strawberries	Cream
		Limes	Cane sugar
	Baking	Pancake mix	Maple syrup
		Baking soda	Vinegar
		Baker's yeast	Flour
		Muffin cups	Flour and chocolate chips

*The system observes an item bought w and can then recommend another item to buy a. For each bought item w, there is one other concrete item a that yields the maximum utility when recommended (indicated in the last column of the table). Additionally, each bought item belongs to a category and a less specific super-category. Recommending the best-selling item of the corresponding category or super-category yields sub-optimal but non-zero utility values. A depiction of the utility function U(w, a) is shown in **Figure 2A**.*

one specific action that leads to the highest possible utility but zero utility for all other world-states. At the same time there exist more abstract actions that are sub-optimal but still "good" for a set of world-states. See the legend of **Figure 2** for more details on the example.

Figure 2B shows the result $p^*(a|w)$ obtained through Blahut-Arimoto iterations [equations (9) and (10)] for $\beta = 1.3$. For each world-state (on the x-axis) the probability over all actions (y-axis) corresponds to one column in the plot and is color-coded in red. For this particular value of β the agent cannot afford to pick the specific actions for most of the world-states (except for the last three world-states) in order to stay within the limit on the maximum allowed rate. Rather, the agent recommends the best-selling items of the corresponding category which allows for a lower rate by having identical policies (i.e., columns in the plot) for sets of world-states. The optimal policies thus lead to abstractions, where several different world-states elicit identical responses of the agent. Importantly, the abstractions are not induced because some stimuli are more similar than others under some utility-free measure and they are also not the result of a *post hoc* aggregation or clustering scheme. Rather, the abstractions are shaped by the utility function and appear as a consequence of bounded rational decision-making in the given task.

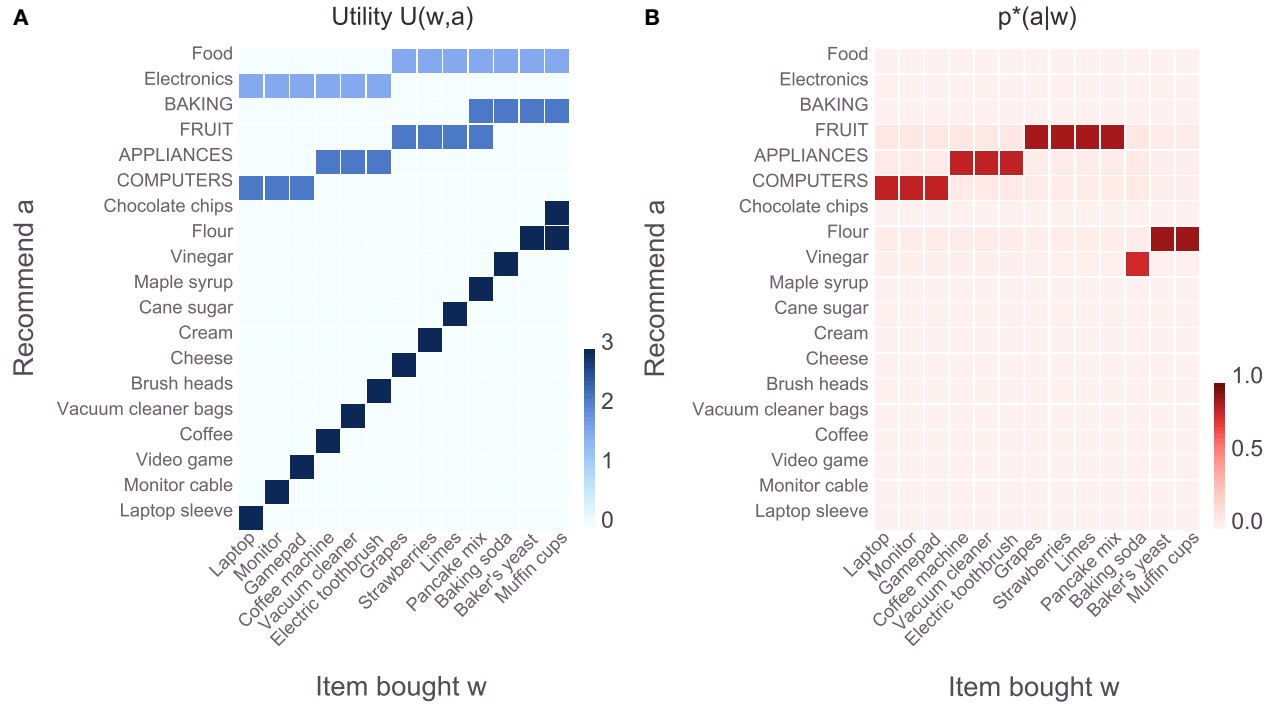

FIGURE 2 | Task setup and solution $p^*(a|w)$ for $\beta = 1.3$. (A) Utility function $U(w, a)$ for the recommender system task. The recommender system observes an item w bought by a customer and recommends another item a to buy to the customer. For each item bought there is another concrete item that has a high chance of being bought by the customer. Therefore, recommending the correct concrete item leads to the maximum utility of 3. However, each item also belongs to a category (indicated by capital letters) and recommending the best-selling item of the corresponding category leads to a utility of 2.2. Finally, each item also belongs to a super-category (either "food" or "electronics") and recommending the best-selling item of the corresponding super-category leads to a utility of 1.6. There is one item (muffin cups) where two concrete items can be recommended and both yield maximum utility. Additionally there is one item (pancake mix) where the recommendation of the best-selling item of both categories "fruit" and "baking" yields the same utility. **(B)** Solution $p^*(a|w)$ [equation (9)] for $\beta = 1.3$. Due to the low β, the computational resources of the recommender system are quite limited and it cannot recommend the highest scoring items (except in the last three columns). Instead, it saves computational effort by applying the same policy to multiple items.

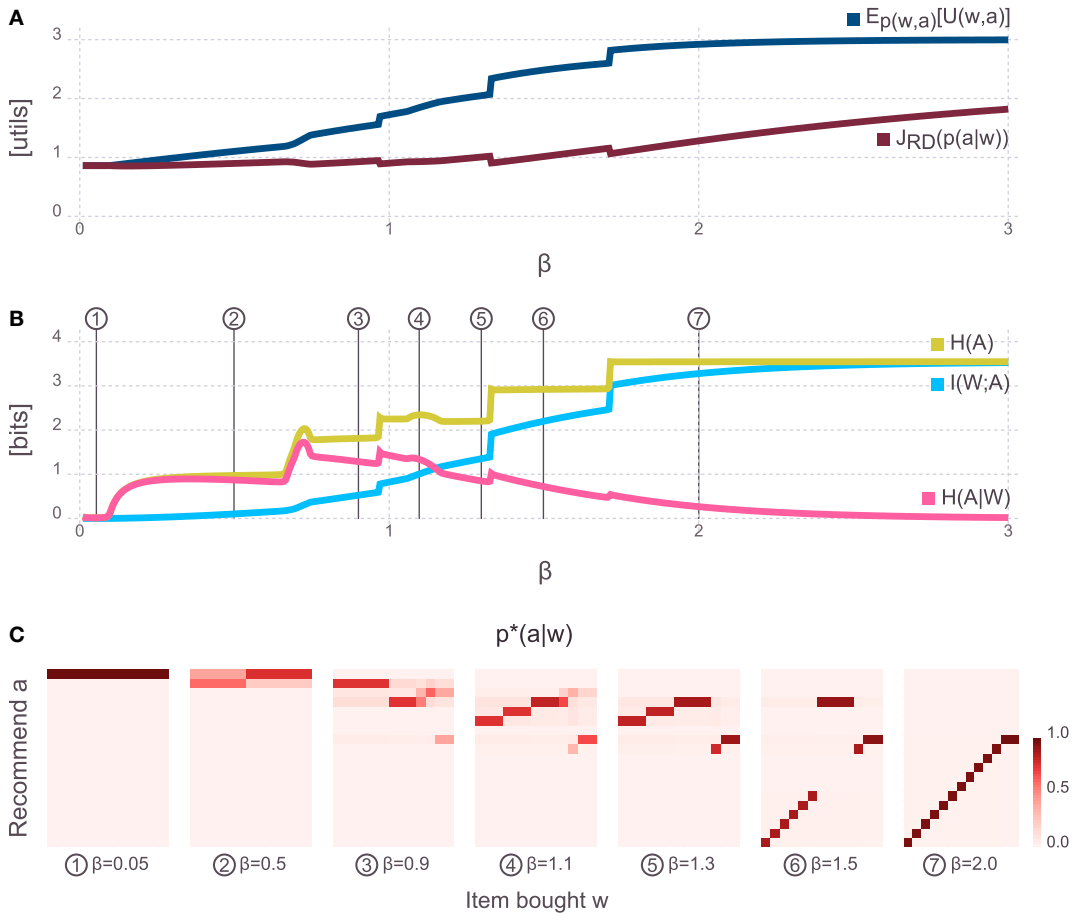

FIGURE 3 | Sweep over β values. (A) Evolution of expected utility and value of the objective $J_{RD}(p(a|w))$ [equation (7)]. The expected utility is monotonically increasing, the objective also increases with increasing β but not monotonically. In both curves, there are some sharp kinks which indicate a change in level of abstraction (see panels below). **(B)** Evolution of information quantities. The same sharp kinks as in the panel above can also be seen in all three curves. Interestingly, $H(A)$ has mostly flat plateaus in between these kinks which means that the behavior of the agent does not change qualitatively along the plateau. At the same time, $H(A|W)$ goes down, indicating that behavior becomes less and less stochastic until a phase transition occurs where $H(A)$ changes. **(C)** Solutions $p^*(a|w)$ [equation (9)] that illustrate the agents behavior on each of the plateaus of $H(A)$. It can be seen that on each plateau (and with each kink), the level of abstraction changes.

Figure 3A shows the expected utility $\mathbf{E}_{p(w,a)}[U(w, a)]$ and the rate-distortion objective $J_{\mathrm{RD}}(p(a|w))$ as a function of the inverse temperature β. The plot shows that by increasing β the expected utility increases monotonically, whereas the objective $J_{\mathrm{RD}}(p(a|w))$ also shows a trend to increase but not monotonically. Interestingly, there are a few sharp transitions at the same points in both curves. The same steep transitions are also found in **Figure 3B,** which shows the mutual information and its decomposition into the entropic terms $I(W, A) = H(A) - H(A|W)$ as a function of β. The line corresponding to the entropy over actions $H(A)$ shows flat plateaus in between these phase transitions. **Figure 3C** illustrates solutions $p^*(a|w)$ for β values corresponding to points on each of the plateaus (labels for bought and recommended items have been omitted for visual compactness but are identical to the plot in **Figure 2B**). Surprisingly, most of the solutions correspond to different levels of abstraction – from fully abstract for $\beta \to 0$, then going through several levels of abstraction and getting more and more specific up to the case $\beta \to \infty$ where the conditional entropy $H(A|W)$ goes to zero implying that the conditionals $p^*(a|w)$ become deterministic and identical to the maximum expected utility solutions. Within a plateau of $H(A)$, the entropy over actions does not change but the conditional entropy $H(A|W)$ tends to decrease with increasing β. This means that qualitatively the behavior along a plateau does not change in the sense that across all world-states the same subset of actions is used. However, the stochasticity within this subset of actions decreases with increasing β (until at some point a phase-transition occurs). Changing the temperature leads to a natural emergence of different levels of abstraction – levels that emerge from the agent-environment interaction structure described by the utility function. Each level of abstraction corresponds to one plateau in $H(A)$.

In general, abstractions are formed by reducing the information content of an entity until it only contains relevant information. For a discrete random variable $w \in \mathcal{W}$, this translates into forming a partitioning over the space $\mathcal{W}$ where "similar" elements are grouped into the same subset of $\mathcal{W}$ and become indistinguishable within the subset. In physics, changing the granularity of a

partitioning to a coarser level is known as *coarse-graining* which reduces the resolution of the space $\mathcal{W}$ in a non-uniform manner. Here, the partitioning emerges in $p^*(a|w)$ as a *soft-partitioning* (see Still and Crutchfield, 2007), where "similar" world-states w get mapped to an action a (or a subset of actions) and essentially become indistinguishable. Readers are encouraged to interactively explore the example in the Supplementary Jupyter Notebook "2-RateDistortionForDecisionMaking."

In analogy to rate-distortion theory where the rate-distortion function serves as an information-theoretic characterization of a system, one can define the *rate-utility function*

$$U(R) = \max_{p(a|w):I(W;A)\leq R} \mathbf{E}_{p(w,a)}[U(w,a)]. \tag{11}$$

where the expected utility is a function of the information processing rate $I(W;A)$. If the decision-maker is conceptualized as a communication channel between world-states and actions, the rate $I(W;A)$ defines the minimally required capacity of that channel. The rate-utility function thus specifies the minimum required capacity for computing actions given a certain expected utility target, or analogously the maximally achievable expected utility given a certain information processing capacity. The rate-utility curve is obtained by varying the inverse temperature β (corresponding to different values of R) and plotting the expected utility as a function of the rate. The resulting plot is shown in **Figure 4**, where the solid line denotes the rate-utility curve and the shaded region corresponds to systems that are theoretically infeasible and cannot be achieved regardless of the implementation. Systems in the white region are sub-optimal, meaning that they could either achieve the same performance with a lower rate or given their limits on computational capacity they could theoretically achieve higher performance. This curve is interesting for both designing systems as well as characterizing the degree of sub-optimality of given systems.

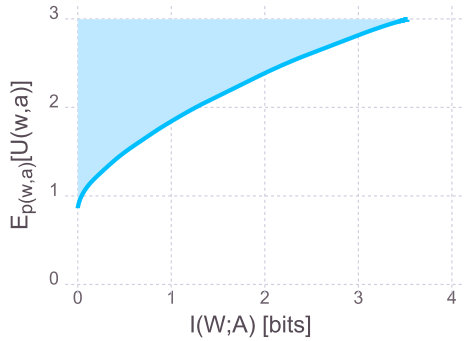

FIGURE 4 | Rate-utility curve. Analogously to the rate-distortion curve in rate-distortion theory, the rate-utility curve shows the minimally required information processing rate to achieve a certain level of expected utility or dually, the maximally achievable expected utility, given a certain rate. Systems that optimally trade-off expected utility against cost of computation lie exactly on the curve. Systems in the shaded region are theoretically impossible and cannot be realized. Systems that lie in the white region of the figure are sub-optimal in the sense that they could achieve a higher expected utility given their computational resources or they could achieve the same expected utility with lower resources.

3. SERIAL INFORMATION-PROCESSING HIERARCHIES

In this section, we apply the rate-distortion principle for decision-making to a serial perception-action system. We design two stages: a perceptual stage $p(x|w)$ that maps world-states w to observations x and an action stage $p(a|x)$ that maps observations x to actions a. Note that the world-state w does not necessarily have to be considered as a latent variable but could in general also be an observation from a previous processing stage. The action stage implements a bounded rational decision-maker (similar to the one presented in the previous section) that optimally trades off expected utility against cost of computation [see equation (7)]. Classically, the perceptual stage might be designed to represent w as faithfully as possible, given the computational limitations of the perceptual stage. Here, we show that trading off expected utility against the cost of information processing on *both* the perceptual and the action stage leads to bounded-optimal perception that does not necessarily represent w as faithfully as possible but rather extracts the most relevant information about w such that the action stage can work most efficiently. As a result, bounded-optimal perception will be tightly coupled to the action stage and will be shaped by the utility function as well as the computational capacity of the action channel.

3.1. Optimal Perception is Shaped by Action

To model a perceptual channel we extend the model from Section 2.2 as follows: The agent is no longer capable of fully observing the state of the world W but using its sensors it is capable to form a percept X as $p(x|w)$ which then allows for adaptation of behavior according to $p(a|x)$. The three random variables for world-state, percept, and action form a serial chain of channels, one channel from world-states to percepts expressed by $p(x|w)$ and another channel from percepts to actions expressed by $p(a|x)$ which implies the following conditional independence

$$p(w,x,a) = p(w)p(x|w)p(a|x),$$

that is also expressed by the graphical model $W \to X \to A$. We assume that $p(w)$ is given and the utility function depends on the world-state and the action $U(w,a)$. Note that mathematically, the results are identical for $U(w,x,a)$, but in this paper we consider the utility independent of the internal percept x.

Classically, inference and decision-making are separated – for instance, by first performing Bayesian inference over the state of the world w using the observation x and then choosing an action a according to the maximum expected utility principle. The MEU action-selection principle can be replaced by a bounded rational model for decision-making that takes into account the computational cost of transforming a (optimal) prior behavior $p_0(a)$ to a posterior behavior $p(a|x)$ as shown in Section 2.

$$\text{Bayesian inference :} \quad p(w|x) = \frac{p(w)p(x|w)}{\sum_w p(w)p(x|w)} \tag{12}$$

$$\text{Bounded rational decision :} \quad p^*(a|x) = \frac{1}{Z(x)} p(a) e^{\beta_2 U(x,a)} \tag{13}$$

where $U(x, a) = \Sigma_w p(w|x)U(w, a)$ is the expectation of the utility under the Bayesian posterior over w given x. Note that the bounded rational decision-maker in equation (13) is identical to the rate-distortion decision-maker introduced in Section 2 that minimizes the trade-off given by equation (7) by implementing equation (9). It includes the MEU solution as a special case for $\beta_2 \to \infty$. Here, the inverse temperature is denoted by β_2 (instead of β as in the previous section) for notational reasons that ensure consistency with later results of this section.

In equation (12), the choice of the likelihood model $p(x|w)$ remains unspecified and the question is where does it come from? In general, it is chosen by the designer of a system and the choice is often driven by bandwidth or memory constraints. In purely descriptive scenarios, the likelihood model is determined by the sensory setup of a given system and $p(x|w)$ is obtained by fitting it to data of the real system. In the following, we present a particular choice of $p(x|w)$ that is fundamentally grounded on the principle that any transformation of behavior or beliefs is costly (which is identical to the assumption of limited-rate information processing channels) and this cost should be traded off against gains in expected utility. Remarkably, equations (12) and (13) drop out naturally from the principle.

Given the graphical model: $W \to X \to A$, we consider an information processing channel between W and X and another one between X and A and introduce different rate-limits on these channels, i.e., the information processing price on the perceptual level $\frac{1}{\beta_1}$ can be different from the price of information processing on the action level $\frac{1}{\beta_2}$. Formally, we set up the following variational problem:

$$\begin{aligned} &\arg\max_{p(x|w),p(a|x)} \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A) \\ &\quad = \arg\max_{p(x|w),p(a|x)} J_{\text{ser}}(p(x|w), p(a|x)). \end{aligned} \tag{14}$$

Similar to the rate-distortion case, the solution is given by the following set of four self-consistent equations:

$$p^*(x|w) = \frac{1}{Z(w)} p(x) exp\left(\beta_1 \Delta F_{\text{ser}(w,x)}\right) \tag{15}$$

$$p(x) = \sum_w p(w) p^*(x|w) \tag{16}$$

$$p^*(a|x) = \frac{1}{Z(x)} p(a) exp\left(\beta_2 \sum_w p(w|x) U(w,a)\right) \tag{17}$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|x), \tag{18}$$

where $Z(w)$ and $Z(x)$ denote the corresponding normalization constants or partition sums. The conditional probability $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{ser}}(w, x)$ is the free energy difference of the action stage:

$$\Delta F_{\text{ser}(w,x)} := \mathbf{E}_{p^*(a|x)}[U(w,a)] - \frac{1}{\beta_2} D_{\text{KL}}(p^*(a|x)||p(a)), \tag{19}$$

see also equation (5). More details on the derivation of the solution equations can be found in the Supplementary Methods Section 2.

The bounded-optimal perceptual model is given by equation (15). It follows the typical structure of a bounded rational solution consisting of a prior times the exponential of the utility multiplied by the inverse temperature. Compare equation (9) to see that the downstream free-energy trade-off $\Delta F_{\text{ser}(w,x)}$ now plays the role of the utility function for the perceptual model. The distribution $p^*(x|w)$ thus optimizes the downstream free-energy difference in a bounded rational fashion, that is taking into account the computational resources of the perceptual channel. Therefore, the optimal percept becomes tightly coupled to the agent-environment interaction structure as described by the utility function or in other words: the optimal percept is shaped by the embodiment of the agent and, importantly, is not simply a maximally faithful representation of W through X given the limited rate of the perceptual channel. A second interesting observation is that the action stage given by equation (17) turns out to be a bounded rational decision-maker using the Bayesian posterior $p(w|x)$ for inferring the true world-state w given the observation x. This is identical to equation (13) (using the optimal prior $p(a) = \Sigma_{w,x} p(w)p^*(x|w)p^*(a|x)$), even though the latter was explicitly modeled by first performing Bayesian inference over the world-state w given the percept x [equation (12)] and then performing bounded rational decision-making [equation (13)], whereas the same principle drops out naturally in equation (17) as a result of optimizing equation (14).

3.2. Illustrative Example

In this section, we design a hand-crafted perceptual model $p_\lambda(x|w)$ with precision-parameter λ, that drives a subsequent bounded rational decision-maker that maps an observation x to a distribution over actions $p(a|x)$ in order to maximize expected utility while not exceeding a constraint on the rate of the action channel. The latter is implemented by following equation (13) and setting β_2 according to the limit on the rate $I(X; A)$. We compare the bounded rational actor with hand-crafted perception against a bounded-optimal actor that maximizes equation (14) by implementing the four corresponding self-consistent equations (15)–(18). Importantly, the perceptual model $p^*(x|w)$ of the bounded-optimal actor maximizes the downstream free-energy trade-off of the action stage $\Delta F_{\text{ser}(w,x)}$ which leads to a tight coupling between perception and action that is not present in the hand-crafted model of perception. The action stage is identical in both models and given by equation (17).

