

Universitätsklinikum Hamburg-Eppendorf

Zentrum für Experimentelle Medizin

Institut für Neurophysiologie und Pathophysiologie

Direktor: Prof. Dr. Andreas K. Engel

Flexible sensorimotor linkage during decision-making in changing environments

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Keno Hagena

aus Henstedt-Ulzburg

Hamburg, 2021

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: **31.03.2022**

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: **Prof. Dr. Götz Thomalla**

Prüfungsausschuss, zweite/r Gutachter/in: **Prof. Dr. Tobias Donner**

Contents

1	Introduction	4
1.1	Information integration in decision making	6
1.2	Decision-making in uncertain environments	8
1.3	Hierarchical decision-making	10
1.4	Neuromodulation of decision-making in uncertain environments	12
1.5	Challenges of brainstem fMRI and pupil diameter as a proxy for neuro- modulatory arousal systems	14
2	Results	16
2.1	Task, participants, descriptive behavior	17
2.2	Normative model of belief updating captures participants’s behavior . . .	18
2.3	Cortical correlates of sensorimotor decision	23
2.4	Cortical correlates of the belief updating process	27
2.5	Brainstem nuclei reflect sensorimotor decision in “Inferred rule” trials . .	31
2.6	Pupil responses reflect change-point probability of evidence samples . .	32
3	Discussion	35
3.1	Evidence accumulation in changing environments	36
3.2	Correlates of sensorimotor choice and task rule	38
3.3	Encoding of task rule and coupling of decision processes	40
3.4	Role of neuromodulatory brainstem centers	42
3.5	Outlook	45
4	Methods	46
4.1	Participants	46
4.2	Task	47
4.3	Stimuli	48
4.4	Procedure	48
4.5	Behavioral model	49
4.6	Model fitting	50
4.7	Alternative models	51
4.8	Computational parameters of the belief updating process	52
4.9	fMRI data acquisition	52

4.10 FMRI preprocessing	53
4.11 General linear model	54
4.12 Cortical regression results	54
4.13 Delineation of brainstem ROIs	55
4.14 Pupil data acquisition and preprocessing	55
4.15 Epoch based analyses	55
5 Summary	57
6 Abbreviations	59
7 Supplementary Materials	60
8 Acknowledgements	62
9 Eidesstattliche Versicherung (Declaration of academic integrity)	64

1 Introduction

When making simple perceptual decisions we produce specific responses based on an interpretation of incoming sensory information (??). Imagine walking on a street crowded with people and seeing a remotely familiar person from distance. You scan characteristic traits of the person's face, clothes, and gait in order to discern if you indeed know the person in question. The outcome of this decision then elicits a specific response, e.g. greeting. During this process, the brain needs to pass information from cue specific sensory neurons to action effector specific neurons through several intermediate stages with integrative function, that have been characterized in extensive studies of behaving non-human primates (?????). The flow of information during such simple perceptual decisions can be approximated as follows: Neurons in sensory cortical areas track the *evidence value* of presented stimuli, for example, sensory information regarding the color or shape of the face. Neurons in higher cortical areas - such as in the parietal and frontal cortices - integrate the subsequently incoming pieces of information over time (?). This accumulated evidence can be conceptualized as a *decision variable*. Finally, in motor and motor preparatory cortices, a signature of this *decision variable* can be traced in the form of a gradual build-up of choice predictive activity (??). Once this activity reaches a threshold, the decision-maker commits to a decision in the form of a prompted motor response.

In distinct contexts, decision-makers can be required to come up with different responses given the exact same sensory input. Adaptive cognitive behavior thus requires i) the arbitrary mapping of sensory inputs to motor outputs and ii) the ability to change these mappings flexibly when contextual changes demand it. For example, in the case of discerning a familiar person among the crowd, the elicited response can drastically vary depending on the contextual setting or on the current status of your relationship with that person. In such scenarios, the information flow during decision-formation needs to be routed flexibly from cue-specific sensory neurons to action effector-specific neurons. Critically, in the real world, decisions are corrupted by various sources of uncertainty: "Expected uncertainty" arises from the inherent noisiness of sensory information itself. Moreover, environmental contexts are often unstable and can undergo sudden changes giving rise to a second form of uncertainty coined "unexpected uncertainty" (?). When the context changes unexpectedly, which sensorimotor mapping

is currently appropriate can itself be subject to a decision-making process that results in a covert belief state. Thus, the information flow from sensory neurons to action effector-specific neurons depends on the belief state regarding the current state of the uncertain environmental context. A change in belief state potentially entails a complete reversal of information flow between cue-specific sensory neurons and action effector-specific motor neurons.