We designed the following example where the actor is an animal in a predator-prey scenario. The actor has sensors to detect the size of other animals it encounters. In this simplified scenario, animals can only belong to one of three size-groups and their size correlates with their hearing-abilities:

- Small animals (insects): either 2, 3, or 4 size-units cannot hear very well.
- Medium-sized animals (rodents): either 6, 7, or 8 size-units can hear quite well.
- Large animals (cats of prey): either 10, 11, or 12 size-units can hear quite well.

The actor has a sensor for detecting the size of an animal, however, depending on the capacity of the perceptual channel this sensor will either be more or less noisy. To survive, the actor can

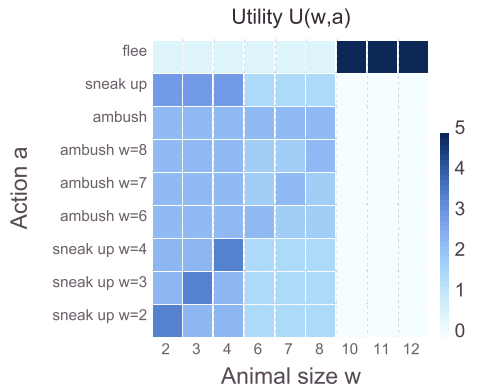

FIGURE 5 | Predator-prey utility function *U(w, a)*. The actor encounters different groups of animals – small, medium-sized, and large – and should hunt small and medium-sized animals for food and flee from large animals. Animals within a group have a similar size *w*: small animals ∈ {2,3,4} size-units, medium-sized animals ∈ {6,7,8} size-units, and large animals ∈ {10,11,12} size-units. For hunting – denoted by the action *a* – the actor can either ambush the other animal or alternatively sneak up to the other animal and strike once close enough. In general, sneaking up works well on small animals as they cannot hear very well, whereas medium-sized animals can hear quite well which decreases the chance of a successful sneak-up. The ambush works equally well on both groups of prey. Large animals should never be engaged, but the actor should flee in order to avoid falling prey to them. In addition to the generic hunting patterns that work equally well for all members of a group, for small animals there are specific sneak-up patterns that have an increased chance of success if applied to the correct animal but a decreased chance of success when applied to the wrong animal. For medium-sized animals, there are specific ambush behaviors that have the same chance of success as the generic ambush behavior when applied to the correct animal but a decreased chance of success when applied to the wrong animal. Since the specific ambush behaviors for the medium-sized animals have no advantage over the generic ambush behavior but require more information processing, any bounded rational actor should never choose the specific ambush behaviors. This somewhat artificial construction is meant to illustrate that bounded rational actors do not add unnecessary complexity to their behavior. See main text for more details on the example.

hunt animals from both the small and the medium-sized group for food. On the other hand, it can fall prey to animals of the large group. The actor has three basic actions:

- Ambush: steadily wait for the other animal to get close and then strike.
- Sneak-up: slowly move closer to the animal and then strike.
- Flee: quickly move away from the other animal.

The advantage of the ambush is that it is silent, however, the risk is that the animal might not move toward the position of the ambush – it works equally well on animals from the small and medium-sized group. The sneak-up is not silent but does not rely on the other animal coincidentally getting closer – it works better than the ambush for small-sized animals but the opposite is true for medium-sized animals. If the actor encounters a large animal the only sensible action is to flee in order to avoid falling prey to the large animal. Besides these generic actions, the actor also has a repertoire of more specific hunting patterns – see **Figure 5** which shows the full details of the utility function for the predator-prey scenario. The exact numeric values are found in the Supplementary Jupyter Notebook "3-SerialHierarchy."

The hand-crafted model of perception is specified by $p_\lambda(x|w)$, where the observed size x corresponds to the actual size of the animal w corrupted by noise. The precision-parameter λ governs the noise-level and thus the quality of the perceptual channel which can be measured with $I(W;X)$. In particular, the observation o is a discretized noisy version of w with precision λ:

$$x|w,\lambda \sim \text{round}(\mathcal{N}_{\text{trunc}}(w, 1/\lambda)), \tag{20}$$

where the set of world-states is given by all possible animal sizes $w \in \mathcal{W} = \{2,3,4,6,7,8,10,11,12\}$ and the set of possible observations is given by $x \in \mathcal{X} = \{1,2,3,\ldots,11,12,13\}$. To avoid a boundary-bias due to the limited interval $\mathcal{X}$ we reject and re-sample all values of x that would fall outside of $\mathcal{X}$. For $\lambda \to \infty$, the perceptual channel is very precise, and there is no uncertainty about the true value of w after observing x. However, such a channel incurs a large computational effort as the mutual information $I(W;X)$ is maximal in this case. If the perceptual channel has a smaller capacity than required to uniquely map each w to an x, the rate must be reduced by lowering the precision λ. Medium precision will mostly lead to within-group confusion whereas low precision will also lead to across-group confusion and corresponds to perceptual channels with a very low rate $I(W;X)$.

The results in **Figure 6** show solutions when having large computational resources on both the perception and action channel. As the figure clearly shows, the hand-crafted model $p_\lambda(x|w)$ looks quite different from the bounded-optimal solution $p^*(x|w)$, even though the rate on the perceptual channel is identical in both cases (given by the mutual information $I(W;X) \approx 2$ bits). The difference is that the bounded-optimal percept spends the two bits mainly on discriminating between specific animals of the small group and on discriminating between medium-sized and large animals. It does not discriminate between specific sizes within the latter two groups. This makes sense, as there is no gain in utility by applying any specific actions to specific animals in the medium- or large-sized group. **Figure 6** also shows the overall-behavior from the point of view of an external observer $p(a|w)$, which is computed as follows

$$p_\lambda(a|w) = \sum_x p_\lambda(x|w)p_\lambda(a|x) \quad \text{and}$$
$$p^*(a|w) = \sum_x p^*(x|w)p^*(a|x) \tag{21}$$

The overall-behavior in the bounded-optimal case is more deterministic, leading to a higher expected utility in the bounded-optimal case. The distributions $p_\lambda(a|x)$ and $p^*(a|w)$ are not shown in the figure but can easily be inspected in the Supplementary Jupyter Notebook "3-SerialHierarchy." If the price of information processing on the perceptual channel in the hand-crafted model is the same as in the bounded-optimal model (given by β_1), then the overall objective $J_{\text{ser}}(p(x|w), p(a|x))$ is larger for the bounded-optimal case compared to the hand-crafted case, implying that the bounded optimal actor achieves a better trade-off between expected utility and computational cost. The crucial insight of this example is that the optimal percept depends on the utility function, where in this particular case it does for instance make no sense to waste computational resources on discriminating

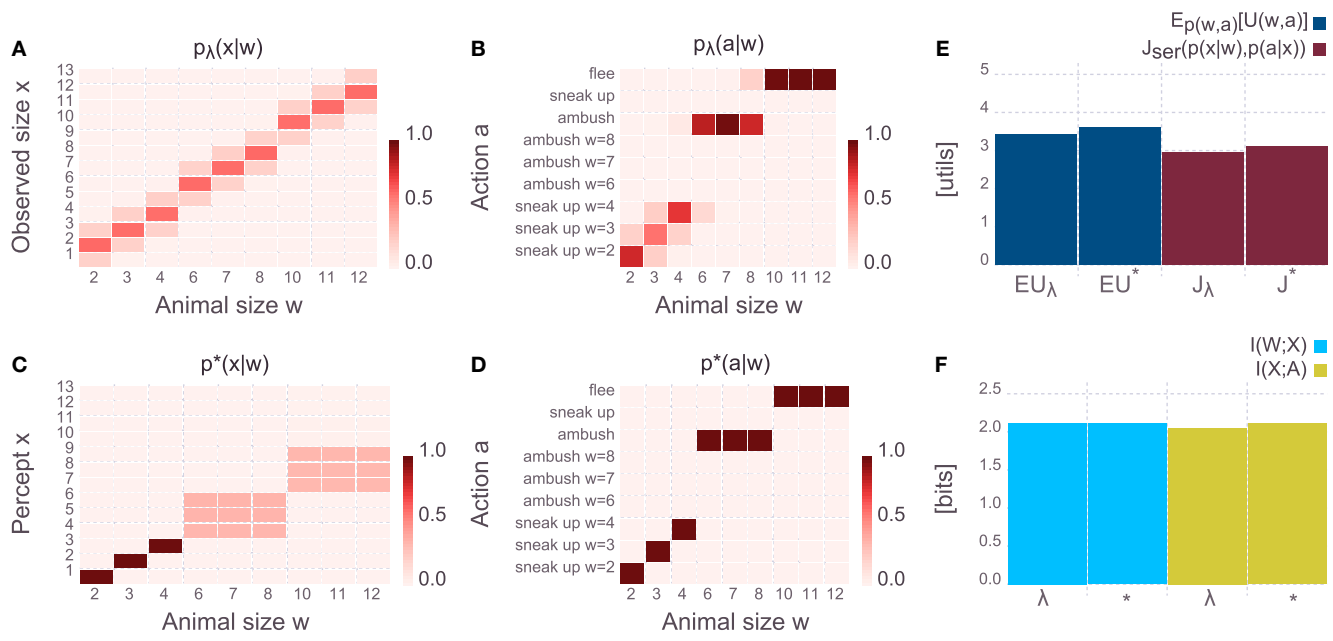

FIGURE 6 | Comparing hand-crafted perception against bounded-optimal perception (parameters are given above the panels). Both the perceptual and the action channel have large computational resources (λ or β_1 for the perceptual channel, respectively, and β_2 for the action channel). **(A)** Hand-crafted perceptual model $p_\lambda(x|w)$ [equation (20)] – the observed animal size is a noisy version of the actual size of the animal, but the noise-level is quite low (due to large λ). **(B)** Overall-behavior of the system $p_\lambda(a|w)$ [equation (21)] with the hand-crafted perception. For small animals, the specific sneak-up patterns are used, but the noisiness of the perception does not allow to deterministically pick the correct pattern. **(C)** Bounded-optimal perceptual model $p^*(x|w)$ [equation (15)] – the bounded-optimal model does not distinguish between the individual medium- or large-sized animals as all of them are treated with the respective generic action. Instead, the computational resources are spent on precisely discriminating between the different small animals. Note that this bounded-optimal perceptual model spends exactly the same computation as the hand-crafted model [see **(F)**, blue bars]. **(D)** Overall-behavior $p^*(a|w)$ [equation (21)] of the bounded-optimal system. Compared to the hand-crafted system, the bounded-optimal perception allows to precisely discriminate between the different small animals which is also reflected in the overall-behavior. The bounded-optimal system thus achieves a higher expected utility **(E)**. **(E)** Comparison of expected utility (dark blue, first two columns) and value of the objective [equation (14), dark red, last two columns] for the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). Since both systems spend roughly the same amount of computation [see **(F)**], but the bounded-optimal model achieves a higher expected utility, the overall objective is in favor of the bounded-optimal system. **(F)** Comparison of information processing effort in the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). Both systems spend (roughly) the same amount of computation, however, the bounded-optimal system spends it in a more clever way – compare **(A,C)**. The distributions $p_\lambda(a|x)$ and $p^*(a|w)$ are not shown in the figure but can easily be inspected in the Supplementary Jupyter Notebook "3-SerialHierarchy."

between the specific animals of the large group because the optimal response (flee with certainty) is identical to all of them. In the Supplementary Jupyter Notebook"3-SerialHierarchy" the utility function can easily be switched while keeping all other parameters identical in order to observe how the bounded-optimal percept changes accordingly. Note that the bounded-optimal behavior $p^*(a|w)$ shown in **Figure 6D** yields the highest possible expected utility in this task setup – there is no behavior that would lead to a higher expected utility (though there are other solutions that lead to the same expected utility).

The bounded-optimal percept depends not only on the utility function but also on the behavioral richness of the actor which is governed by the rate on the action channel $I(X; A)$. In **Figure 7** we show the results of the same setup as in **Figure 6** with the only change being the significantly increased price for information processing in the action stage (as specified by $\beta_2 = 1$ bit per util whereas it used to be $\beta_2 = 10$ bits per util in the previous figure). The hand-crafted perceptual model is unaffected by this change of the action stage, but the bounded-optimal model of perception has changed compared to the previous figure and now reflects the limited behavioral richness. As shown in $p^*(a|w)$ in **Figure 7**, the actor is no longer capable of applying different actions to animals of the small group and animals of the medium-sized group. Accordingly, the bounded-optimal percept does not waste computational resources for discriminating between small and medium-sized animals since the downstream policy is identical for both groups of animals. In terms of expected utility, both the hand-crafted model as well as the bounded-optimal decision-maker score equally at ≈3 utils. However, the bounded-optimal model does so by using lower computational resources and thus scoring better on the overall trade-off $J_{\text{ser}}(p(x|w), p(a|x))$.

In **Figure 8** we again use large resources on the action channel $\beta_2 = 10$ (as in the first example in **Figure 6**), but now the resources on the perceptual channel are limited by setting $\beta_1 = 1$ (compared to $\beta_1 = 8$ in the first case). Accordingly, the precision of the hand-crafted perceptual model is tuned to $\lambda = 0.4$ (compared to $\lambda = 1.65$ in the first case) such that it has the same rate $I(W; X)$ as the bounded-optimal model. By comparing the two panels for $p_\lambda(x|w)$ and $p^*(x|w)$, it can clearly be seen that the bounded-optimal perceptual model now spends its scarce

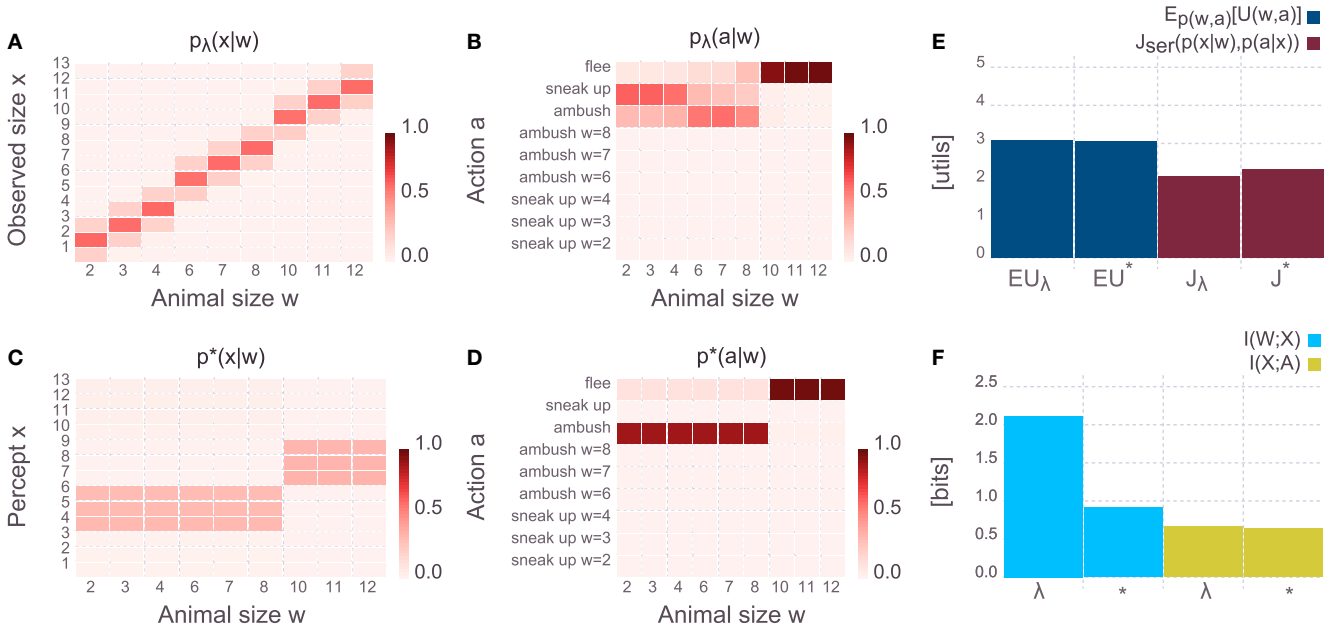

FIGURE 7 | Comparing hand-crafted perception against bounded-optimal perception (parameters are given above the panels). Compared to **Figure 6**, the action channel of both the hand-crafted and the bounded-optimal system now has low computational resources (through the low β_2 for the action channel). **(A)** Hand-crafted perceptual model $p_\lambda(x|w)$ [equation (20)] – the hand-crafted perception is unaffected by the increased limit in computational resources on the action channel. **(B)** Overall-behavior of the system $p_\lambda(a|w)$ [equation (21)] with the hand-crafted perception. Even though perception was unaffected by the increased limit on the action channel, the overall-behavior is severely affected. **(C)** Bounded-optimal perceptual model $p^*(x|w)$ [equation (15)] – due to the low information processing rate on the action channel [see **(F)**, yellow bars], the bounded-optimal perception adjusts accordingly and only discriminates between predator and prey animals. **(D)** Overall-behavior $p^*(a|w)$ [equation (21)] of the bounded-optimal system. Compared to the hand-crafted system, the bounded-optimal system distinguishes sharply between predator and prey animals. Note, however, that both systems achieve an identical expected utility [**(E)**, dark blue bars]. **(E)** Comparison of expected utility (dark blue, first two columns) and value of the objective [equation (14), dark red, last two columns] for the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). Both systems achieve the same expected utility, but the bounded-optimal system does so with requiring less computation on the perceptual channel [see **(F)**, blue bars]. **(F)** Comparison of information processing effort in the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). Due to the low computational resources on the action channel, the corresponding information processing rate (yellow bars) is quite low. The bounded-optimal perception adjusts accordingly and requires a much lower rate compared to the hand-crafted model of perception that does not adjust. The distributions $p_\lambda(a|x)$ and $p^*(a|x)$ are not shown in the figure but can easily be inspected in the Supplementary Jupyter Notebook "3-SerialHierarchy."

resources to reliably discriminate between large animals and all other animals. The overall behavioral policies $p(a|w)$ reflect the limited perceptual capacity in both cases, however, the bounded-optimal case scores a higher expected utility of ≈3 utils compared to the hand-crafted case. The overall objective $J_{\text{ser}}(p(x|w), p(a|x))$ is also higher for the bounded-optimal model, indicating that this model should be preferred because it finds a better trade-off between expected utility and information processing cost.

Note that in all three examples the optimal percept $p^*(x|w)$ often leads to a uniform mapping of an exclusive subset of world-states w to the same set of percepts x. Importantly, these percepts do not directly correspond to an observed animal size as in the case of the hand-crafted model of perception. Rather, the optimal percepts often encode more abstract concepts such as medium- or large-sized animal (as in **Figure 6**) or predator and prey animal (as in **Figures 7** and **8**). In a sense, abstractions similar to the ones shown in the recommender system example in the previous section (**Figure 3**) emerge in the predator-prey example as well but now they also manifest themselves in the form of abstract percepts. Crucially, the abstract percepts allow for more efficient information processing further downstream in the decision-making part of the system. The formation of these abstract percepts is driven by the embodiment of the agent and reflects certain aspects of the utility function of the agent. For instance, unlike the actor in **Figure 6**, the actors in **Figures 7** and **8** would not "understand" the concept of medium-sized animals as it is of no use to them: with their very limited resources it is most important for them to have the two perceptual concepts of predator and prey. Note that the cardinality of X in the bounded-optimal model of perception is fixed in all examples in order to allow for easy comparison against the hand-crafted model, but it could be reduced further without any consequences (up to a certain point) – this can be explored in the Supplementary Jupyter Notebook "3-SerialHierarchy."

The solutions shown in this section were obtained by iterating the self-consistent equations until numerical convergence. Since there is no convergence-proof, it cannot be fully ruled out that the solutions are sub-optimal with respect to the objective. However, the point of the simulation results shown here is to allow for easier interpretation of the theoretical results and highlight

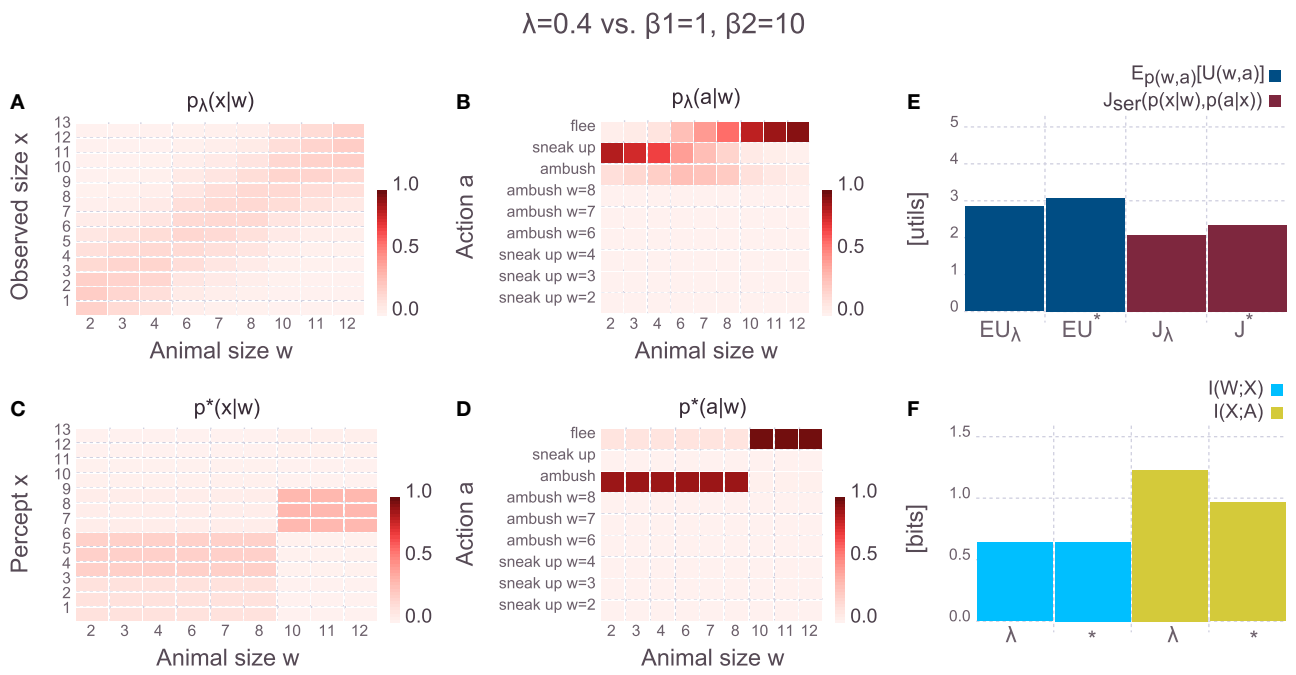

FIGURE 8 | Comparing hand-crafted perception against bounded-optimal perception (parameters are given above the panels). Compared to **Figure 6**, the perceptual channel of both the hand-crafted and the bounded-optimal system now has low computational resources (through the low λ and β_1 respectively). **(A)** Hand-crafted perceptual model $p_\lambda(x|w)$ [equation (20)] – due to the low precision λ, the noise for the hand-crafted perception has increased dramatically. **(B)** Overall-behavior of the system $p_\lambda(a|w)$ [equation (21)] with the hand-crafted perception. Even though the action-part of the system still has large computational resources the overall-behavior is severely affected due to the bad perceptual channel. **(C)** Bounded-optimal perceptual model $p^*(x|w)$ [equation (15)] – due to the low computational resources on the perceptual channel (β_1) the system can only distinguish between predator and prey animal (which is the most important information). Note how the percept in the bounded-optimal case no longer corresponds to some observed animal size but rather to a more abstract concept such as predator or prey animal. The parameters were chosen such that both the hand-crafted and the bounded-optimal system spend the same amount of information processing on the perceptual channel [see **(F)**, blue bars]. **(D)** Overall-behavior $p^*(a|w)$ [equation (21)] of the bounded-optimal system. Compared to the hand-crafted system, the bounded-optimal system distinguishes sharply between predator and prey animals. This is only possible because the bounded-optimal perception spends its scarce resource to exactly perform this distinction. **(E)** Comparison of expected utility (dark blue, first two columns) and value of the objective [equation (14), dark red, last two columns] for the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). The bounded-optimal system achieves a higher expected utility even though the amount of information processing is not larger compared to the hand-crafted model [see **(F)**]. Rather, the bounded-optimal model spends its scarce resources more optimally. **(F)** Comparison of information processing effort in the hand-crafted system (denoted by λ) and the bounded-optimal system (denoted by *). Note that both perceptual channels (blue bars) require the same information processing rate, but the bounded-optimal perceptual model processes the more important information (predator versus prey) which allows the subsequent action channel to perform better [see **(E)**, dark blue bars). The distributions $p_\lambda(a|x)$ and $p^*(a|x)$ are not shown in the figure but can easily be inspected in the Supplementary Jupyter Notebook "3-SerialHierarchy."

certain aspects of the theoretical findings. We discuss this issue in Section 5.2.

4. PARALLEL INFORMATION-PROCESSING HIERARCHIES

Rational decision-making requires searching through a set of alternatives a and picking the option with the highest expected utility. Bounded rational decision-making replaces the "hard maximum" operation with a soft selection mechanism where the first action that satisfies a certain level of expected utility is picked. A parallel hierarchical architecture allows for a prior partitioning of the search space which reduces the effective size of the search space and thus speeds up the search process. For instance, consider a medical system that consists of general practitioners and specialist doctors. The general doctor can restrict the search space for a particular ailment of a patient by determining which specialist the patient should see. The specialist doctor in turn can determine the exact disease. This leads to a two-level decision-making hierarchy consisting of a high-level partitioning that allows for making a subsequent low-level decision with reduced (search) effort. In statistics, the partitioning that is induced by the high-level decision is often referred to as a *model* and is commonly expressed as a probability distribution over the search space $p(a|m)$ (where m indicates the model) which also allows for a soft-partitioning. The advantage of hierarchical architectures is that the computation that leads to the high-level reduction of the search space can be stored in the model (or in a set of parameters in case of a parametric model). This computation can later be re-used by using the correct model (or set of parameters) in order to perform the low-level computation more efficiently. Interestingly, it should be most economic to put the most re-usable, and thus more abstract, information into the models $p(a|m)$ which leads to a hierarchy of abstractions. However, in order to make sure that the correct model is used, another deliberation process $p(m|w)$ is required (where w indicates the observed stimulus or data). Another problem is how to chose the partitioning to be most effective. In this section, we address both problems from a bounded rational point

of view. We show that the bounded optimal solution $p^*(a|m)$ trades off the computational cost for choosing a model m against the reduction in computational cost for the low-level decision.

To keep the notation consistent across all sections of the paper we denote the model m in the rest of the paper with the variable x. This is in contrast to Section 3, where x played the role of a percept. The advantage of this notation is that it allows to easily see similarities and differences of the information terms and solution equations of the different cases. In particular, in Section 5 we present a unifying case that includes the serial and parallel case as special cases – by keeping the notation consistent this can easily be seen.

4.1. Optimal Partitioning of the Search Space

Constructing a two-level decision-making hierarchy requires the following three components: high-level models $p(a|x)$, a model selection mechanism $p(x|w)$ and a low-level decision maker $p(a|w, x)$ (w denotes the observed world-state, x indicates a particular model and a is an action). The first two distributions are free to be chosen by the designer of the system, for $p(a|w, x)$ a maximum expected utility decision-maker is the optimal choice if computational costs are neglected. Here, we take computational cost into account and replace the MEU decision-maker with a bounded rational decision-maker that includes MEU as a special case ($\beta_3 \to \infty$) – the bounded rational decision-maker optimizes equation (7) by implementing equation (9). In the following we show how all parts of the hierarchical architecture:

1. Selection of model (or expert): $p(x|w)$ (22)
2. Prior knowledge of model (or expert): $p(a|x)$ (23)
3. Bounded rational decision of model (or expert):

$$p^*(a|w,x) = \frac{1}{Z(w,x)} p(a|x) e^{\beta_3 U(w,a)} \tag{24}$$

emerge from optimally trading off computational cost against gains in utility. Importantly, $p(a|x)$ plays the role of a prior distribution for the bounded rational decision-maker and reflects the high-level partitioning of the search space.

The optimization principle that leads to the bounded-optimal hierarchy trades off expected utility against the computational cost of model selection $I(W; X)$ and the cost of the low-level decision using the model as a prior $I(W; A|X)$:

$$\begin{aligned} \underset{p(x|w),p(a|w,x)}{\arg\max} \; & \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_3} I(W;A|X) \\ &= \underset{p(x|w),p(a|w,x)}{\arg\max} \; J_{\text{par}}(p(x|w), p(a|w,x)). \end{aligned} \tag{25}$$

The set of self-consistent solutions is given by

$$p^*(x|w) = \frac{1}{Z(w)} p(x) exp\left(\beta_1 \Delta F_{\text{par}(w,x)}\right) \tag{26}$$

$$p(x) = \sum_w p(w) p^*(x|w) \tag{27}$$

$$p^*(a|w,x) = \frac{1}{Z(w,x)} p^*(a|x) exp\left(\beta_3 U(w,a)\right) \tag{28}$$

$$p^*(a|x) = \sum_w p(w|x) p^*(a|w,x), \tag{29}$$

where $Z(w)$ and $Z(w, x)$ denote the corresponding normalization constants or partition sums. $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{par}(w,x)}$ is the free energy difference of the low-level stage:

$$\Delta F_{\text{par}}(w,x) := \mathbf{E}_{p^*(a|w,x)}[U(w,a)] - \frac{1}{\beta_3} D_{\text{KL}}(p^*(a|w,x)||p^*(a|x)), \tag{30}$$

see equation (5). More details on the derivation of the solution equations can be found in the Supplementary Methods Section 3. By comparing the solution equations (26)–(29) with equations (22)–(24) the hierarchical structure of the bounded-optimal solution can be seen clearly. The bounded-optimal model selector in equation (26) maximizes the downstream free-energy trade-off $\Delta F_{\text{par}(w,x)}$ in a bounded rational fashion and is similar to the optimal perceptual model of the serial case [equation (15)]. This means that the optimal model selection mechanism is shaped by the utility function as well as the computational process on the low-level stage of the hierarchy (governed by β_3) but also by the computational cost of model selection (governed by β_1). The optimal low-level decision-maker given by equation (28) turns out to be exactly a bounded rational decision-maker with $p(a|x)$ as a prior – identical to the low-level decision-maker that was motivated in equation (24). Importantly, the bounded-optimal solution provides a principled way of designing the models $p(a|x)$ [see equation (29)]. According to the equation, the optimal model $p^*(a|x)$ is given by a Bayesian mixture over optimal solutions $p^*(a|w, x)$ where w is known. The Bayesian mixture turns out to be the optimal compressor of actions for unknown w under the belief $p(w|x)$.

4.2. Illustrative Example

To illustrate the formation of bounded-optimal models, we designed the following example: in a simplified environment, only three diseases can occur – a heart disease or one of two possible lung diseases. Each of the diseases comes in two possible types (e.g., type 1 or type 2 diabetes). Depending on how much information is available on the symptoms of a patient, diseases can be treated according to the specific type (which is most effective) or with respect to the disease category (which is less effective but requires less information). See **Figure 9** for a plot of the utility function and a detailed description of the example. The goal is to design a medical analysis hierarchy that initiates the best possible treatment, given its limitations. The hierarchy consists of an automated medical system that can cheaply take standard measurements to partially assess a patient's disease category. Additionally, the patient is then sent to a specialist who can manually perform more elaborate measurements if necessary to further narrow down the patient's precise disease type and recommend a treatment. The automated system should be designed in a way that minimizes the additional measurements required by the specialists. More formally, the automated system delivers a first diagnosis x given the patient's precise disease type w according to $p(x|w)$. The first diagnosis narrows down the possible treatments a according to a model $p(a|x)$. For each x, a specialist can further reduce uncertainty about the correct treatment by performing more measurements $p(a|w, x)$. We compare the optimal design of the automated system and the corresponding optimal

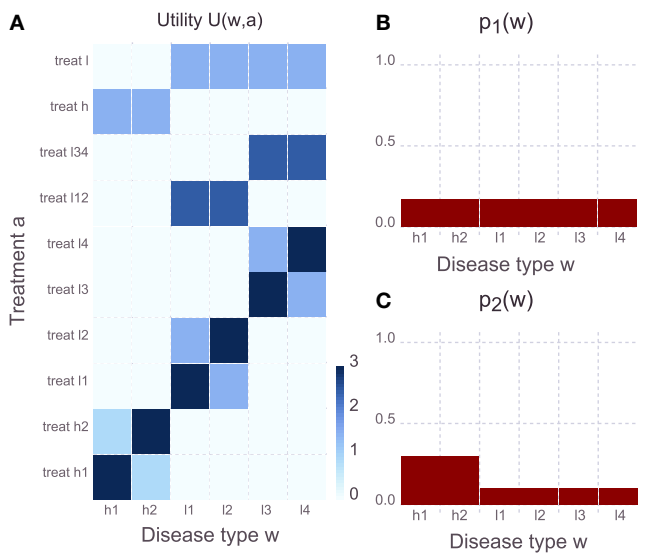

FIGURE 9 | Task setup of the medical system example. **(A)** Utility function. The disease category heart disease comes in the two types $h1$ or $h2$. There are two different categories of lung diseases that come in types $l1$ or $l2$ (lung disease A) or types $l3$ and $l4$ (lung disease B). The goal of the medical system is to gather information about the disease type and then initiate a treatment. Each of the disease types is best treated with a specific treatment. However, when the specific treatment is applied to the wrong disease type of the same disease category it is less effective. Additionally there are general treatments for the heart- and the two lung diseases that are slightly worse than the specific treatments but can be applied with less knowledge about the disease type. For lung diseases, there is a general treatment that works even if the specific lung disease is not known; however, it is less effective compared to the treatments where the disease but not the precise type is known. **(B)** Distribution of disease types in environment 1 – all diseases appear with equal probability. **(C)** Distribution of disease types in environment 2 where the heart diseases have an increased probability and the two corresponding disease types appear with higher chance.

treatment recommendations to the specialist $p^\star(a|x)$ according to equation (29) in two different environments: one, where all disease types occur with equal probability (**Figure 9B**) versus two, where heart diseases occur with increased chance (**Figure 9C**). For this example, the number of different high-level diagnoses X is set to $|\mathcal{X}| = 3$ which also means that there can be three different treatment recommendations $p(a|x)$. Since in the example the total budget for performing measurements is quite low (reflected by β_1, β_3 both being quite low), the whole system (automated plus specialists) can in general not gather enough information about the symptoms to treat every disease type with the correct specific treatment. Rather, the low budget has to be spent on gathering the most important information.

Figure 10 shows bounded-optimal hierarchies for the medical system in both environments. The top row in **Figure 10** shows the optimal hierarchy for the environment where all diseases appear with equal probability: the automated system $p^\star(x|w)$ (see **Figure 10A**) distinguishes between a heart disease, lung disease A and lung disease B, which means that there is one treatment recommendation for heart diseases and one treatment recommendation for each of the two possible lung diseases respectively (see the three columns of $p^\star(a|x)$ in **Figure 10B**). Since the general treatment for the heart disease works less effective than the general treatments for the two lung diseases, the (very limited) budget of the specialists is completely spent on finding the correct specific heart treatment. Both lung diseases are treated with their respective general treatments since the two lung specialists have no budget for additional measurements. Since the automated system already distinguishes between the two lung diseases, it can narrow down the possible treatments to a delta over the correct general treatment, thus requiring no additional measurements by the lung specialists (shown by the two columns in $p^\star(a|x)$ that have a delta over the treatment).

The bottom row in **Figure 10** shows the optimal hierarchy for the environment where heart diseases appear with higher probability. In this case it is optimal to redesign the automated system to distinguish between the two types of the heart disease $h1$, $h2$, and lung diseases in general (see $p^\star(x|w)$ in **Figure 10D** of the figure). This means that there are now treatment recommendations $p^\star(a|x)$ for $h1$ and $h2$ that do not require any more measurements by the specialists (shown by the delta over a treatment in the first two columns of $p^\star(a|x)$ in **Figure 10E**) and there is another treatment recommendation for lung diseases. The corresponding specialist can use the limited budget to perform additional measurements to distinguish between the two categories of lung disease (but not between the four possible types as this would require more measurements than the budget allows). The example illustrates how the bounded-optimal decision-making hierarchy is shaped by the environment and emerges from optimizing the trade-off between expected utility and overall information processing cost. Readers can interactively explore the example in the Supplementary Jupyter Notebook "4-ParallelHierarchy" – in particular by changing the information processing costs of the specialists β_3 or changing the number of specialists by increasing or decreasing the cardinality of X.

The solutions shown in this section were obtained by iterating the self-consistent equations until numerical convergence. Since there is no convergence-proof, it cannot be fully ruled out that the solutions are sub-optimal with respect to the objective. However, the point of the simulation results shown here is to allow for easier interpretation of the theoretical results and highlight certain aspects of the theoretical findings. We discuss this issue in Section 5.2.

4.3. Comparing Parallel and Serial Information Processing

In order to achieve a certain expected utility, a certain overall rate $I(W;A)$ is needed. In the one-step rate-distortion case (Section 2) the channel from w to a must have a capacity larger or equal to that rate. In the serial case (Section 3) there is a channel from w to x and another channel from x to a. Both serial channels must at least have a capacity of $I(W;A)$ in order to achieve the same overall rate, as the following inequality always holds for the serial case

$$I(W;A) \leq \min\{I(W;X), I(X;A)\}.$$

In contrast, the parallel architecture allows for computing a certain overall rate $I(W;A)$ using channels with a lower capacity

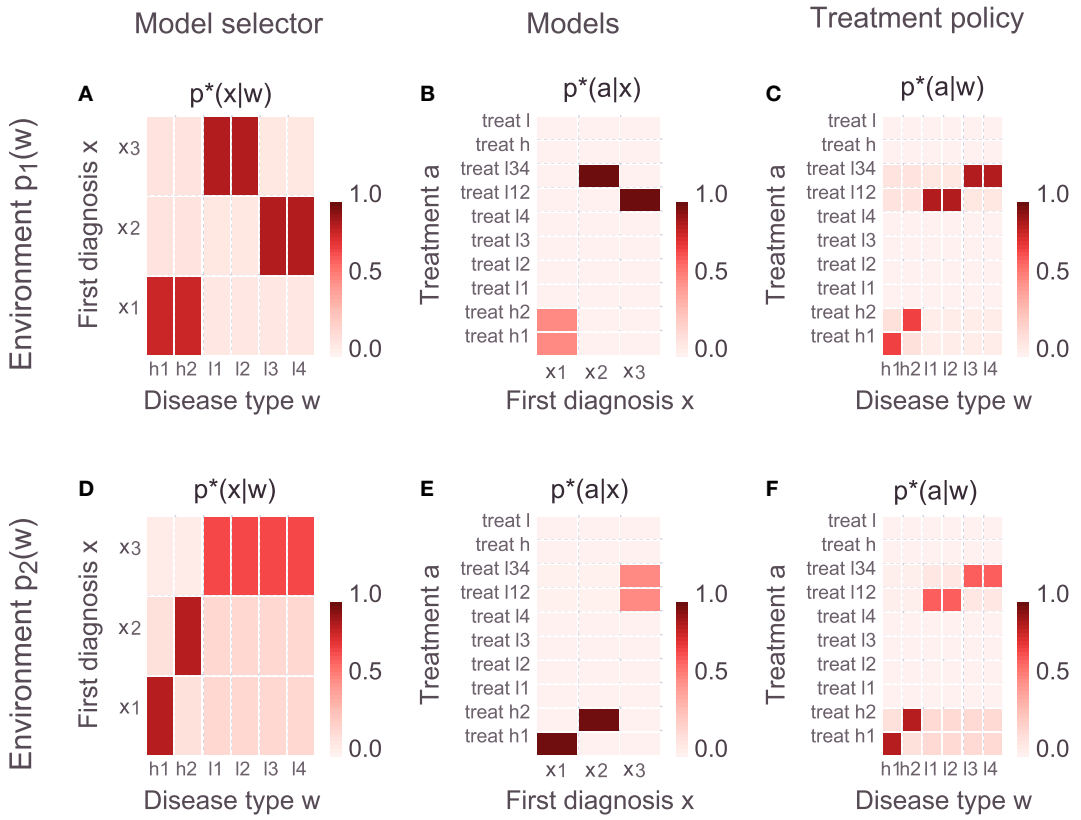

FIGURE 10 | Bounded-optimal hierarchies for two different environments (different *p(w)*). Top row: all disease types are equally probable, bottom row: heart diseases and thus *h*1 and *h*2 are more probable. **(A,D)** Optimal model (or specialist) selectors [equation (26)]. In the uniform environment it is optimal to have one specialist for each disease category (heart, lung A, and lung B) and have a model selector that maps specific disease types to the corresponding specialist *x*. In the non-uniform environment it is optimal to have one specialist for each of the two types of the heart disease (*h*1, *h*2) and another one for all lung diseases. The corresponding model selector in the bottom row reflects this change in the environment. **(B,E)** Optimal models [equation (29), each column corresponds to one specialist]. The resources of the specialists are very limited due to a very low β_3, meaning that the average deviation of $p^*(a|w, x)$ from $p^*(a|x)$ must be small. In the uniform environment it is optimal that the heart specialist spends all the resources for discriminating between the two heart disease types because using the general heart treatment on both types is less effective compared to using one of the lung disease treatments on both corresponding types. In the non-uniform environment this is reversed as the specialists for *h*1 and *h*2 do not need any more measurements to determine the correct treatment, however, the remaining budget (of 1 bit) is spent on discriminating between the two categories of lung diseases (but is insufficient for discriminating between the four possible lung disease types). **(C,F)** Overall-behavior of the hierarchical system $p^*(a|w) = \Sigma_x p^*(x|w)p^*(a|w, x)$. The distributions $p^*(a|w, x)$ [equation (28)] are not shown in the figure but can easily be investigated in the Supplementary Jupyter Notebook "4-ParallelHierarchy."

because the contribution in reducing uncertainty about a on each level of the hierarchy splits up as follows:

$$\underbrace{I(W,X;A)}_{\text{total reduction}} = \underbrace{I(X;A)}_{\text{high-level}} + \underbrace{I(W;A|X)}_{\text{low-level}},$$

which implies $I(W, X; A) \geq I(X; A)$. In particular, if the low-level step contributes information then $I(W; A|X) > 0$ and the previous inequality becomes strict. The same argument also holds when considering $I(W; A)$ (see Section 5.1).