An influential theory maintains that the brain accomplishes adaptive flexibility of decision-making through a top-down signal carrying information about the mapping of motor responses onto perceptual interpretations of the sensory signal (????). However, previous studies that entailed different stimulus-response mappings and, thus, required a flexibly routed flow of information from sensory to motor areas explicitly instructed the rewarded stimulus-response mapping (??). In contrast, in more ecological scenarios this contingency is not instructed but highly uncertain and requires the decision-maker to entertain an ongoingly updated belief about it. We hypothesize, that flexible behavioral adaptation under uncertainty about a hidden state of the environment requires i) a belief updating process, which keeps track of the evidence for competing hypotheses regarding the state of the environment (i.e. the context) and ii) a simple decision-making process which interprets and reacts to sensory cues dependent on current belief state regarding the context. These two processes should be coupled functionally in the sense that the latent belief state regarding the context shapes the information flow from cue-specific sensory neurons to action effector-specific motor neurons in the low-level decision-making process.

In the below study, we were interested in investigating how the brain flexibly changes the flow of information of a decision-making process based on a volatile belief regarding the state of the environment. We, therefore, constructed a task, in which the same visual stimuli required different responses depending on context-dependent response rules. The rule-defining context was uncertain and could potentially change at any given time. Specifically, participants had to infer from a noisy stream of cues presented on the timescale of tens of seconds, which of two possible task rules was the currently active one. Throughout the task, a simple visual grating was shown and the currently active task rule had to be applied. Critically, the rewarded rule could switch at any time without being explicitly signaled to the participants. Our task was partially derived from a recent normative Bayesian model of evidence accumulation in changing

environments (?). In this framework, a belief variable about a hidden state of the world is updated with each new arriving piece of evidence by discounting the prior belief and adding the new evidence. The discounting of the prior belief depends on the expected "hazard rate", i.e. the probability of the state of the world to change. In our task, hidden states of the world represented the task rules, that had to be inferred and applied. The following rationale motivated our task design: First, the visual-choice task required a complete reversal of information flow between task-relevant neural populations in visual and motor cortical areas. Second, the volatility and uncertainty of the rewarded task rule made some form of belief signal necessary. Third, using the normative Bayesian model of evidence accumulation by Glaze et al. (2015) enabled us to track this hidden belief variable behaviorally. And finally, the different timescales of the slow belief updating process and the fast sensorimotor decisions allowed a disentanglement of the neural correlates of these intertwined processes.

By combining this novel task with 3 Tesla functional magnetic resonance imaging (fMRI) and pupillometry we tried to probe into two questions. First, we were interested in the brain areas that are involved in the two coupled decision-making processes. Second, based on a large body of evidence, positing that neuromodulatory brainstem nuclei are involved in monitoring the statistical nature of the environment (????) and providing a signal for adaptive behavior accordingly (?), we aimed to investigate the role of neuromodulatory brainstem centers in the interplay between these two levels of decision making. This thesis will be structured as follows. First, I will briefly summarize the progress made over the last decades in the field of decision neuroscience, particularly focusing on decision-making in uncertain environments. Next, I will review three neuromodulatory systems and their respective effect on cognition and decision-making. I will, moreover, highlight challenges in measuring neural activity in the brainstem and briefly review the role of pupillometry as a proxy for activity in brainstem arousal systems. This introductory section is followed by results, methods, and discussion of the conducted study, which constitutes the centerpiece of this thesis.

1.1 Information integration in decision making

Decision-making can be understood as the process of deliberation between multiple alternatives leading to the commitment to a choice. We continuously need to take a plethora of decisions varying in scope and timescale. It comprises complex decisions

("Is the suspect in a judicial trial guilty?") as well as ordinary everyday decisions ("Should I wear a rain jacket today?").

In most decisions, we have to deal with various sources of uncertainty. One substantial source of uncertainty - termed "expected uncertainty" (?) - originates from the inherent unreliability (i.e. noisiness) of the information we gather in our decision-making. For example, when deciding whether to wear a rain jacket today we might base our decision on the weather forecast we checked in the morning, which is inherently probabilistic and thus "noisy". A fruitful strategy for reducing the "noise" of our decisions caused by this kind of uncertainty is to base our decisions on multiple pieces and different sources of evidence (???). For example, before dressing, we might supplement our evidence gathered from the forecast with a glimpse out of the window to check if the sky is blue or cloudy right now.