In many scenarios the maximum capacity of a single processing element is limited and it is desirable to spread the total processing load on several elements that require a lower capacity. For instance, there could be technical reasons why processing elements with 5 bits of capacity can easily be manufactured but processing elements with a capacity of 10 bits cannot be manufactured or are disproportionally more costly to produce. In the one-step case and the serial case the only way to stay below a certain capacity limit is by tuning β until the required rate is below the capacity – however, in both cases this also decreases the overall rate $I(W; A)$. In the parallel hierarchical case several building blocks with a limited capacity can be used to produce an overall rate $I(W; A)$ larger than the capacity of each processing block.

Splitting of information processing load onto several processing blocks is illustrated in **Figure 11**, where the one-step and parallel hierarchical solutions to the medical example are compared. In this example, the price of information processing in the one-step case is quite low ($\beta = 10$ bits per util) such that the corresponding solution leads to a deterministic mapping of each w to the best a (see **Figure 11A**). Doing so requires $I(W; A) \approx 2.6$ bits (see **Figure 11B**). Now assume for the sake of this example that processing elements, where information processing cost is reduced (15 bits per util), could be manufactured, but the maximum capacity of these elements is 1.58 bits. In the one-step case these processing elements can only be used if it is acceptable to reduce the rate $I(W; A)$ to 1.58 bits (by tuning β) which

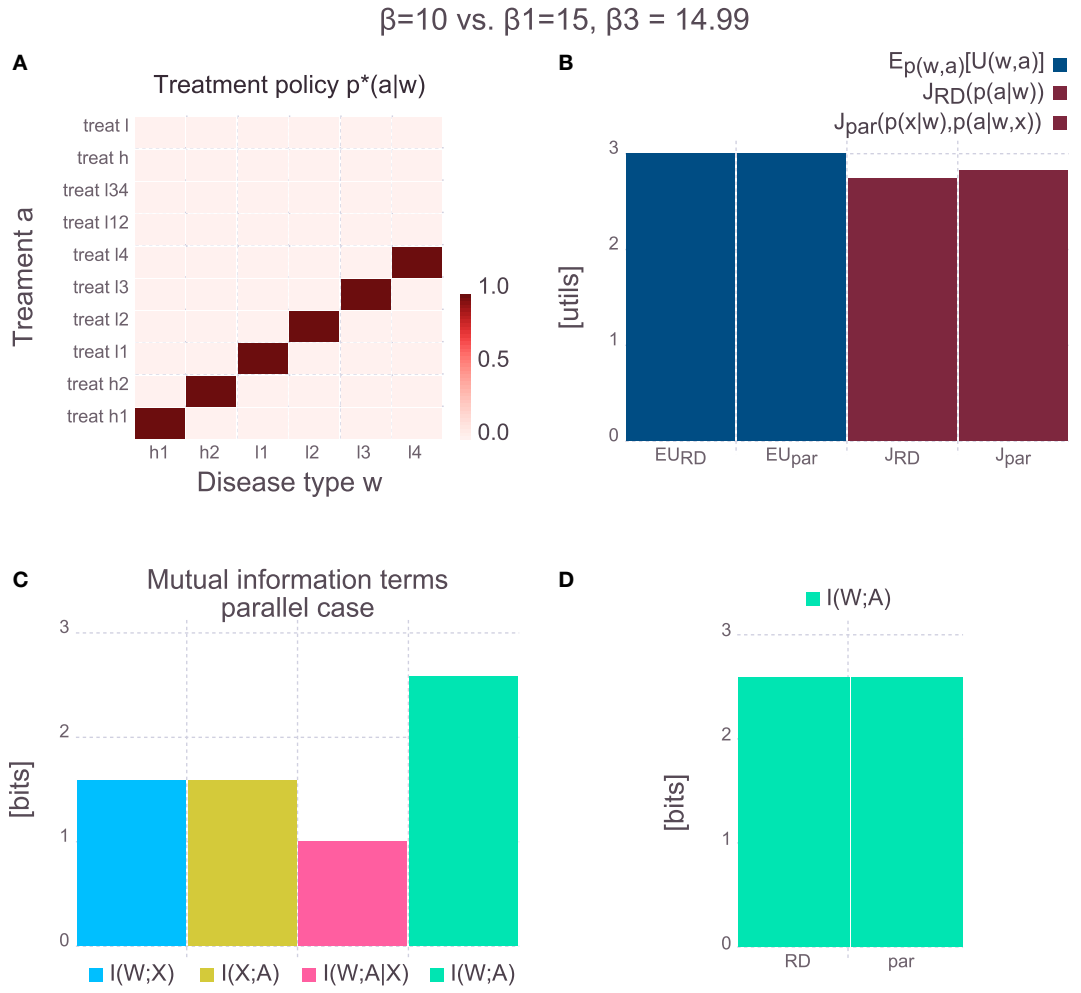

FIGURE 11 | Bounded-optimal parallel hierarchy (denoted by "par") versus one-step (rate-distortion) case (denoted by "RD"). **(A)** Overall-behavior $p^*(a|w) = \Sigma_x p^*(x|w)p^*(a|w, x)$. The overall-behavior is identical for the one-step case and the parallel hierarchical solution. Due to the low price of information processing each w is deterministically mapped to the best a. **(B)** Comparison of performance of one-step ("RD") and parallel case ("par"). Both solutions achieve the same expected utility but the hierarchical solution has a better trade-off between expected utility and information processing cost (because the price of information processing is lower in the hierarchical case $\frac{1}{\beta}_1$, $\frac{1}{\beta}_3$ compared to $\frac{1}{\beta}$ in the one-step case) since in the hierarchical case the cheap but capacity-limited processing elements can be used – see main text). $J_{RD}(p(a|w))$ corresponds to equation (7) and $J_{par}(p(x|w), p(a|w, x))$ corresponds to equation (25). **(C)** Mutual information terms of the parallel case $I(W; X)$, $I(X; A)$, $I(W; A|X)$, and $I(W; A)$ (from left to right). The individual processing channels in the hierarchical case are all within the capacity limit of 1.58 bits, but the overall processed information $I(W; A)$ is ≈2.6 bits. **(D)** Comparison of "effective" information $I(W; A)$ processed in the parallel case (denoted by "par") and the information processed in the one-step (rate-distortion) case (denoted by "RD"). $I(W; A)$ is identical for the one-step and the hierarchical case, but in the parallel case it was computed in a distributed fashion, where each computational step (or channel) has a rate lower than $I(A; W)$ [see **(C)**]. The figure can be reproduced with the Supplementary Jupyter Notebook "5-DistributionOfInformationProcessing."

would imply a lower expected utility. However, in the parallel hierarchical case the new processing elements can be used (see **Figure 11C**) which leads to a reduced price of information processing ($\beta_1 = 15$, $\beta_3 = 14.999$). In conjunction the new processing elements process the same effective information $I(W; A)$ (see **Figure 11D**) and achieve the same expected utility as the one-step case (see **Figure 11B**). However, since the price for information processing is lower on the more limited elements, the overall trade-off between expected utility and information processing cost is in favor of the parallel hierarchical architecture. Note that in this example it is important that the cardinality of X is limited (in this case $|\mathcal{X}| = 3$) and $\beta_1 > \beta_3$. We discuss this in the next paragraphs.

In the parallel hierarchical case, there are two possible pathways from w to a:

$$\begin{aligned} \text{Two-stage serial pathway} \quad & I(W; X) \rightarrow I(X; A) \\ \text{Parallel pathway} \quad & I(W; A|X) \end{aligned}$$

Note that $I(X; A)$ does not appear in the objective [equation (25)]; however, it is crucial for distributing information processing on both levels of the hierarchy (see more analysis of the medical system example in **Figure 11**). $I(X; A)$ measures the average adaptation effort for going from $p(a)$ to $p(a|x)$. In the parallel hierarchical case it is a measure of how much the different models $p(a|x)$ narrow down the search space compared to the average

$p(a) = \Sigma_x p(x)p(a|x)$. If all models are equal $p(a|x) = p(a)\,\forall x$ the mutual information $I(X; A)$ is zero. Note, however, that a large $I(X; A)$ is rendered useless by a low $I(W; X)$ and vice versa – if the model selector is very bad, even the best models are not useful and vice versa.

Since there is no cost for having a large rate $I(X; A)$, the overall throughput of the serial pathway is effectively governed by β_1 as it affects the rate $I(W; X)$. Similarly, β_3 governs the rate on the parallel pathway $I(W; A|X)$. As a result, whenever one of the two inverse temperatures β_1 and β_3 is larger than the other, it becomes more economic to shift all the information processing to the cheaper pathway (either serial or parallel) thus rendering the other pathway obsolete. The only scenario where it can be advantageous to use both pathways (and distribute computation) is when the cheaper pathway has insufficient capacity and the more expensive pathway is used to take on additional computational load that cannot be handled by the cheap pathway alone. Effectively, this translates into the constraint that the serial pathway must be cheaper $\beta_1 > \beta_3$ and additionally the serial pathway must be limited in its capacity by limiting the cardinality $|\mathcal{X}|$ (see Supplementary Methods Section 5 for a detailed discussion).

Note the important difference between changing the cardinality of X which governs the channel capacity of the serial pathway (that is the maximally possible rates $I(W; X)$, $I(X; A)$) but has no influence on the price of information processing and changing β_1 which governs the price of processing $I(W; X)$ and hence affects the actual rate on the serial pathway but has no effect on the capacity of the channels of the serial pathway.

$$I(W; X) = H(X) - H(X|W) \leq \mathbf{H}(\mathbf{X}) \tag{31}$$

$$I(X; A) = H(A) - H(A|X) = H(X) - H(X|A) \leq \mathbf{H}(\mathbf{X}) \tag{32}$$

$$I(W; A|X) = H(A|X) - H(A|W, X) \leq H(A|X) \tag{33}$$

$H(X)$ is an upper bound for both $I(W; X)$ but also $I(X; A)$ and the upper bound of $H(X)$ itself is a function of $|\mathcal{X}|$. Note that since there is no cost associated with $I(X; A)$ it is generally desirable to maximize $I(X; A)$ at least such that $I(X; A) \geq I(W; X)$. To do so $H(A|X)$ must be pushed toward zero [equation (32)] – however, this simultaneously pushes the upper bound for $I(W; A|X)$ toward zero [equation (33)]. In case of a sufficiently limited $H(X)$ (through a low $|\mathcal{X}|$), $I(X; A)$ cannot be fully maximized, therefore leaving a non-zero upper bound for $I(W; A|X)$.

In the example shown in **Figure 11** information processing is performed on both the serial pathway ($I(W; X)$ and $I(X; A)$) but also on the parallel pathway ($I(W; A|X)$) because the constraints for distribution of information processing are fulfilled: $\beta_1 > \beta_3$ and the capacity (that is the maximum rate $I(W; X)$ and $I(X; A)$) of the serial pathway is limited by the (low) cardinality $|\mathcal{X}| = 3$. The cardinality of X for the example can easily be changed in the Supplementary Jupyter Notebook "4-ParallelHierarchy" – if it is for instance increased to $|\mathcal{X}| = 6$ while keeping all other parameters the same, the whole information processing load will be entirely on the serial pathway and $I(W; A|X) = 0$. Alternatively to limiting the cardinality of X, a cost for $I(X; A)$ could be introduced to limit the computational resources for computing $p(a|x)$ from $p(a)$. This is explored in Section 5.

5. TOWARD MORE GENERAL ARCHITECTURES

In the serial case in Section 3, information processing cost arises from adapting $p(x)$ to $p(x|w)$ and $p(a)$ to $p(a|x)$, and the average informational effort is measured by $I(W; X)$ and $I(X; A)$. In the parallel hierarchical case in Section 4 the two information processing terms considered are $I(W; X)$ and $I(W; A|X)$, where the latter measures the average informational effort for adapting from $p(a|x)$ to $p(a|w, x)$. In this section, we present a mathematically unifying case that considers all three mutual information terms and includes the serial and the parallel case as special cases. This unifying formulation might also be a starting point for generalizing toward more than three random variables as the corresponding objective function could easily be extended to include more variables.

The general case uses the same factorization of the three variables W, X, A as the parallel case: $p(w, x, a) = p(w)p(x|w)p(a|w, x)$. Given this factorization, the KL-divergence between the joint $p(w, x, a)$ and the product of all three marginals, also known as the total correlation $C(W, X, A)$, leads to:

$$\begin{aligned} C(W, X, A) &= D_{\mathrm{KL}}(p(w, x, a)||p(w)p(x)p(a)) \\ &= H(W) + H(X) + H(A) - H(W, X, A) \qquad (34) \\ &= \sum_{w,x,a} p(w, x, a) \log \frac{p(w, x, a)}{p(w)p(x)p(a)} \\ &= \sum_{w,x} p(w, x) \log \frac{p(x|w)}{p(x)} \\ &\quad + \sum_{w,a,x} p(w, x, a) \log \frac{p(a|w, x)}{p(a)} \\ &= I(W; X) + I(W, X; A) \qquad (35) \\ &= I(W; X) + I(X; A) + I(W; A|X). \qquad (36) \end{aligned}$$

The total correlation (Watanabe, 1960), also called multivariate constraint (Garner, 1962) or multiinformation (Studený and Vejnarová, 1998), is the sum of the three information processing terms considered in the serial and parallel case. The general objective is formed by assigning different prices to each of the terms and trading off the resulting information processing cost against the expected utility:

$$\begin{aligned} &\max_{p(x|w), p(a|w,x)} \mathbf{E}_{p(w,x,a)}[U(w, a)] - \frac{1}{\beta_1} I(W; X) \\ &\quad - \frac{1}{\beta_2} I(X; A) - \frac{1}{\beta_3} I(W; A|X) \\ &\quad = \max_{p(x|w), p(a|w,x)} J_{\mathrm{gen}}(p(x|w), p(a|w, x)). \end{aligned} \tag{37}$$

Identical to the parallel hierarchical case, the general case has two information processing pathways that allow for splitting up the total computational load: a serial pathway consisting of the two stages $I(W; X)$ and $I(X; A)$ and a parallel pathway $I(W; A|X)$. If any of the pathways is cheaper than the other one, it is more economical to shift all the computation to the cheaper pathway. However, the capacity of the serial pathway can be limited, for

example by reducing the cardinality of X. In such a case the parallel pathway can take on additional computational load, leading to a parallel hierarchical information processing architecture.

The solution to the general objective is given by the following set of five self-consistent equations (the detailed derivation of the solutions is included in the Supplementary Methods Section 1):

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp\Big(\beta_1 \Delta F_{\text{gen}}(w,x) - \Big(\frac{\beta_1}{\beta_3} - \frac{\beta_1}{\beta_2}\Big) D_{\text{KL}}(p^*(a|w,x)||p^*(a|x))\Big) \quad (38)$$

$$p(x) = \sum_w p(w) p^*(x|w) \quad (39)$$

$$p^*(a|w,x) = \frac{1}{Z(w,x)} p^*(a|x) \exp\left(\beta_3 U(w,a) - \frac{\beta_3}{\beta_2} \log \frac{p^*(a|x)}{p(a)}\right) \quad (40)$$

$$p^*(a|x) = \sum_w p(w|x) p^*(a|w,x) \quad (41)$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|w,x), \quad (42)$$

where $Z(w)$ and $Z(w, x)$ denote the corresponding normalization constants or partition sums. The conditional distribution $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{gen}(w,x)}$ is the free energy difference

$$\Delta F_{\text{gen}}(w,x) := \mathbf{E}_{p^*(a|w,x)}[U(w,a)] - \tfrac{1}{\beta}_2 D_{\text{KL}}(p^*(a|w,x)||p(a)).$$

For $\beta_3 < \beta_2$ the KL-term in equation (38) has a positive sign, implying that the KL-divergence is a utility instead of a cost which makes sense if computation on $I(W; A|X)$ is cheaper than computation on $I(A|X)$. For $\beta_3 > \beta_2$ the KL-term gets a negative sign, implying that the KL-divergence is a cost, as a result of computation on $I(W; A|X)$ being more expensive than computation on $I(A|X)$.

Equation 38 can also be rewritten as (see Supplementary Methods Section 1.2):

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp\Big(\beta_1 \Delta F_{\text{par}}(w,x) - \frac{\beta_1}{\beta_2} \sum_a p^*(a|w,x) \log \tfrac{p^*(a|x)}{p(a)}\Big) \quad (43)$$

where $\Delta F_{\text{par}(w,x)}$ is the same free energy difference as in the parallel case

$$\Delta F_{\text{par}}(w,x) := \mathbf{E}_{p^*(a|w,x)}[U(w,a)] - \tfrac{1}{\beta}_3 D_{\text{KL}}(p^*(a|w,x)||p^*(a|x)) \quad (44)$$

see equation (30).

Comparing the objective in equation (37) with the objective of the parallel case in equation (25), it can be seen that by setting $\beta_2 \to \infty$ the two objective functions become equal and the implicit assumption that in the parallel case there is no cost for going from $p(a)$ to $p(a|x)$ (as the latter is considered a prior) is made explicit. The solution equations of the general case also collapse to the solutions of the parallel case by letting $\beta_2 \to \infty$: compare equations (43) and (40) against equations (26) and (28). The general case thus also allows for designing more realistic hierarchical cases where there is a small cost for switching models $p^*(a|x)$ that arises, for instance, from loading a certain set of parameters or switching to a particular sampler or reading the model from memory. Similarly, the serial case can be recovered by $\beta_3 \to 0$. The special cases of the general objective are summarized in **Table 2**.

TABLE 2 | Recovery of special cases from the general, unifying case by specific settings of the inverse temperatures.

Case	β_1	β_2	β_3	(inverse) price per transformation
General	β_1	β_2	β_3	β_1: $p(x) \to p(x\|w)$ β_2: $p(a) \to p(a\|x)$ β_3: $p(a\|x) \to p(a\|w,x)$
Total correlation	β	β	β	β: $p(x) \to p(x\|w)$ β: $p(a) \to p(a\|w,x)$
Degenerate TC	β_1	β	β	β_1: $p(x) \to p(x\|w)$ β: $p(a) \to p(a\|w,x)$
Serial	β_1	β_2	$\to 0$	β_1: $p(x) \to p(x\|w)$ β_2: $p(a) \to p(a\|x)$ $p(a\|w,x) = p(a\|x)\ \forall w$ $I(W; A\|X) = 0$
Parallel	β_1	$\to \infty$	β_3	β_1: $p(x) \to p(x\|w)$ β_3: $p(a\|x) \to p(a\|w,x)$
Joint (x,a)	β	$\to \infty$	β	β: $p(x,a) \to p(x,a\|w)$

The table shows how to set the inverse temperatures in the general case to recover particular special cases. The last column shows for all cases which probability-transformations are considered as computational effort and the corresponding (inverse) price. The case "degenerate total correlation" is not described in the main paper, but is outlined in the Supplementary Methods Section 4 – it could be relevant in a two-dimensional decision-making scenario, that is when x is considered one dimension of the decision and a is considered the other dimension. This implies that the utility function also depends on x: U(w,x,a). Similarly, the case "joint (x,a)" is only described in the Supplementary Methods Section 5 and describes how the one-step (rate-distortion) case is related to the general case.

5.1. Effective Information Throughput $I(W; A)$

The amount of information processing that effectively contributes toward achieving a high expected utility is measured by $I(W; A)$ which does not directly appear in the objective of the general case (nor the serial and parallel case). However, the effective information throughput of the system is given by

$$I(W;A) = I(W;X;A) + I(W;A|X) \quad (45)$$

$$= I(W;X) + I(X;A) - I(X;W,A) + I(W;A|X) \quad (46)$$

$$= C(W,X,A) - I(X;W,A) \quad (47)$$

where $I(W; X; A)$ denotes the multivariate mutual information (MMI; Yeung, 1991). The first equation above is obtained by re-ordering the definition of the MMI $I(K; L; M) = I(K; M) - I(K; M|L)$. Note that in the serial hierarchical case $I(W; A|X) = 0$ always holds. The equations above also show how the total correlation $C(W, X, A)$ and the MMI $I(W; X; A)$ are related.

The multivariate mutual information is upper-bounded by $I(W; X; A) \leq \min\{I(W; X), I(W; A), I(X; A)\}$ (see (Yeung, 1991)). Using the bound in equation (45) leads to and upper bound for the effective information throughput:

$$I(W;A) \leq \min\{I(W;X), I(X;A)\} + I(W;A|X) \quad (48)$$

Equation 48 shows how information processing in the general case can be distributed between a two-stage serial pathway (consisting of $I(W;X)$ and $I(X;A)$) and a parallel pathway ($I(W;A|X)$). The general case forms a parallel hierarchy similar to Section 4, but it allows to associate a cost with $I(X;A)$ (which is a measure of how costly it is to switch models). Importantly the discussion on splitting up information processing between both levels of the parallel hierarchy as in Section 4.3 also holds for the general case.

5.2. Iterating the Self-Consistent Equations

For the simulation results shown in this paper the corresponding set of self-consistent equations was iterated until convergence (by checking that the total change in probability distributions between two iteration steps is below a certain threshold – see code underlying the Supplementary Notebooks for details). This is inspired by the Blahut-Arimoto scheme that is proven to converge to the global maximum in the rate-distortion case (Csiszar, 1974; Cover and Thomas, 1991) (Section 2). Unfortunately there is no such proof for iterating the sets of self-consistent equations of the general, serial or parallel case. It is not clear whether the optimization problems are still convex and have a global solution, nor is it clear that iterating the self-consistent equations would converge toward these global solutions. A convexity and convergence analysis is certainly among the most important steps for future investigations of the principles presented here. At this point, we can only report empirical observations and interested readers are encouraged to explore convergence behavior using the Supplementary Jupyter Notebooks (which include plots that show convergence behavior across iterations) but also the underlying code (published in the Supplementary Material).

6. DISCUSSION AND CONCLUSION

The overarching principle behind this paper is the consistent application of the trade-off of gains in expected utility against the computational cost that these gains require. Here, computational cost is defined as the average effort of computational adaptation (measured by the mutual information) multiplied by the price of information processing. This definition is motivated by first principles (Mattsson and Weibull, 2002; Ortega and Braun, 2010; Ortega and Braun, 2011) and is grounded in a thermodynamic framework for decision-making (Ortega and Braun, 2013). Mathematically, the basic principle is identical to the principle behind rate-distortion theory, the information-theoretic framework for lossy compression (Genewein and Braun, 2013; Still, 2014). This connection is no coincidence as bounded rational decision-making can be cast as a lossy compression problem in lossy compression the goal is to transmit the most relevant information (given the limited channel capacity) in order to minimize a distortion-function. In bounded rational decision-making the goal is to process the most relevant information in order maximize a utility function, given the limitations on information processing. In Section 2, we have shown how different levels of behavioral abstraction can be induced by different computational limitations. The authors in (van Dijk and Polani, 2013) use the *Relevant Information* method, which is a particular application of rate-distortion theory and find a very similar emergence of "natural abstractions" and "ritualized behavior" when studying goal-directed behavior in the MDP case. We have shown how the basic principle can be extended to more complex cases and that analytic solutions can be obtained for these cases. Importantly, the solutions allow for interesting interpretations, highlighting how the same fundamental trade-off can lead to systems that elegantly solve more complex problems. For instance, when designing a perception-action system, the perceptual part of the system can easily be understood as a lossy compressor, but the corresponding distortion-function is not intuitively clear. We have shown in Section 3 how the extended lossy compression principle leads to a well-defined distortion-function for the perceptual part of the system that optimizes the downstream trade-off between expected utility and computational cost. In a similar fashion we have shown in Section 4 how the problem of designing bounded-optimal decision-making hierarchies is fundamentally equal to designing a distributed lossy compressor (that is spread over both levels of the hierarchy).

In the serial hierarchy in Section 3, we compared a perceptual channel that performs Bayesian inference against a bounded-optimal perceptual channel that optimizes the downstream free energy trade-off. We found that the difference between both models of perception was that in one case the likelihood model $p(x|w)$ was unspecified (Bayesian inference) whereas in the other case it was well defined (bounded-optimal solution). Perception is often conceptualized as (Bayesian) inference, however, given our findings there is a subtle but important difference. In our model of a perception-action system, the goal of the perceptual model $p(x|w)$ is to extract the most relevant information from w for choosing an action according to $p(a|x)$, given the computational limitations of the system. In plain inference, the goal is to predict w from x very well and the likelihood model is thus chosen to maximize predictive power. In many cases the two objectives coincide as achieving a large expected utility often requires precise knowledge about w. However, this must not always be the case and in particular for systems where computational limitations play a large role, the (limited) computational resources can often be spent more economically which allows for a higher expected utility at the cost of not being able to predict w from x that well. An interesting machine-learning application of the serial principle could be the design of optimal features for classification.

In Section 4, we showed how parallel bounded-optimal decision-making hierarchies can emerge from solving the trade-off between utility and cost of computation. We found that the condition for parallel hierarchies to being optimal solutions was that the price for model selection is lower than the price for processing information on the low level of the hierarchy ($\beta_1 > \beta_3$). At the same time, the upper level of the hierarchy must be limited in capacity (for instance through the cardinality $|\mathcal{X}|$). Intuitively this makes sense and fits with the general observation that often hardware that allows for cheap information processing is itself quite expensive to build (low signal to noise ratio, etc.). Therefore the amount of hardware that allows for cheap information processing is likely to be quite limited. It remains an open question whether this is a fundamental constraint for hierarchies being optimal solutions or whether there are other arguments in

favor of hierarchical architectures. In changing environments, for example, the overall change required to adapt a system is smaller for a hierarchical system, compared to a flat system, because the more abstract levels of the hierarchy might require little or no change at all. It could also be that the upper levels of a hierarchical model based on our principle contain more transferable knowledge that can be applied to novel but similar tasks. Changing the task corresponds to changing the utility-function, which requires a non-equilibrium analysis (Grau-Moya and Braun, 2013) that we leave for future investigation.

In our simulations, we initialize $p(x|w)$ and either $p(a|w, x)$ (parallel hierarchies) or $p(a|x)$ (serial hierarchies) and iterate the equations until (numerical) convergence. We found sometimes that the solutions can be sensitive to the initialization. This hints at the problem being non-convex or the iteration-scheme being prone to get stuck in local optima or plateaus. In particular, we find that in the serial hierarchy with low cost of computation, a sparse, diagonal-like initialization of $p(x|w)$ works much better than a random initialization. For the parallel hierarchies, we found that a random or sparse initialization of $p(x|w)$ combined with a uniform initialization of $p(a|w, x)$ works most reliably. Additionally we found that in the hierarchical case if β_3 is slightly larger than β_1 the iterations converge to sub-optimal solutions where both pathways are used instead of shifting all the computation to the parallel pathway. The toy simulations presented here are illustrative examples only and numerically efficient implementations of the iteration-schemes are beyond the scope of the current paper. These problems might be addressed by other solution schemes like sampling-based or parametric model-based solutions. Nevertheless, these other solution schemes (that potentially do not even require the sets of analytical solutions) can benefit from the interpretations given by the analytic solution equations in this paper.

The ability to form abstractions is thought of as a hallmark of intelligence, both in cognitive tasks and in basic sensorimotor behaviors (Kemp et al., 2007; Braun et al., 2010a,b; Gershman and Niv, 2010; Tenenbaum et al., 2011; Genewein and Braun, 2012). Traditionally, the formation of abstractions is conceptualized as being computationally costly because particular entities have to be grouped together by neglecting irrelevant information. Recently, abstractions that arise from sensory evolution and hierarchical behaviors have been studied from an information-theoretic perspective (Salge and Polani, 2009; Van Dijk et al., 2011). Here, we study abstractions in the process of decision-making, where "similar" situations elicit the same behavior when partially ignoring the current situational context. Extending our principle to hierarchies with more than two levels might provide novel points of view on the formation of hierarchies in biological systems, such as the early visual system (DiCarlo et al., 2012). One fundamental prediction, based on our current work is that the formation of abstractions and concepts should be heavily shaped by the agent-environment structure (the utility function). Following the work of (Simon, 1972) decision-making with limited information-processing resources has been studied extensively in psychology, economics, political science, industrial organization, computer science, and artificial intelligence research. In this paper, we use an information-theoretic model of decision-making under resource constraints (McKelvey and Palfrey, 1995; Kappen, 2005; Wolpert, 2006; Todorov, 2009; Peters et al., 2010; Theodorou et al., 2010; Rubin et al., 2012). In particular, Braun et al. (2011) and Ortega and Braun (2011, 2012, 2013) present a framework in which gain in expected utility is traded off against the adaptation cost of changing from an initial behavior to a posterior behavior. The variational problem that arises due to this trade-off has the same mathematical form as the minimization of a *free energy difference* functional in thermodynamics. Here, we discuss the close connection between the thermodynamic decision-making framework (Ortega and Braun, 2013) and rate-distortion theory which is an information-theoretic framework for lossy compression. The problem in lossy compression is essentially the problem of separating structure from noise and is thus highly related to finding abstractions (Tishby et al., 1999; Still and Crutchfield, 2007; Still et al., 2010). In the context of decision-making the rate-distortion framework can be applied by conceptualizing the decision-maker as a channel from observations to actions *with limited capacity*, which is known in economics as the framework of "Rational Inattention" (Sims, 2003).

The rate-distortion principle and all the extended principles presented in this paper measure computational cost with the mutual information which is an abstract measure that quantifies the average KL-divergence. The mutual information measures the actual transformation of probabilities and thus provides a lower bound for any possible implementation. In fact, different implementations could perform the same transformation more or less efficiently which should reflect in the price of information processing but not the amount of information processed. The advantage of using a generic measure is that the principle is universal and can be applied to any system. The downside of this is that it cannot be directly used to analyze specific implementations. In practice it can be hard to determine how difficult or "costly" it is to implement a certain transformation of probability distributions. Rather, the price for information processing is often set implicitly, for instance by certain computation-time constraints or by constraining the number of samples, etc. When applying the principle to a specific implementation it might be required to derive a novel, specific solution scheme for the corresponding optimization problem. In Leibfried and Braun (2015), for instance, the authors apply the rate-distortion principle for decision-making to a spiking neuron model by deriving a gradient-based update rule for tuning the parameters of the model (the weights of the neuron). In their case, the price of information processing β appears directly in the parameter update equations which leads to an interesting regularizer for the (online) parameter update rule.

The fundamental trade-off between large expected utility and low computational cost appears in many domains such as machine learning, AI, economics, computational biology or neuroscience, and many solutions, such as heuristics, sampling-based approaches, and model-based approximation schemes, exist (Gershman et al., 2015; Jordan and Mitchell, 2015; Parkes and Wellman, 2015). One of the exciting prospects of such an approach

is that it might provide a common ground for research-questions from artificial intelligence and neuroscience, thus partially unifying the two fields that share common origins but have drifted apart over the last decades (Gershman et al., 2015). The main contribution of this paper is to advance a principled mathematical framework that formalizes the problem objective such that the trade-off between large expected utility and low computational cost and its solutions can be addressed in both a qualitative but also quantitative way. The main finding is that the consistent application of the principle beyond simple one-stage information processing systems leads to non-trivial solutions that address questions like optimal likelihood model design or the design of optimal decision-making hierarchies. Since the mathematics can easily be extended to more variables while the underlying principle remains the same, we believe that the formulation presented in this paper is a good candidate for a general underlying objective that is also applicable to biological organisms and evolutionary processes. We find the principle an interesting starting point for solving timely problems in machine learning, robotics, and AI but also for providing an interesting novel angle for research in computational neuroscience and biology. The principle also provides a promising basis for the design and analysis of guided self-organizing systems as most of the inner structure of systems following our principle is emergent (and thus self-organized) but ultimately aimed at solving particular tasks (through the utility function).

AUTHORS CONTRIBUTION

TG and DB conceived the project, TG and JG performed simulations, TG and FL did analysis and derivations, and TG, FL, JG, and DB wrote the paper.

FUNDING

This study was supported by the DFG, Emmy Noether grant BR4164/1-1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/frobt.2015.00027

A Supplementary Methods provides detailed steps to derive the solution to the general case (Section 5) and how to rewrite the solution equations of the general case. Additionally it outlines how to derive the solutions for the serial and the parallel case. It also provides the set of self-consistent equations for the "degenerate total correlation" and "total correlation" case that drop out mathematically from the general case but are not used in this paper (see **Table 2**). The Supplementary Methods provides details to the discussion on the different information processing pathways of the parallel case (Section 4.3). Finally, it contains the proof for the inequality based on equation (6).

The simulations underlying the results presented in this paper are published as supplementary material using Jupyter (http://jupyter.org/) notebooks. The notebooks are considered part of the results of this paper and readers are encouraged to use the notebooks to interactively explore the examples and concepts presented here. The underlying code is written in Julia (Bezanson et al., 2014) and uses the Gadfly package (http://gadflyjl.org/) for visualization. At the time of writing, the notebooks can be run with a local installation of Jupyter and Julia or without any installation in a web-browser through the JuliaBox project (https://www.juliabox.org/). The notebooks and code at the time of publication are provided in a supplementary.zip file but also under (Genewein, 2015). The notebooks and the code behind the notebooks as well as information on different methods to run the notebooks will be kept up-to-date in the accompanying GitHub repository: https://github.com/tgenewein/BoundedRationalityAbstractionAndHierarchicalDecisionMaking. If compatibility issues with future Julia versions are encountered, please refer to the GitHub repository and feel free to submit an issue. A readme-file on how to run the notebooks (with or without installation) is also provided in the supplementary data as well as in the GitHub repository.

The following notebooks are provided:

- "1-FreeEnergyForBoundedRationalDecisionMaking": Illustrates the results of Section 1 and reproduces **Figure 1**.
- "2-RateDistortionForDecisionMaking": Illustrates the results of Section 4 (the recommender system example) and reproduces **Figures 2–4**. The notebook can be used as a general template for setting up any of the examples presented in the paper and solving it using Blahut-Arimoto.
- "S1-SampleBasedBlahutArimoto": A simple proof-of-concept implementation of sample-based Blahut-Arimoto iterations. Due to space-constraints, this part has been omitted from the paper, but interested readers can find a short theoretical part on the sampling approach in the notebook. Additionally, the notebook shows an implementation of the sampling scheme and applies it to a toy example.
- "3-SerialHierarchy": Illustrates the comparison between hand-crafted perception and bounded-optimal perception in the serial case (Section 3) using the predator-prey example. The notebook reproduces **Figures 5–8**. The notebook allows to easily modify the parameters (e.g., inverse temperatures) of the example or to switch to a different utility function. It can also be used to see how the parallel or general case solution for the predator-prey example would look like.
- "4-ParallelHierarchy": Illustrates the emergence of bounded-optimal hierarchies in two different environments of the medical system example as presented in Section 4 and reproduces **Figures 9** and **10**. The notebook can be used to easily explore the different information processing pathways in the parallel case but also to compare any two cases against each other (because it compares two general case solutions and they can be tuned to all of the special cases).
- "5-DistributionOfInformationProcessing": Compares the parallel hierarchical solution to the medical example to the one-step (rate-distortion) case as shown in **Figure 11**. Since it implements the parallel case through the general case, it also allows to compare any other case to the one-step solution.

REFERENCES

Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory* 18, 14–20. doi:10.1109/TIT.1972.1054753

Ashby, W. R. (1956). *An Introduction to Cybcernetics*. London: Chapman & Hall.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2014). Julia: a fresh approach to numerical computin. *arXiv preprint arXiv:1411*.1607.

Bishop, C. M. (2006). "Sampling methods," in *Pattern Recognition and Machine Learning, Number 4 in Information Science and Statistics*, Chap. 11 (New York: Springer).

Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory* 18, 460–473. doi:10.1109/TIT.1972.1054855

Braun, D. A., Mehring, C., and Wolpert, D. M. (2010a). Structure learning in action. *Behav. Brain Res.* 206, 157–165. doi:10.1016/j.bbr.2009.08.031

Braun, D. A., Waldert, S., Aertsen, A., Wolpert, D. M., and Mehring, C. (2010b). Structure learning in a sensorimotor association task. *PLoS ONE* 5:e8973. doi:10.1371/journal.pone.0008973

Braun, D. A., and Ortega, P. A. (2014). Information-theoretic bounded rationality and epsilon-optimality. *Entropy* 16, 4662–4676. doi:10.3390/e16084662

Braun, D. A., Ortega, P. A., Theodorou, E., and Schaal, S. (2011). "Path integral control and bounded rationality," in *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (Piscataway: IEEE), 202–209.

Burns, E., Ruml, W., and Do, M. B. (2013). Heuristic search when time matters. *J. Artif. Intell. Res.* 47, 697–740. doi:10.1613/jair.4047

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NY: Princeton University Press.

Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. Hoboken: John Wiley & Sons.

Csiszar, I. (1974). On the computation of rate-distortion functions. *IEEE Trans. Inf. Theory* 20, 122–124. doi:10.1109/TIT.1974.1055146

Csiszár, I., and Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Stat. Decis.* 1, 205–237.

Daniel, C., Neumann, G., and Peters, J. (2012). "Hierarchical relative entropy policy search," in *International Conference on Artificial Intelligence and Statistics*. La Palma.

Daniel, C., Neumann, G., and Peters, J. (2013). "Autonomous reinforcement learning with hierarchical REPS," in *International Joint Conference on Neural Networks*. Dallas.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi:10.1016/j.neuron.2012.01.010

Fox, C. W., and Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artif. Intell. Rev.* 38, 85–95. doi:10.1007/s10462-011-9236-8

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi:10.1038/nrn2787

Garner, W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York: Wiley.

Genewein, T. (2015). Bounded rationality, abstraction and hierarchical decision-making: an information-theoretic optimality principle: supplementary code (v1.1.0). *Zenodo*. doi:10.5281/zenodo.32410

Genewein, T., and Braun, D. A. (2012). A sensorimotor paradigm for Bayesian model selection. *Front. Hum. Neurosci.* 6:291. doi:10.3389/fnhum.2012.00291

Genewein, T., and Braun, D. A. (2013). Abstraction in decision-makers with limited information processing capabilities. *arXiv preprint arXiv:1312.4353*.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278. doi:10.1126/science.aac6076

Gershman, S. J., and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256. doi:10.1016/j.conb.2010.02.008

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi:10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., and Todd, P. M. (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

Grau-Moya, J., and Braun, D. A. (2013). Bounded rational decision-making in changing environments. *arXiv preprint arXiv:1312.6726*.

Horvitz, E. (1988). "Reasoning under varying and uncertain resource constraints," in *AAAI*, Vol. 88 (Palo Alto: AAAI), 111–116.

Horvitz, E., and Zilberstein, S. (2001). Computational tradeoffs under bounded resources. *Artif. Intell.* 126, 1–4. doi:10.1016/S0004-3702(01)00051-0

Horvitz, E. J., Cooper, G. F., and Heckerman, D. E. (1989). "Reflection and action under scarce resources: theoretical principles and empirical study," in *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Vol. 2 (Detroit: Morgan Kaufmann Publishers, Inc.), 1121–1127.

Howes, A., Lewis, R. L., and Vera, A. (2009). Rational adaptation under task and processing constraints: implications for testing theories of cognition and action. *Psychol. Rev.* 116, 717–751. doi:10.1037/a0017187

Janssen, C. P., Brumby, D. P., Dowell, J., Chater, N., and Howes, A. (2011). Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Top. Cogn. Sci.* 3, 123–139. doi:10.1111/j.1756-8765.2010.01125.x

Jones, B. D. (2003). Bounded rationality and political science: lessons from public administration and public policy. *J. Public Adm. Res. Theory* 13, 395–412. doi:10.1093/jopart/mug028

Jordan, M., and Mitchell, T. (2015). Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. doi:10.1126/science.aaa8415

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi:10.1257/000282803322655392

Kappen, H. J. (2005). Linear theory for control of nonlinear stochastic systems. *Phys. Rev. Lett.* 95, 200–201. doi:10.1103/PhysRevLett.95.200201

Kappen, H. J., Gómez, V., and Opper, M. (2012). Optimal control as a graphical model inference problem. *Mach. Learn.* 87, 159–182. doi:10.1007/s10994-012-5278-7

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Dev. Sci.* 10, 307–321. doi:10.1111/j.1467-7687.2007.00585.x

Leibfried, F., and Braun, D. A. (2015). A reward-maximizing spiking neuron as a bounded rational decision maker. *Neural Comput.* 27, 1686–1720. doi:10.1162/NECO_a_00758

Levy, R. P., Reali, F., and Griffiths, T. L. (2009). "Modeling the effects of memory on human online sentence processing with particle filters," in *Advances in Neural Information Processing Systems* (Vancouver: NIPS), 937–944.

Lewis, R. L., Howes, A., and Singh, S. (2014). Computational rationality: linking mechanism and behavior through bounded utility maximization. *Top. Cogn. Sci.* 6, 279–311. doi:10.1111/tops.12086

Lieder, F., Griffiths, T., and Goodman, N. (2012). "Burn-in, bias, and the rationality of anchoring," in *Advances in Neural Information Processing Systems* (Lake Tahoe: NIPS), 2690–2798.

Lipman, B. (1995). Information processing and bounded rationality: a survey. *Can. J. Econ.* 28, 42–67. doi:10.2307/136022

Mattsson, L. G., and Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games Econ. Behav.* 41, 61–78. doi:10.1016/S0899-8256(02)00014-3

McKelvey, R. D., and Palfrey, T. R. (1995). Quantal response equilibria for normal-form games. *Games Econ. Behav.* 10, 6–38. doi:10.1006/game.1995.1023

Neal, R. M. (2003). Slice sampling. *Ann. Stat.* 31, 705–767. doi:10.1214/aos/1056562461

Neymotin, S. A., Chadderdon, G. L., Kerr, C. C., Francis, J. T., and Lytton, W. W. (2013). Reinforcement learning of two-joint virtual arm reaching in a computer model of sensorimotor cortex. *Neural Comput.* 25, 3263–3293. doi:10.1162/NECO_a_00521

Ortega, P., and Braun, D. (2010). "A conversion between utility and information," in *Third Conference on Artificial General Intelligence (AGI 2010)* (Lugano: Atlantis Press), 115–120.

Ortega, P. A., and Braun, D. A. (2014). Generalized Thompson sampling for sequential decision-making and causal inference. *Complex Adapt. Syst. Model.* 2, 269–274. doi:10.1186/2194-3206-2-2

Ortega, P. A., Braun, D. A., and Tishby, N. (2014). "Monte Carlo methods for exact & efficient solution of the generalized optimality equations," in *Proceedings of IEEE International Conference on Robotics and Automation*. Hong Kong.

Ortega, P. A., and Braun, D. A. (2011). "Information, utility and bounded rationality," in *Proceedings of the 4th International Conference on Artificial General Intelligence* (Mountain View: Springer-Verlag), 269–274.

Ortega, P. A., and Braun, D. A. (2012). "Free energy and the generalized optimality equations for sequential decision making," in *Journal of Machine Learning Research: Workshop and Conference Proceedings* (Edinburgh: JMLR W&C Proceedings), 1–10.

Ortega, P. A., and Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 469, 2153.

Palmer, S. E., Marre, O., Berry, M. J., and Bialek, W. (2015). Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* 112(22), 6908–6913. doi: 10.1073/pnas.1506855112

Parkes, D. C., and Wellman, M. P. (2015). Economic reasoning and artificial intelligence. *Science* 349, 267–272. doi:10.1126/science.aaa8403

Peters, J., Mülling, K., and Altun, Y. (2010). "Relative entropy policy search," in *AAAI*. Atlanta.

Ramsey, F. P. (1931). "Truth and probability," in *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite (New York, NY: Harcourt, Brace and Co), 156–198.

Rawlik, K., Toussaint, M., and Vijayakumar, S. (2012). "On stochastic optimal control and reinforcement learning by approximate inference," in *Proceedings Robotics: Science and Systems*. Sydney.

Rubin, J., Shamir, O., and Tishby, N. (2012). "Trading value and information in mdps," in *Decision Making with Imperfect Decision Makers* (Springer), 57–74.

Rubinstein, A. (1998). *Modeling Bounded Rationality*. Cambridge: MIT Press.

Russell, S. (1995). "Rationality and intelligence," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. C. Mellish (San Francisco, CA: Morgan Kaufmann), 950–957.

Russell, S. J., and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.

Russell, S. J., and Subramanian, D. (1995). Provably bounded-optimal agents. *J. Artif. Intell. Res.* 2, 575–609.

Salge, C., and Polani, D. (2009). Information-driven organization of visual receptive fields. *Adv. Complex Syst.* 12, 311–326. doi:10.1142/S0219525909002234

Sanborn, A. N., Griffiths, T. L., and Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychol. Rev.* 117, 1144. doi:10.1037/a0020511

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x

Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ.* 69, 99–118. doi:10.2307/1884852

Simon, H. A. (1972). Theories of bounded rationality. *Decis. Organ.* 1, 161–176.

Sims, C. A. (2003). Implications of rational inattention. *J. Monet. Econ.* 50, 665–690. doi:10.1016/S0304-3932(03)00029-1

Sims, C. A. (2005). "Rational inattention: a research agenda," in *Deutsche Bundesbank Spring Conference, Number 4*. Berlin.

Sims, C. A. (2006). Rational inattention: beyond the linear-quadratic case. *Am. Econ. Rev.* 96, 158–163. doi:10.1257/000282806777212431

Sims, C. A. (2010). "Rational inattention and monetary economics," in *Handbook of Monetary Economics*, Vol. 3, Chap. 4 (Elsevier), 155–181.

Spiegler, R. (2011). *Bounded Rationality and Industrial Organization*. Oxford: Oxford University Press.

Still, S. (2009). Information-theoretic approach to interactive learning. *Europhys. Lett.* 85, 28005. doi:10.1209/0295-5075/85/28005

Still, S. (2014). "Lossy is lazy," in *Workshop on Information Theoretic Methods in Science and Engineering* (Helsinki: University of Helsinki), 17–21.

Still, S., and Crutchfield, J. P. (2007). Structure or noise? *arXiv preprint arXiv:0708.0654*.

Still, S., Crutchfield, J. P., and Ellison, C. J. (2010). Optimal causal inference: estimating stored information and approximating causal architecture. *Chaos* 20, 037111. doi:10.1063/1.3489885

Studený, M., and Vejnarová, J. (1998). "The multiinformation function as a tool for measuring stochastic dependence," in *Learning in Graphical Models* (New York: Springer), 261–297.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi:10.1126/science.1192788

Theodorou, E., Buchli, J., and Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *J. Mach. Learn. Res.* 11, 3137–3181.

Tishby, N., Pereira, F. C., and Bialek, W. (1999). "The information bottleneck method," in *The 37th Annual Allerton Conference on Communication, Control, and Computing*.

Tishby, N., and Polani, D. (2011). "Information theory of decisions and actions," in *Perception-Action Cycle*, Chap. 19 (New York: Springer), 601–636.

Tkačik, G., and Bialek, W. (2014). Information processing in living systems. *arXiv preprint arXiv:1412.8752*.

Todorov, E. (2007). "Linearly-solvable Markov decision problems," in *Advances in Neural Information Processing Systems* (Vancouver: NIPS), 1369–1376.

Todorov, E. (2009). Efficient computation of optimal actions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11478–11483. doi:10.1073/pnas.0710743106

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi:10.1126/science.185.4157.1124

van Dijk, S. G., and Polani, D. (2013). Informational constraints-driven organization in goal-directed behavior. *Adv. Complex Syst.* 16:1350016. doi:10.1142/S0219525913500161

Van Dijk, S. G., Polani, D., and Nehaniv, C. L. (2011). "Hierarchical behaviours: getting the most bang for your bit," in *Advances in Artificial Life: Darwin Meets von Neumann*, eds. R. Goebel, J. Siekmann, and W. Wahlster (New York: Springer), 342–349.

Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Vul, E., Alvarez, G., Tenenbaum, J. B., and Black, M. J. (2009). "Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model," in *Advances in Neural Information Processing Systems* (New York: Wiley), 1955–1963.

Vul, E., Goodman, N., Griffiths, T. L., and Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cogn. Sci.* 38, 599–637. doi:10.1111/cogs.12101

Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* 4, 66–82. doi:10.1147/rd.41.0066

Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*, Vol. 25. Cambridge: MIT press.

Wolpert, D. H. (2006). "Information theory-the bridge connecting bounded rational game theory and statistical physics," in *Complex Engineered Systems*, eds D. Braha, A. A. Minai, and Y. Bar-Yam (New York: Springer), 262–290.

Yeung, R. W. (1991). A new outlook on Shannon's information measures. *IEEE Trans. Inf. Theory* 37, 466–474. doi:10.1109/18.79902

Yeung, R. W. (2008). *Information Theory and Network Coding*. New York: Springer.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers

Supplementary Methods: Bounded rationality, abstraction and hierarchical decision-making: an information-theoretic optimality principle

Tim Genewein [1,2,3,*], **Felix Leibfried** [1,2,3], **Jordi Grau-Moya** [1,2,3] **and Daniel Braun** [1,2]

*Correspondence:
Tim Genewein
tim.genewein@tuebingen.mpg.de

1 SOLVING THE VARIATIONAL PROBLEM - GENERAL CASE

The variational problem in Section 5 of the main paper is given by the following constrained optimization problem

$$\begin{aligned}\underset{p(x|w),p(a|w,x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] \quad &\text{subject to} \quad I(W;X)\leq R_1,\ I(X;A)\leq R_2,\ I(W;A|X)\leq R_3 \\ &\text{and} \quad \forall w \sum_x p(x|w)=1,\ \forall w,x \sum_a p(a|w,x)=1,\end{aligned}$$

where R_1, R_2 and R_3 denote given upper bounds on the respective mutual information terms. Translating this constrained optimization problem into an unconstrained optimization problem yields the corresponding Lagrangian $L(p(x|w),p(a|w,x))$

$$\begin{aligned}L(p(x|w),p(a|w,x)) = &\sum_{w,x,a} p(w)p(x|w)p(a|w,x)U(w,a) \\ &+\frac{1}{\beta_1}\left(R_1-\sum_w p(w)\sum_x p(x|w)\log\frac{p(x|w)}{p(x)}\right) \\ &+\frac{1}{\beta_2}\left(R_2-\sum_w p(w)\sum_x p(x|w)\sum_a p(a|w,x)\log\frac{p(a|x)}{p(a)}\right) \\ &+\frac{1}{\beta_3}\left(R_3-\sum_w p(w)\sum_x p(x|w)\sum_a p(a|w,x)\log\frac{p(a|w,x)}{p(a|x)}\right) \\ &+\sum_w \lambda_1(w)\left(\sum_x p(x|w)-1\right)+\sum_{w,x}\lambda_2(w,x)\left(\sum_a p(a|w,x)-1\right),\end{aligned}$$

where $\frac{1}{\beta_1}$, $\frac{1}{\beta_2}$, $\frac{1}{\beta_3}$, $\lambda_1(w)$ and $\lambda_2(w,x)$ are Lagrange multipliers that capture the inequality and equality constraints of the original constrained problem formulation. Strictly speaking, it must also be ensured that the values of the distributions $p(x|w)$ and $p(a|w,x)$ are nonnegative - to do so, the variation is over the set of values that fulfill this constraint.

In order to maximize the variational problem, the derivative of $L(p(x|w), p(a|w,x))$ with respect to one particular $p(\tilde{x}|\tilde{w})$ has to equal zero and is given by

$$\begin{aligned}\frac{\partial L(p(x|w),p(a|w,x))}{\partial p(\tilde{x}|\tilde{w})} = \; & p(\tilde{w})\sum_a p(a|\tilde{w},\tilde{x})U(\tilde{w},a) \\ & - \frac{1}{\beta}_1 \underbrace{\left(p(\tilde{w})\log\frac{p(\tilde{x}|\tilde{w})}{p(\tilde{x})}\right)}_{=\frac{\partial}{\partial p(\tilde{x}|\tilde{w})}I(W;X)} \\ & - \frac{1}{\beta}_2 \underbrace{\left(p(\tilde{w})\sum_a p(a|\tilde{w},\tilde{x})\log\frac{p(a|\tilde{x})}{p(a)} - p(\tilde{w})\right)}_{=\frac{\partial}{\partial p(\tilde{x}|\tilde{w})}I(X;A)} \\ & - \frac{1}{\beta}_3 \underbrace{\left(p(\tilde{w})\sum_a p(a|\tilde{w},\tilde{x})\log\frac{p(a|\tilde{w},\tilde{x})}{p(a|\tilde{x})}\right)}_{=\frac{\partial}{\partial p(\tilde{x}|\tilde{w})}I(W;A|X)} \\ & + \lambda_1(\tilde{w}) = 0,\end{aligned}$$

The partial derivatives of the three mutual information terms are non-trivial and are explained in more detail in Supplementary Section 1.1. Resolving the upper equation with respect to $p(\tilde{x}|\tilde{w})$ leads to the following expression

$$p(\tilde{x}|\tilde{w}) = p(\tilde{x})\exp\left(\beta_1 T(\tilde{w},\tilde{x}) + \beta_1\frac{1}{\beta_2} + \beta_1\frac{\lambda_1(\tilde{w})}{p(\tilde{w})}\right), \tag{1}$$

where $T(\tilde{w},\tilde{x}) = \sum_a p(a|\tilde{w},\tilde{x})\left(U(\tilde{w},a) - \frac{1}{\beta_2}\log\frac{p(a|\tilde{x})}{p(a)} - \frac{1}{\beta_3}\log\frac{p(a|\tilde{w},\tilde{x})}{p(a|\tilde{x})}\right)$. Plugging Equation (1) into the equality constraint $\sum_x p(x|\tilde{w}) = 1$ yields

$$\beta_1\frac{1}{\beta_2} + \beta_1\frac{\lambda_1(\tilde{w})}{p(\tilde{w})} = -\log\sum_x p(x)\exp(\beta_1 T(\tilde{w},x)). \tag{2}$$

Inserting the result from Equation (2) back into Equation (1) leads to the final result in Equations (43) and (44) from the main paper:

$$p(\tilde{x}|\tilde{w}) = \frac{p(\tilde{x})\exp(\beta_1 T(\tilde{w},\tilde{x}))}{\sum_x p(x)\exp(\beta_1 T(\tilde{w},x))}. \tag{3}$$

Note that the utility-term $T(\tilde{w},\tilde{x})$ can be expressed in two different ways that both allow for different interpretations. We show both forms in Supplementary Section 1.2.

The derivative of $L(p(x|w), p(a|w,x))$ with respect to one particular $p(\tilde{a}|\tilde{w},\tilde{x})$ is given by

$$\begin{aligned}\frac{\partial L(p(x|w),p(a|w,x))}{\partial p(\tilde{a}|\tilde{w},\tilde{x})} = \; & p(\tilde{w})p(\tilde{x}|\tilde{w})U(\tilde{w},\tilde{a}) \\ & - \frac{1}{\beta_2}\underbrace{\left(p(\tilde{w})p(\tilde{x}|\tilde{w})\log\frac{p(\tilde{a}|\tilde{x})}{p(\tilde{a})}\right)}_{=\frac{\partial}{\partial p(\tilde{a}|\tilde{w},\tilde{x})}I(X;A)} \\ & - \frac{1}{\beta_3}\underbrace{\left(p(\tilde{w})p(\tilde{x}|\tilde{w})\log\frac{p(\tilde{a}|\tilde{w},\tilde{x})}{p(\tilde{a}|\tilde{x})}\right)}_{=\frac{\partial}{\partial p(\tilde{a}|\tilde{w},\tilde{x})}I(W;A|X)} \\ & + \lambda_2(\tilde{w},\tilde{x}) = 0,\end{aligned}$$

where the partial derivatives of mutual information terms follow similar rules as outlined in Supplementary Section 1.1. Resolving with respect to $p(\tilde{a}|\tilde{w},\tilde{x})$ leads to

$$p(\tilde{a}|\tilde{w},\tilde{x}) = p(\tilde{a}|\tilde{x})\exp\left(\beta_3 T(\tilde{w},\tilde{x},\tilde{a}) + \beta_3\frac{\lambda_2(\tilde{w},\tilde{x})}{p(\tilde{w})p(\tilde{x}|\tilde{w})}\right), \tag{4}$$

where we defined $T(\tilde{w},\tilde{x},\tilde{a}) = U(\tilde{w},\tilde{a}) - \frac{1}{\beta_2}\log\frac{p(\tilde{a}|\tilde{x})}{p(\tilde{a})}$. Inserting Equation (4) into the equality constraint $\sum_a p(a|\tilde{w},\tilde{x}) = 1$ yields

$$\beta_3\frac{\lambda_2(\tilde{w},\tilde{x})}{p(\tilde{w})p(\tilde{x}|\tilde{w})} = -\log\sum_a p(a|\tilde{x})\exp(\beta_3 T(\tilde{w},\tilde{x},a)). \tag{5}$$

Plugging the result from Equation (5) back into Equation (4) leads to the final result in Equation (40) from the main paper:

$$p(\tilde{a}|\tilde{w},\tilde{x}) = \frac{p(\tilde{a}|\tilde{x})\exp(\beta_3 T(\tilde{w},\tilde{x},\tilde{a}))}{\sum_a p(a|\tilde{x})\exp(\beta_3 T(\tilde{w},\tilde{x},a))}.$$

Note that it is not possible to express the Lagrange multipliers $\frac{1}{\beta_1}$, $\frac{1}{\beta_2}$ and $\frac{1}{\beta_3}$ in closed analytic form which is why they are treated as hyperparameters. The unconstrained optimization problem may hence be simply formalized as

$$\underset{p(x|w),p(a|w,x)}{\arg\max}\; \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1}I(W;X) - \frac{1}{\beta_2}I(X;A) - \frac{1}{\beta_3}I(W;A|X), \tag{6}$$

which does not require to specify upper bounds R_1, R_2 and R_3 on mutual information terms explicitly. This way of formulating the problem corresponds to our formulation in Equation (37) from the main text.

1.1 Computing partial derivatives of mutual information terms

$$\begin{aligned}\frac{\partial}{\partial p(\tilde{x}|\tilde{w})} I(W;X) = \; & p(\tilde{w}) \log \frac{p(\tilde{x}|\tilde{w})}{p(\tilde{x})} \\ & + \underbrace{p(\tilde{w})p(\tilde{x}|\tilde{w}) \frac{\partial}{\partial p(\tilde{x}|\tilde{w})} \log p(\tilde{x}|\tilde{w})}_{= p(\tilde{w})p(\tilde{x}|\tilde{w}) \frac{1}{p(\tilde{x}|\tilde{w})} = p(\tilde{w})} \\ & - \underbrace{\sum_w p(w)p(\tilde{x}|w) \frac{\partial}{\partial p(\tilde{x}|\tilde{w})} \log p(\tilde{x})}_{= \sum_w p(w)p(\tilde{x}|w) \frac{1}{p(\tilde{x})} p(\tilde{w}) = p(\tilde{w})} \\ = \; & p(\tilde{w}) \log \frac{p(\tilde{x}|\tilde{w})}{p(\tilde{x})}\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial p(\tilde{x}|\tilde{w})} I(X;A) = \; & p(\tilde{w}) \sum_a p(a|\tilde{w},\tilde{x}) \log \frac{p(a|\tilde{x})}{p(a)} \\ & + \underbrace{\sum_w p(w)p(\tilde{x}|w) \sum_a p(a|w,\tilde{x}) \frac{\partial}{\partial p(\tilde{x}|\tilde{w})} (\log p(\tilde{x},a) - \log p(\tilde{x}))}_{\sum_w p(w)p(\tilde{x}|w) \sum_a p(a|w,\tilde{x}) \left(\frac{1}{p(\tilde{x},a)} p(\tilde{w})p(a|\tilde{w},\tilde{x}) - \frac{1}{p(\tilde{x})} p(\tilde{w}) \right) = p(\tilde{w}) \sum_a p(\tilde{x},a) \left(\frac{1}{p(\tilde{x},a)} p(a|\tilde{w},\tilde{x}) - \frac{1}{p(\tilde{x})} \right) = 0} \\ & - \underbrace{\sum_w p(w) \sum_x p(x|w) \sum_a p(a|w,x) \frac{\partial}{\partial p(\tilde{x}|\tilde{w})} \log p(a)}_{\sum_w p(w) \sum_x p(x|w) \sum_a p(a|w,x) \frac{1}{p(a)} p(\tilde{w})p(a|\tilde{w},\tilde{x}) = p(\tilde{w}) \sum_a p(a) \frac{1}{p(a)} p(a|\tilde{w},\tilde{x}) = p(\tilde{w})} \\ = \; & p(\tilde{w}) \sum_a p(a|\tilde{w},\tilde{x}) \log \frac{p(a|\tilde{x})}{p(a)} - p(\tilde{w})\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial p(\tilde{x}|\tilde{w})} I(W;A|X) = \; & p(\tilde{w}) \sum_a p(a|\tilde{w},\tilde{x}) \log \frac{p(a|\tilde{w},\tilde{x})}{p(a|\tilde{x})} \\ & - \underbrace{\sum_w p(w)p(\tilde{x}|w) \sum_a p(a|w,\tilde{x}) \frac{\partial}{\partial p(\tilde{x}|\tilde{w})} (\log p(\tilde{x},a) - \log p(\tilde{x}))}_{\sum_w p(w)p(\tilde{x}|w) \sum_a p(a|w,\tilde{x}) \left(\frac{1}{p(\tilde{x},a)} p(\tilde{w})p(a|\tilde{w},\tilde{x}) - \frac{1}{p(\tilde{x})} p(\tilde{w}) \right) = p(\tilde{w}) \sum_a p(\tilde{x},a) \left(\frac{1}{p(\tilde{x},a)} p(a|\tilde{w},\tilde{x}) - \frac{1}{p(\tilde{x})} \right) = 0} \\ = \; & p(\tilde{w}) \sum_a p(a|\tilde{w},\tilde{x}) \log \frac{p(a|\tilde{w},\tilde{x})}{p(a|\tilde{x})}\end{aligned}$$

1.2 Two ways of expressing the utility-term $T(\tilde{w}, \tilde{x})$

In the result in Equation (3), the term $T(\tilde{w}, \tilde{x})$ plays the role of a utility-term that is maximized in a bounded-rational fashion. It is given by

$$T(\tilde{w}, \tilde{x}) = \sum_a p(a|\tilde{w}, \tilde{x}) \left(U(\tilde{w}, a) - \frac{1}{\beta_2} \log \frac{p(a|\tilde{x})}{p(a)} - \frac{1}{\beta_3} \log \frac{p(a|\tilde{w}, \tilde{x})}{p(a|\tilde{x})} \right)$$

which can be written as:

$$T(\tilde{w}, \tilde{x}) = \Delta F_{\text{par}}(\tilde{w}, \tilde{x}) - \frac{1}{\beta_2} \sum_a p(a|\tilde{w}, \tilde{x}) \log \frac{p(a|\tilde{x})}{p(a)} \tag{7}$$

where $\Delta F_{\text{par}}(\tilde{w}, \tilde{x})$ is the free energy difference of the action stage:

$$\Delta F_{\text{par}}(\tilde{w}, \tilde{x}) := \mathbf{E}_{p(a|\tilde{w},\tilde{x})}[U(\tilde{w}, a)] - \frac{1}{\beta_3} D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a|\tilde{x})).$$

The second term in Equation (7) looks almost like a KL-divergence. It can be multiplied with $\frac{p(a|\tilde{w},\tilde{x})}{p(a|\tilde{w},\tilde{x})}$ and then decomposed into two KL-divergences

$$\begin{aligned}
&- \frac{1}{\beta_2} \sum_a p(a|\tilde{w}, \tilde{x}) \log \left(\frac{p(a|\tilde{x})}{p(a)} \frac{p(a|\tilde{w}, \tilde{x})}{p(a|\tilde{w}, \tilde{x})} \right) = \\
&- \frac{1}{\beta_2} \sum_a p(a|\tilde{w}, \tilde{x}) \left(\log \frac{p(a|\tilde{x})}{p(a|\tilde{w}, \tilde{x})} + \log \frac{p(a|\tilde{w}, \tilde{x})}{p(a)} \right) = \\
&- \frac{1}{\beta_2} \left(\sum_a p(a|\tilde{w}, \tilde{x}) \log \frac{p(a|\tilde{w}, \tilde{x})}{p(a)} - \sum_a p(a|\tilde{w}, \tilde{x}) \log \frac{p(a|\tilde{w}, \tilde{x})}{p(a|\tilde{x})} \right) = \\
&- \frac{1}{\beta_2} \left(D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a)) - D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a|\tilde{x})) \right).
\end{aligned}$$

Then Equation (7) can be rewritten as

$$T(\tilde{w}, \tilde{x}) = \Delta F_{\text{gen}}(\tilde{w}, \tilde{x}) - \left(\frac{1}{\beta_3} - \frac{1}{\beta_2} \right) D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a|\tilde{x})), \tag{8}$$

with

$$\Delta F_{\text{gen}}(\tilde{w}, \tilde{x}) := \mathbf{E}_{p(a|\tilde{w},\tilde{x})}[U(\tilde{w}, a)] - \frac{1}{\beta_2} D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a)),$$

which corresponds to the way the utility term $T(\tilde{w}, \tilde{x})$ is expressed in Equation (38) from the main paper. From Equation (7) it can easily be seen that by $\beta_2 \to \infty$ the solution of the general case becomes equal to the corresponding solution of the hierarchical case:

$$T(\tilde{w}, \tilde{x}) = \Delta F_{\text{par}}(\tilde{w}, \tilde{x}). \tag{9}$$

In Equation (8) it can be seen that the utility-term consists of the free energy $\Delta F_{\text{gen}}(\tilde{w}, \tilde{x})$ and $D_{\text{KL}}(p(a|\tilde{w}, \tilde{x})||p(a|\tilde{x}))$ which is either considered as a cost (for $\beta_2 > \beta_3$) or a gain (for $\beta_2 < \beta_3$).

2 VARIATIONAL PROBLEM - SERIAL CASE

The variational problem in Section 3 of the main paper is given by

$$\begin{aligned}\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] \quad &\text{subject to} \quad I(W;X) \leq R_1,\ I(X;A) \leq R_2 \\ &\text{and} \qquad \forall w\ \sum_x p(x|w) = 1,\ \forall x\ \sum_a p(a|x) = 1,\end{aligned}$$

where R_1 and R_2 denote given upper bounds on the respective mutual information terms. Translating this constrained optimization problem into an unconstrained optimization problem yields the corresponding Lagrangian $L(p(x|w), p(a|x))$

$$\begin{aligned}L(p(x|w),p(a|x)) = &\sum_{w,x,a} p(w)p(x|w)p(a|x)U(w,a) \\ &+ \frac{1}{\beta}_1 \left(R_1 - \sum_w p(w) \sum_x p(x|w) \log \frac{p(x|w)}{p(x)} \right) \\ &+ \frac{1}{\beta}_2 \left(R_2 - \sum_w p(w) \sum_x p(x|w) \sum_a p(a|x) \log \frac{p(a|x)}{p(a)} \right) \\ &+ \sum_w \lambda_1(w) \left(\sum_x p(x|w) - 1 \right) + \sum_x \lambda_2(x) \left(\sum_a p(a|x) - 1 \right),\end{aligned}$$

with Lagrange multipliers $\frac{1}{\beta_1}$, $\frac{1}{\beta_2}$, $\lambda_1(w)$ and $\lambda_2(x)$. Optimal solutions can then be found by following similar steps as outlined in Supplementary Section 1.

Note, that it is not possible to find closed form analytic solutions for the Lagrange multipliers $\frac{1}{\beta_1}$ and $\frac{1}{\beta_2}$. Therefore they are treated as hyperparameters, hence allowing to express the optimization problem as

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta}_1 I(W;X) - \frac{1}{\beta}_2 I(X;A), \tag{10}$$

which renders the explicit specification of mutual information upper bounds R_1 and R_2 needless and is our preferred way of formulating the problem in the main text—see Equation (14) there.

3 VARIATIONAL PROBLEM - PARALLEL CASE

The variational problem in Section 4 of the main paper is given by

$$\begin{aligned}\underset{p(x|w),p(a|w,x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] \quad &\text{subject to} \quad I(W;X) \leq R_1,\ I(W;A|X) \leq R_3 \\ &\text{and} \qquad \forall w\ \sum_x p(x|w) = 1,\ \forall w,x\ \sum_a p(a|w,x) = 1,\end{aligned}$$

where R_1 and R_3 denote given upper bounds on the respective mutual information terms. Translating this constrained optimization problem into an unconstrained optimization problem yields the corresponding

Lagrangian $L(p(x|w), p(a|w,x))$

$$\begin{aligned} L(p(x|w), p(a|w,x)) = & \sum_{w,x,a} p(w)p(x|w)p(a|w,x)U(w,a) \\ & + \frac{1}{\beta_1}\left(R_1 - \sum_w p(w)\sum_x p(x|w)\log\frac{p(x|w)}{p(x)}\right) \\ & + \frac{1}{\beta_3}\left(R_3 - \sum_w p(w)\sum_x p(x|w)\sum_a p(a|w,x)\log\frac{p(a|w,x)}{p(a|x)}\right) \\ & + \sum_w \lambda_1(w)\left(\sum_x p(x|w) - 1\right) + \sum_{w,x}\lambda_2(w,x)\left(\sum_a p(a|w,x) - 1\right), \end{aligned}$$

with Lagrange multipliers $\frac{1}{\beta_1}$, $\frac{1}{\beta_3}$, $\lambda_1(w)$ and $\lambda_2(w,x)$. Optimal solutions can then be found by following similar steps as outlined in Supplementary Section 1 or alternatively by modifying the solution of the general case given in Equations (38) to (42) in the main text by taking the limit $\beta_2 \to \infty$.

Note, that it is not possible to find closed analytic form solutions for the Lagrange multipliers $\frac{1}{\beta_1}$ and $\frac{1}{\beta_3}$. Therefore they are treated as hyperparameters, hence allowing to express the optimization problem as

$$\arg\max_{p(x|w,a),p(a|x)} \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_3} I(W;A|X), \tag{11}$$

which renders the explicit specification of mutual information upper bounds R_1 and R_3 needless and is our preferred way of formulating the problem in the main text—see Equation (25) there.

4 DEGENERATE TOTAL CORRELATION AND TOTAL CORRELATION

With $\beta_2 = \beta_3 = \beta$, the unconstrained objective of the general case (Equation (37) in the main manuscript) reduces to the objective of the degenerate total correlation case

$$\arg\max_{p(x|w),p(a|w,x)} \mathbf{E}_{p(w,x,a)}[U(w,x,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta} I(W,X;A), \tag{12}$$

where we made use of a utility function that explicitly depends on x as well, not only on w and a. The two corresponding mutual information terms that are considered as information processing costs have the following form:

$$\begin{aligned} I(W;X) &= \sum_w p(w)\sum_x p(x|w)\log\frac{p(x|w)}{p(x)}, \\ I(W,X;A) &= \sum_w p(w)\sum_x p(x|w)\sum_a p(a|w,x)\log\frac{p(a|w,x)}{p(a)}. \end{aligned}$$

This objective implies information processing cost for changing from $p(x)$ to $p(x|w)$ and a different cost for changing from $p(a)$ to $p(a|w,x)$. Compared to the general case, the information processing costs for changing from $p(a)$ to $p(a|x)$ and the cost for changing from $p(a|x)$ to $p(a|w,x)$ do not need to be specified, but rather the "overall" cost for changing from $p(a)$ to $p(a|w,x)$.

By inserting $\beta_2 = \beta_3 = \beta$ in Equations (38) to (42) from the main paper, we obtain the following solution

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp\left(\beta_1 \Delta F_{\text{gen}}(w,x)\right) \tag{13}$$

$$p(x) = \sum_w p(w) p^*(x|w) \tag{14}$$

$$p^*(a|w,x) = \frac{1}{Z(w,x)} p^*(a|x) \exp\left(\beta U(w,x,a) - \log \frac{p^*(a|x)}{p(a)}\right) = \frac{1}{Z(w,x)} p(a) \exp\left(\beta U(w,x,a)\right) \tag{15}$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|w,x), \tag{16}$$

which does no longer require the explicit representation of $p^*(a|x)$. $Z(w)$ and $Z(w,x)$ denote the corresponding normalization constants or partition sums and $\Delta F_{\text{gen}}(w,x)$ is given according to the main paper:

$$\Delta F_{\text{gen}}(w,x) := \mathbf{E}_{p^*(a|w,x)}[U(w,x,a)] - \frac{1}{\beta} D_{\text{KL}}(p^*(a|w,x)||p(a)).$$

The solution of the total correlation case can be easily obtained from the equations above by setting $\beta_1 = \beta$.

5 DISTRIBUTING COMPUTATION ON THE DIFFERENT INFORMATION PROCESSING PATHWAYS OF THE PARALLEL HIERARCHICAL CASE

In the parallel hierarchical case (Section 4 in the main paper), there are two possible pathways from w to a:

$$\begin{aligned} \text{Two-stage serial pathway} \quad & I(W;X) \to I(X;A) \\ \text{Parallel pathway} \quad & I(W;A|X) \end{aligned}$$

and the total computational load can either be split up between both pathways or be shifted to either pathway exclusively, depending on the information processing prices β_1 and β_3 (see Section 4.3 in the main manuscript). The same also holds true for the general case. The terms appear the following way in the objective of the parallel hierarchical case (Equation (25) in the main manuscript):

$$\underset{p(x|w),p(a|w,x)}{\arg\max} \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_3} I(W;A|X) = \underset{p(x|w),p(a|w,x)}{\arg\max} J_{\text{par}}(p(a|w,x)), p(x|w)). \tag{17}$$

Here, we show how the different temperature-settings lead to different utilization of the parallel and serial pathway and how the parallel case is related to the one-step (rate distortion) case.

5.1 $\beta_1 = \beta_3$ - joint action (x, a)

Consider the original variational problem in the one-step case (see Equation (7) in the main manuscript):

$$\underset{p(a'|w)}{\arg\max} \mathbf{E}_{p(w,a')}[U(w,a')] - \frac{1}{\beta} I(W;A').$$

If the random variable A' is replaced with a joint-variable (X, A), implying $p(a') = p(x,a)$ the following factorization can be used: $p(a'|w) = p((x,a)|w) = p(x|w)p(a|w,x)$. Applying the chain rule for the

mutual information allows to split the mutual information term:

$$\frac{1}{\beta}I(W;A') = \frac{1}{\beta}I(W;(X,A)) = \frac{1}{\beta}I(W;X) + \frac{1}{\beta}I(W;A|X).$$

The two terms $I(W;X)$ and $I(W;A|X)$ appear exactly as the computational effort in the objective of the parallel case (Equation (17) but with different temperatures β_1 and β_3 instead of β. The equation above therefore makes it obvious that if the temperatures β_1 and β_3 in Equation (17) of the main manuscript are equal, then the parallel case becomes mathematically equivalent to the one-step case. Note that in this case it does not matter whether the information is processed exclusively on the parallel pathway through $p(a|w,x)$ or exclusively on the serial pathway through $p(x|w)$ and $p(a|x)$ or through any mixture in-between. All solutions are equal in terms of expected utility and value of the objective $J_{\text{par}}(p(a|w,x)), p(x|w))$. However, empirically we find that iterating the self-consistent equations in the case that both temperatures are equal favors solutions where $I(W;A|X) = 0$.

5.2 $\beta_1 < \beta_3$ - parallel pathway only

If the price of model selection $\frac{1}{\beta}_1$ is larger than the price of processing on the lower level of the hierarchy $\frac{1}{\beta}_3$, then all computation is shifted to the lower level of the hierarchy, thus not requiring a model selection mechanism anymore ($I(W;X) = 0$) and the models are rendered obsolete by $p(a|x) = p(a)\ \forall x$ (implying $I(X;A) = 0$). Additionally $p(a|w,x) = p(a|w)\ \forall x$ since x does not carry any useful information. The cost that arises through $I(W;A|X)$ then becomes the average KL-divergence between $p(a)$ and $p(a|w)$ which is exactly identical to the one-step case. Thus, for $\beta_1 < \beta_3$ the upper level of the hierarchy is not used and the one-step case (Section 2.2 in the main manuscript) is recovered. This is not shown in any of the examples in the paper but can be explored in the Supplementary Jupyter Notebook "4-ParallelHierarcy".

5.3 $\beta_1 > \beta_3$ - serial pathway only

If the price for information processing on the low level of the hierarchy $\frac{1}{\beta}_3$ is larger than the price of model selection $\frac{1}{\beta}_1$, all computation is shifted to the model selection and the models. This means that $I(W;A|X) = 0$ which is achieved by $p(a|w,x) = p(a|x)\ \forall w$ meaning that the models $p(a|x)$ become the final policy. Note that this reproduces the same structure as in the serial case (Section 3 in the main manuscript) where information has to pass the two serial stages $I(W;X)$ and $I(X;A)$. The crucial difference is that there is no cost for $I(X;A)$ in the parallel case which corresponds to $\beta_2 \to \infty$ in the serial case, meaning that processing on the second stage of the sequence is for free. In this setting both the parallel and the serial case are exactly equivalent to the one-step case in terms of expected utility and value of the respective objective functions. However, the conceptual difference is that in the serial case $p(a)$ is considered a prior and there is a cost for computing $p(a|x)$. In the parallel case $p(a|x)$ plays the role of a prior and there is a cost for computing $p(a|w,x)$. This elegantly hides the implicit assumption that there is no cost for going from $p(a)$ to $p(a|x)$ in the parallel case, but this computation is implicitly carried out during iteration of the self-consistent equations. In the parallel case the result of this computation can be stored in $p(a|x)$ and can then be re-used without having to perform the computation again. In the serial case the assumption is that only $p(a)$ can be stored and computing $p(a|x)$ incurs a computational cost. It would be equally valid to assume that $p(a|x)$ can be stored in the serial case as well which would correspond to setting $\beta_2 \to \infty$ under which condition the two cases actually become identical.

5.4 $\beta_1 > \beta_3$ and limited cardinality $|\mathcal{X}|$ - both pathways simultaneously

If $\beta_1 > \beta_3$ it is cheaper to move all processing to the serial pathway. However, if the capacity of the serial pathway (that is the maximally possible rates $I(W;X$ and $I(X;A))$ is limited through a low cardinality $|\mathcal{X}|$, it is most economical to fully use the capacity of the serial pathway and then use the parallel pathway to take on additional computational load (see Main Manuscript Section 4.3 for more details).

6 EXPONENTIAL OF KULLBACK-LEIBLER DIVERGENCE IS A LOWER BOUND FOR THE SAMPLING COMPLEXITY

In Equation (6) of Section 2.1 in the main manuscript the average number of samples to draw from a prior $p_0(a)$ in order to obtain one sample from $p^*(a|w)$ (according to Equation (4) in the main manuscript) with a rejection-sampling scheme is given as

$$\overline{\#Samples}(w) = \frac{\exp(\beta T(w))}{Z(w)}. \tag{18}$$

From thermodynamics it is known that the partition sum $Z(w)$ and the (negative) free energy difference $\Delta F(w)$ are related through

$$\Delta F(w) = \frac{1}{\beta} \log Z(w),$$

where $\Delta F(w) = \mathbf{E}_{p^*(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}(p^*(a|w)||p_0(a))$(as given by Equation (5) in the main manuscript). Rewriting to $Z(w) = \exp(\beta \Delta F(w))$ and plugging into Equation (18) yields the final inequality

$$\begin{aligned}\overline{\#Samples}(w) &= \exp(\beta(T(w) - \Delta F(w))) \\ &= \underbrace{\exp\left(\beta(T(w) - \mathbf{E}_{p^*(a|w)}[U(w,a)])\right)}_{\geq 1} \cdot \exp\left(D_{\mathrm{KL}}(p^*(a|w)||p_0(a))\right)\end{aligned}$$

where the second line uses the definition of the aspiration value $T(w) \geq \max_a U(w,a)$ (see main manuscript). This finally leads to

$$\overline{\#Samples}(w) \geq \exp\left(D_{\mathrm{KL}}(p^*(a|w)||p_0(a))\right) \tag{19}$$

Bibliography

[Abe and Watanabe, 2011] Abe, K. and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature neuroscience*, 14(8):1067–1074.

[Abeele and Bock, 2001] Abeele, S. and Bock, O. (2001). Mechanisms for sensorimotor adaptation to rotated visual input. *Experimental Brain Research*, 139(2):248–253.

[Acuna and Schrater, 2010] Acuna, D. E. and Schrater, P. (2010). Structure learning in human sequential decision-making. *PLoS Computational Biology*, 6(12):12.

[Adolph, 2008] Adolph, K. E. (2008). Learning to move. *Current Directions in Psychological Science*, 17(3):213–218.

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.

[Anderson, 1991] Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409.

[Arimoto, 1972] Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *Information Theory, IEEE Transactions on*, 18(1):14–20.

[Ashby and Maddox, 2005] Ashby, F. G. and Maddox, W. T. (2005). Human category learning. *Annual Reviews Psychology*, 56:149–178.

[Astrom and Wittenmark, 1994] Astrom, K. J. and Wittenmark, B. (1994). *Adaptive Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition.

[Bailey and Thomas, 2001] Bailey, A. M. and Thomas, R. K. (2001). The effects of nucleus basalis magnocellularis lesions in long-evans hooded rats on two learning set formation tasks, delayed matching-to-sample learning, and open-field activity. *Behavioral Neuroscience*, 115(2):328.

[Bathellier et al., 2012] Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2):435–449.

[Beckers et al., 2012] Beckers, G. J., Bolhuis, J. J., Okanoya, K., and Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport*, 23(3):139–145.

[Bellman, 1957] Bellman, R. (1957). Dynamic programming. *Princeton University Press, Princeton, NJ*.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, volume 4. Springer.

[Blahut, 1972] Blahut, R. E. (1972). Computation of channel capacity and rate-distortion functions. *Information Theory, IEEE Transactions on*, 18(4):460–473.

[Bock et al., 2001] Bock, O., Schneider, S., and Bloomberg, J. (2001). Conditions for interference versus facilitation during sequential sensorimotor adaptation. *Experimental Brain Research*, 138(3):359–365.

[Botvinick, 2008] Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in cognitive sciences*, 12(5):201–208.

[Botvinick, 2012] Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current opinion in neurobiology*, 22(6):956–962.

[Botvinick et al., 2009] Botvinick, M. M., Niv, Y., and Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3):262–280.

[Braun et al., 2009a] Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009a). Learning optimal adaptation strategies in unpredictable motor tasks. *The Journal of Neuroscience*, 29(20):6472–6478.

[Braun et al., 2009b] Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009b). Motor task variation induces structural learning. *Current Biology*, 19(4):352–357.

[Braun et al., 2010a] Braun, D. A., Mehring, C., and Wolpert, D. M. (2010a). Structure learning in action. *Behavioural Brain Research*, 206(2):157–165.

[Braun et al., 2011] Braun, D. A., Ortega, P. A., Theodorou, E., and Schaal, S. (2011). Path integral control and bounded rationality. In *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on*, pages 202–209. IEEE.

[Braun et al., 2010b] Braun, D. A., Waldert, S., Aertsen, A., Wolpert, D. M., and Mehring, C. (2010b). Structure learning in a sensorimotor association task. *PLoS ONE*, 5(1):8.

[Brown and Kane, 1988] Brown, A. L. and Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, 20(4):493–523.

[Carey, 1985] Carey, S. (1985). *Conceptual Change in Childhood*. MIT Press, Cambridge MA.

[Carey and Bartlett, 1978] Carey, S. and Bartlett, E. (1978). Acquiring a single new word.

[Chater et al., 2006] Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences*, 10(7):287–291.

[Cheke et al., 2011] Cheke, L. G., Bird, C. D., and Clayton, N. S. (2011). Tool-use and instrumental learning in the eurasian jay (garrulus glandarius). *Animal cognition*, 14(3):441–455.

[Chomsky, 1965] Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT press, Cambridge MA.

[Cohen et al., 2005] Cohen, H. S., Bloomberg, J. J., and Mulavara, A. P. (2005). Obstacle avoidance in novel visual environments improved by variable practice training. *Perceptual and motor skills*, 101(3):853–861.

[Cover and Thomas, 1991] Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons, Hoboken, NJ.

[Daniel et al., 2012] Daniel, C., Neumann, G., and Peters, J. (2012). Hierarchical relative entropy policy search. In *International Conference on Artificial Intelligence and Statistics*.

[Daniel et al., 2013] Daniel, C., Neumann, G., and Peters, J. (2013). Autonomous reinforcement learning with hierarchical REPS. In *International Joint Conference on Neural Networks*.

[Davidson and Wolpert, 2003] Davidson, P. R. and Wolpert, D. M. (2003). Motor learning and prediction in a variable environment. *Current opinion in neurobiology*, 13(2):232–237.

[DiCarlo et al., 2012] DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3):415–434.

[Diedrichsen, 2007] Diedrichsen, J. (2007). Optimal task-dependent changes of bimanual feedback control and adaptation. *Current Biology*, 17(19):1675–1679.

[Diedrichsen et al., 2010] Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2010). The coordination of movement: optimal feedback control and beyond. *Trends in cognitive sciences*, 14(1):31–39.

[Doya, 2007] Doya, K. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press, Cambridge MA.

[Druckmann and Chklovskii, 2012] Druckmann, S. and Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, 22(22):2095–2103.

[Duncan, 1960] Duncan, C. P. (1960). Description of learning to learn in human subjects. *The American journal of psychology*, 73(1):108–114.

[Fox et al., 2015] Fox, R., Pakman, A., and Tishby, N. (2015). G-learning: Taming the noise in reinforcement learning via soft updates. *arXiv preprint arXiv:1512.08562*.

[Franklin and Wolpert, 2011] Franklin, D. W. and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3):425–442.

[Friston, 2002] Friston, K. (2002). Functional integration and inference in the brain. *Progress in neurobiology*, 68(2):113–143.

[Friston, 2005] Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456):815–836.

[Friston, 2010] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

[FUJITA, 1983] FUJITA, K. (1983). Acquisition and transfer of a higher-order conditional discrimination performance in the japanese monkey. *Japanese psychological research*, 25(1):1–8.

[Garner et al., 2006] Garner, J. P., Thogerson, C. M., Würbel, H., Murray, J. D., and Mench, J. A. (2006). Animal neuropsychology: validation of the intra-dimensional extra-dimensional set shifting task for mice. *Behavioural brain research*, 173(1):53–61.

[Genewein and Braun, 2012] Genewein, T. and Braun, D. A. (2012). A sensorimotor paradigm for bayesian model selection. *Frontiers in Human Neuroscience*, 6:291.

[Genewein and Braun, 2013] Genewein, T. and Braun, D. A. (2013). Abstraction in decision-makers with limited information processing capabilities. *NIPS 2013 workshop on planning with information constraints, arXiv:1312.4353*.

[Genewein and Braun, 2014] Genewein, T. and Braun, D. A. (2014). Occam's razor in sensorimotor learning. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1783).

[Genewein et al., 2015a] Genewein, T., Hez, E., Razzaghpanah, Z., and Braun, D. A. (2015a). Structure learning in bayesian sensorimotor integration. *PLoS Computational Biology*, 11(8):e1004369.

[Genewein et al., 2015b] Genewein, T., Leibfried, F., Grau-Moya, J., and Braun, D. A. (2015b). Bounded rationality, abstraction and hierarchical decision-making: an information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2:27.

[Gershman and Niv, 2010] Gershman, S. J. and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, 20(2):251–256.

[Gershman et al., 2009] Gershman, S. J., Pesaran, B., and Daw, N. D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *The Journal of Neuroscience*, 29(43):13524–13531.

[Gershman et al., 2015] Gershman, S. J., Tenenbaum, J. B., and Jäkel, F. (2015). Discovering hierarchical motion structure. *Vision research*.

[Gigerenzer et al., 2008] Gigerenzer, G., Hoffrage, U., and Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to dougherty, franco-watkins, and thomas (2008).

[Gold and Shadlen, 2007] Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574.

[Goodman, 1955] Goodman, N. (1955). *Fact, fiction and forecast*. Harvard University Press, Cambridge, MA.

[Gopnik et al., 2004] Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological review*, 111(1):3.

[Grau-Moya et al., 2016] Grau-Moya, J., Leibfried, F., Genewein, T., and Braun, D. A. (2016). Planning with information-processing constraints and model uncertainty in markov decision processes. *arXiv preprint arXiv:1604.02080*.

[Griffiths et al., 2010] Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, 14(8):357–364.

[Griffiths and Tenenbaum, 2005] Griffiths, T. L. and Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, 51(4):334–384.

[Griffiths and Tenenbaum, 2006] Griffiths, T. L. and Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, 17(9):767–773.

[Halford et al., 1998] Halford, G. S., Bain, J. D., Maybery, M. T., and Andrews, G. (1998). Induction of relational schemas: Common processes in reasoning and complex learning. *Cognitive psychology*, 35(3):201–245.

[Harlow, 1949] Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56(1):51–65.

[Haruno et al., 2001] Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural computation*, 13(10):2201–2220.

[Hick, 1952] Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26.

[Hirsch and Spinelli, 1970] Hirsch, H. V. and Spinelli, D. (1970). Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168(3933):869–871.

[Hochstein and Ahissar, 2002] Hochstein, S. and Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804.

[Hultsch, 1974] Hultsch, D. F. (1974). Learning to learn in adulthood. *Journal of Gerontology*, 29(3):302–309.

[Hyman, 1953] Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of experimental psychology*, 45(3):188.

[Inhelder and Piaget, 1964] Inhelder, B. and Piaget, J. (1964). *The Early Growth of Logic in the Child: Classification and Seriation*. Routledge and Kegan Paul, London.

[Joel et al., 2002] Joel, D., Niv, Y., and Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks*, 15(4):535–547.

[Johnson et al., 2008] Johnson, J. S., Spencer, J. P., and Schöner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology*, 26(2):227–251.

[Jones and Smith, 2002] Jones, S. S. and Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, 5(2):219–232.

[Kappen, 2007] Kappen, H. J. (2007). An introduction to stochastic control theory, path integrals and reinforcement learning. In *Cooperative Behavior in Neural Systems: Ninth Granada Lectures*, volume 887, pages 149–181. AIP Publishing.

[Kappen et al., 2012] Kappen, H. J., Gómez, V., and Opper, M. (2012). Optimal control as a graphical model inference problem. *Machine Learning*, 87(2):159–182.

[Karniel and Mussa-Ivaldi, 2002] Karniel, A. and Mussa-Ivaldi, F. A. (2002). Does the motor control system use multiple models and context switching to cope with a variable environment? *Experimental Brain Research*, 143(4):520–524.

[Kemp, 2008] Kemp, C. (2008). *PhD Thesis*. Massachusetts Institute of Technology, Cambridge MA.

[Kemp et al., 2004] Kemp, C., Perfors, A., and Tenenbaum, J. B. (2004). Learning domain structures. *Proceedings of the 26th annual conference of the cognitive science society*, page 720–725.

[Kemp et al., 2007] Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental Science*, 10(3):307–321.

[Kemp and Tenenbaum, 2009] Kemp, C. and Tenenbaum, J. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, 116(1):20 –58.

[Kemp and Tenenbaum, 2008] Kemp, C. and Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*.

[Kersten et al., 2004] Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.*, 55:271–304.

[Knill and Pouget, 2004] Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719.

[Knill and Richards, 1996] Knill, D. C. and Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press, Cambridge UK.

[Kobak and Mehring, 2012] Kobak, D. and Mehring, C. (2012). Adaptation paths to novel motor tasks are shaped by prior structure learning. *The Journal of Neuroscience*, 32(29):9898–9908.

[Koechlin, 2016] Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current opinion in neurobiology*, 37:1–6.

[Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

[Körding et al., 2007] Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, 2(9):e943.

[Körding and Wolpert, 2004] Körding, K. P. and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247.

[Körding and Wolpert, 2006] Körding, K. P. and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, 10(7):319–326.

[Kousta, 2010] Kousta, S. (2010). Approaches to cognitive modeling. *Trends in cognitive sciences*, 14(8):339.

[Kwisthout and van Rooij, 2013] Kwisthout, J. and van Rooij, I. (2013). Predictive coding and the bayesian brain: Intractability hurdles that are yet to be overcome. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, Austin, TX: Cognitive Science Society*.

[Kwisthout et al., 2011] Kwisthout, J., Wareham, T., and van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, 35(5):779–784.

[Lagnado and Sloman, 2004] Lagnado, D. A. and Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4):856.

[Lagnado and Sloman, 2006] Lagnado, D. A. and Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(3):451.

[Lake et al., 2015] Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.

[Lake et al., 2016] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2016). Building machines that learn and think like people. *arXiv preprint arXiv:1604.00289*.

[Langbein et al., 2007] Langbein, J., Siebert, K., Nürnberg, G., and Manteuffel, G. (2007). Learning to learn during visual discrimination in group housed dwarf goats (capra hircus). *Journal of Comparative Psychology*, 121(4):447.

[Lee and Mumford, 2003] Lee, T. S. and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7):1434–1448.

[Leibfried and Braun, 2015] Leibfried, F. and Braun, D. A. (2015). A reward-maximizing spiking neuron as a bounded rational decision maker. *Neural computation*.

[Leibfried and Braun, 2016] Leibfried, F. and Braun, D. A. (2016). Bounded rational decision-making in feedforward neural networks. *arXiv preprint arXiv:1602.08332*.

[Liu and Pleskac, 2011] Liu, T. and Pleskac, T. J. (2011). Neural correlates of evidence accumulation in a perceptual decision task. *Journal of neurophysiology*, 106(5):2383–2398.

[Logothetis and Sheinberg, 1996] Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, 19(1):577–621.

[Lucas and Griffiths, 2010] Lucas, C. G. and Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical bayesian models. *Cognitive Science*, 34(1):113–147.

[Lucas et al., 2015] Lucas, C. G., Griffiths, T. L., Williams, J. J., and Kalish, M. L. (2015). A rational model of function learning. *Psychonomic bulletin & review*, 22(5):1193–1215.

[MacKay, 2003] MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press, Cambridge, UK.

[Mackintosh and Little, 1969] Mackintosh, N. and Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, 14(1):5–6.

[Markman, 1989] Markman, E. (1989). *Naming and categorization in children*. MIT Press, Cambridge MA.

[Marr, 1982] Marr, D. (1982). Vision: A computational approach.

[McClelland, 1998] McClelland, J. L. (1998). Connectionist models and bayesian inference. *Rational models of cognition*, pages 21–53.

[McClelland et al., 2010] McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., and Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences*, 14(8):348–356.

[McKenzie et al., 2014] McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., and Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1):202–215.

[Medina et al., 2001] Medina, J. F., Garcia, K. S., and Mauk, M. D. (2001). A mechanism for savings in the cerebellum. *The Journal of Neuroscience*, 21(11):4081–4089.

[Mulavara et al., 2009] Mulavara, A. P., Cohen, H. S., and Bloomberg, J. J. (2009). Critical features of training that facilitate adaptive generalization of over ground locomotion. *Gait & posture*, 29(2):242–248.

[Narain et al., 2013a] Narain, D., Mamassian, P., van Beers, R. J., Smeets, J. B., and Brenner, E. (2013a). How the statistics of sequential presentation influence the learning of structure. *PloS one*, 8(4):e62276.

[Narain et al., 2014] Narain, D., Smeets, J. B., Mamassian, P., Brenner, E., and van Beers, R. J. (2014). Structure learning and the occam's razor principle: A new view of human function acquisition. *Frontiers in computational neuroscience*, 8(Article 121).

[Narain et al., 2013b] Narain, D., van Beers, R. J., Smeets, J. B., and Brenner, E. (2013b). Sensorimotor priors in nonstationary environments. *Journal of neurophysiology*, 109(5):1259–1267.

[Navon, 1977] Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, 9(3):353–383.

[Novick and Hurley, 2001] Novick, L. R. and Hurley, S. M. (2001). To matrix, network, or hierarchy: That is the question. *Cognitive Psychology*, 42(2):158–216.

[Ortega and Braun, 2010] Ortega, P. A. and Braun, D. A. (2010). A minimum relative entropy principle for learning and acting. *Journal of Artificial Intelligence Research*, pages 475–511.

[Ortega and Braun, 2013] Ortega, P. A. and Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society A*, 469(2153).

[Ortega et al., 2015] Ortega, P. A., Braun, D. A., Dyer, J., Kim, K.-E., and Tishby, N. (2015). Information-theoretic bounded rationality. *arXiv preprint arXiv:1512.06789*.

[Perfors and Tenenbaum, 2009] Perfors, A. and Tenenbaum, J. (2009). Learning to learn categories. In *Annual Meeting of the Cognitive Science Society (31st: 2009: Amsterdam)*.

[Peters et al., 2010] Peters, J., Mülling, K., and Altun, Y. (2010). Relative entropy policy search. In *AAAI*.

[Pinto et al., 2010] Pinto, N., Majaj, N., Barhomi, Y., Solomon, E., and DiCarlo, J. (2010). Human versus machine: comparing visual object recognition systems on a level playing field. *Cosyne Abstracts*.

[Ploran et al., 2007] Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., and Wheeler, M. E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fmri. *The Journal of Neuroscience*, 27(44):11912–11924.

[Preston et al., 1986] Preston, G., Dickinson, A., and Mackintosh, N. (1986). Contextual conditional discriminations. *The Quarterly Journal of Experimental Psychology*, 38(2):217–237.

[Ramsey, 1931] Ramsey, F. P. (1931). Truth and probability. *The foundations of mathematics and other logical essays*, pages 156–198.

[Rao and Ballard, 1999] Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87.

[Reznikova, 2007] Reznikova, Z. (2007). *Animal intelligence. From individual to social cognition*, volume 37. Cambridge University Press.

[Roberts et al., 1988] Roberts, A., Robbins, T., and Everitt, B. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology*, 40(4):321–341.

[Rokni et al., 2007] Rokni, U., Richardson, A. G., Bizzi, E., and Seung, H. S. (2007). Motor learning with unstable neural representations. *Neuron*, 54(4):653–666.

[Roller et al., 2001] Roller, C. A., Cohen, H. S., Kimball, K. T., and Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor plasticity. *Cognitive brain research*, 12(2):341–352.

[Rosch, 1978] Rosch, E. (1978). Principles of categorization. *Cognition and Categorization*, pages 27–48.

[Rubin et al., 2012] Rubin, J., Shamir, O., and Tishby, N. (2012). Trading value and information in mdps. In *Decision Making with Imperfect Decision Makers*, pages 57–74. Springer.

[Sadtler et al., 2014] Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Byron, M. Y., and Batista, A. P. (2014). Neural constraints on learning. *Nature*, 512(7515):423–426.

[Sanborn et al., 2010] Sanborn, A. N., Griffiths, T. L., and Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, 117(4):1144.

[Sato et al., 2007] Sato, Y., Toyoizumi, T., and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural computation*, 19(12):3335–3355.

[Savage, 1954] Savage, L. J. (1954). *The foundations of statistics*. Wiley, New York.

[Schneiderman et al., 1962] Schneiderman, N., Fuentes, I., and Gormezano, I. (1962). Acquisition and extinction of the classically conditioned eyelid response in the albino rabbit. *Science*, 136(3516):650–652.

[Schrier, 1984] Schrier, A. M. (1984). Learning how to learn: the significance and current status of learning set formation. *Primates*, 25(1):95–102.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

[Seed et al., 2009] Seed, A. M., Call, J., Emery, N. J., and Clayton, N. S. (2009). Chimpanzees solve the trap problem when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(1):23.

[Seidler, 2004] Seidler, R. D. (2004). Multiple motor learning experiences enhance motor adaptability. *Journal of cognitive neuroscience*, 16(1):65–73.

[Seidler, 2007] Seidler, R. D. (2007). Older adults can learn to learn new motor skills. *Behavioural brain research*, 183(1):118–122.

[Shadmehr and Mussa-Ivaldi, 1994] Shadmehr, R. and Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5):3208–3224.

[Shadmehr et al., 2010] Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33:89–108.

[Shea and Morgan, 1979] Shea, J. B. and Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2):179.

[Shultz, 2003] Shultz, T. R. (2003). *Computational developmental psychology*. MIT Press, Cambridge MA.

[Simon, 1955] Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

[Simon, 1972] Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, 1:161–176.

[Sims, 2003] Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

[Sims, 2010] Sims, C. A. (2010). Rational inattention and monetary economics. *Handbook of Monetary Economics*, 3:155–181.

[Steyvers et al., 2003] Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., and Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive science*, 27(3):453–489.

[Still, 2014] Still, S. (2014). Lossy is lazy. In *Proceeedings of the Workshop on Information Theoretic Methods in Science and Engineering*, pages 17–21.

[Still and Crutchfield, 2007] Still, S. and Crutchfield, J. P. (2007). Structure or noise? *arXiv preprint arXiv:0708.0654*.

[Stryker et al., 1978] Stryker, M. P., Sherk, H., Leventhal, A. G., and Hirsch, H. V. (1978). Physiological consequences for the cat's visual cortex of effectively restricting early visual experience with oriented contours. *Journal of Neurophysiology*, 41(4):896–909.

[Tenenbaum and Griffiths, 2001a] Tenenbaum, J. B. and Griffiths, T. L. (2001a). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, 24(04):629–640.

[Tenenbaum and Griffiths, 2001b] Tenenbaum, J. B. and Griffiths, T. L. (2001b). Structure learning in human causal induction. *Advances in neural information processing systems*, pages 59–65.

[Tenenbaum et al., 2006] Tenenbaum, J. B., Griffiths, T. L., and Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends in cognitive sciences*, 10(7):309–318.

[Tenenbaum et al., 2011] Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.

[Tervo et al., 2016] Tervo, D. G. R., Tenenbaum, J. B., and Gershman, S. J. (2016). Toward the neural implementation of structure learning. *Current opinion in neurobiology*, 37:99–105.

[Thelen et al., 2001] Thelen, E., Schöner, G., Scheier, C., and Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and brain sciences*, 24(01):1–34.

[Thompson and Oden, 2000] Thompson, R. K. and Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24(3):363–396.

[Tishby et al., 1999] Tishby, N., Pereira, F. C., and Bialek, W. (1999). The information bottleneck method. *The 37th annual Allerton Conference on Communication, Control, and Computing*.

[Tishby and Polani, 2011] Tishby, N. and Polani, D. (2011). Information theory of decisions and actions. In *Perception-Action Cycle*, chapter 19. Springer.

[Tkačik and Bialek, 2014] Tkačik, G. and Bialek, W. (2014). Information processing in living systems. *arXiv preprint arXiv:1412.8752*.

[Todorov, 2004] Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9):907–915.

[Todorov, 2007] Todorov, E. (2007). Linearly-solvable Markov decision problems. In *Advances in Neural Information Processing Systems*, pages 1369–1376.

[Toni et al., 2009] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202.

[Trobalon et al., 2003] Trobalon, J., Miguelez, D., McLaren, I., and Mackintosh, N. (2003). Intradimensional and extradimensional shifts in spatial learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 29(2):143.

[Tsao, 2014] Tsao, D. (2014). The macaque face patch system: A window into object representation. In *Cold Spring Harbor symposia on quantitative biology*, volume 79, pages 109–114. Cold Spring Harbor Laboratory Press.

[Turnham et al., 2011] Turnham, E. J. A., Braun, D. A., and Wolpert, D. M. (2011). Inferring visuomotor priors for sensorimotor learning. *PLoS Computational Biology*, 7(3):13.

[Van Dijk et al., 2009] Van Dijk, S. G., Polani, D., and Nehaniv, C. L. (2009). Hierarchical behaviours: getting the most bang for your bit. In *Advances in artificial life. Darwin Meets von Neumann*, pages 342–349. Springer.

[Von Neumann and Morgenstern, 1944] Von Neumann, J. and Morgenstern, O. (1944). Theory of games and economic behavior.

[Wasserman et al., 2001] Wasserman, E. A., Fagot, J., and Young, M. E. (2001). Same–different conceptualization by baboons (papio papio): The role of entropy. *Journal of Comparative Psychology*, 115(1):42.

[Weir et al., 2002] Weir, A. A., Chappell, J., and Kacelnik, A. (2002). Shaping of hooks in new caledonian crows. *Science*, 297(5583):981–981.

[Welch et al., 1993] Welch, R. B., Bridgeman, B., Anand, S., and Browman, K. E. (1993). Alternating prism exposure causes dual adaptation and generalization to a novel displacement. *Perception & Psychophysics*, 54(2):195–204.

[Whiten et al., 2005] Whiten, A., Horner, V., and De Waal, F. B. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, 437(7059):737–740.

[Wiener, 1948] Wiener, N. (1948). Cybernetics or control and communication in the animal and the machine. *MIT Press, Cambridge, MA*.

[Wigmore et al., 2002] Wigmore, V., Tong, C., and Flanagan, J. R. (2002). Visuomotor rotations of varying size and direction compete for a single internal model in a motor working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):447.

[Wolpert et al., 2011] Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739–751.

[Wolpert and Flanagan, 2010] Wolpert, D. M. and Flanagan, J. R. (2010). Motor learning. *Current biology*, 20(11):R467–R472.

[Wolpert and Ghahramani, 2000] Wolpert, D. M. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *nature neuroscience*, 3:1212–1217.

[Young and Wasserman, 1997] Young, M. E. and Wasserman, E. A. (1997). Entropy detection by pigeons: Response to mixed visual displays after same–different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 23(2):157.

[Yousif and Diedrichsen, 2012] Yousif, N. and Diedrichsen, J. (2012). Structural learning in feedforward and feedback control. *Journal of neurophysiology*, 108(9):2373–2382.

[Zarahn et al., 2008] Zarahn, E., Weston, G. D., Liang, J., Mazzoni, P., and Krakauer, J. W. (2008). Explaining savings for visuomotor adaptation: linear time-invariant state-space models are not sufficient. *Journal of neurophysiology*, 100(5):2537–2548.