The combination of multiple pieces of evidence in the formation of a decision requires a mechanism of integration. On a conceptual level, decision-making can thus be dissected into the following components: A single piece of evidence is associated with an *evidence value*. For example, seeing a blue sky is associated with a significant *evidence value* in favor of the hypothesis that it will not rain in the next hour. The entirety of currently available pieces of evidence is aggregated into a quantity that keeps track of the current overall balance of evidence. In the context of decision-making theory, this quantity is referred to as the *decision variable* (?). Finally, a *decision rule* defines how the decision variable is transformed into a discrete decision. Different contexts can require different decision rules. For example, in a court ruling - where the consequences of a wrong decision are grave - the conviction of the suspect requires a certain body of proof. The respective decision rule would state, that a decision is taken only after considering all accessible sources of evidence and when the decision variable has reached a certain magnitude. On the contrary, in the above example regarding the choice of clothes for a walk outside, time can be the constraining factor requiring a decision rule with a set time limit and a higher rate of acceptable errors. These principal constituents of a decision - the *evidence value* of consecutive pieces of evidence sequentially integrated into a *decision variable* resulting in the commitment to a decision based on a *decision rule* - have been formalized in numerous models and applied in a range of fields. One prominent example is the sequential probability ratio test (?), of which an independently developed version helped Alan Turing decrypt the setting of

German Enigma machines in World War II, a breakthrough allowing to decode thousands of encrypted military messages per day (??). Critically, related models - most prominently the drift-diffusion model (?) - have been shown to capture behavioral data in a variety of decision-making studies (??). This insight raises the question of how the integration of evidence is mechanistically implemented in the brain.

Seminal studies in non-human primates performing basic (mostly perceptual) decision-making paradigms have generated important mechanistic insights over the last decades. For example, researches showed that the evidence value associated with an ambiguous visual stimulus in a perceptual decision-making task was encoded in the middle temporal area (MT), a part of the extrastriate visual cortex tuned for visual motion direction (???). In higher association cortices (e.g. in the lateral intraparietal sulcus) a signature of the decision variable was found, namely response-selective activity ramping up towards a threshold, upon which the animal indicated the choice through a saccade (???). Non-invasive recordings of neural mass activity in humans complemented these findings. For example, gamma-band activity in magnetoencephalography (MEG) recordings of the visual cortex reflected visual motion coherence of a random dot motion stimulus (?), while a choice-predictive build-up of activity resembled a decision variable in motor regions (????).

To conclude, during decision-making, which naturally requires the accumulation and integration of pieces of evidence, information needs to be passed from sensory areas across association cortices and ultimately to motor cortical areas responsible for eliciting the response. Across this processing hierarchy, a temporally integrated track record of evidence emerges - the decision variable - ultimately resulting in the commitment to a decision. Studying basic decision-making schemes has yielded insights into the neural nodes of this processing hierarchy. Typically, however, studies used a static stimulus-response mapping (e.g. left visual motion reported with left-directed saccade or button-press) and thus a preset route of information flow. In our study, we were interested in how the brain reconfigures itself to reroute this information flow in cases where the stimulus-response mapping can change dynamically.

1.2 Decision-making in uncertain environments

Growing knowledge about the neural basis of the most basic (often perceptual) decisions has paved the way for investigating mechanisms of decision-making under more

ecological conditions. A distinct feature of most real-world decision-making scenarios regards the volatility of the world surrounding us. That is, the state of natural environments can often change covertly and unpredictably. Consider the above example of dressing for a walk outside. You have seen a predominantly blue sky while casting a glance through the window and your flatmate has asserted that the weather was splendid when he went for a walk in the morning. You accumulate these two pieces of information and choose a light summer jacket when leaving the flat. However, on the street, you sense a series of raindrops on your skin and the sky appears clouded. These two sensations - water on your skin and a grey sky - could potentially be explained by a state-change of the environment (i.e. from "good weather" to "rainfall"). This feature of most natural environments - the possibility of sudden unpredictable changes - distinguishes decision-making in many ecological settings from most decision-making studies, in which the underlying signal eliciting the noisy stimulus typically stays static throughout a decision trial (?). Incorporating the possibility of sudden state-changes of the underlying ground truth into experiments introduces a second form of uncertainty termed "unexpected uncertainty", i.e. uncertainty, which is not due to the noisy nature of the stimulus itself, but results from a potential change in the underlying source which generated the sensory signal (?). Whereas variance in a new observation caused by the stochastic nature of the stimulus itself (i.e. "expected uncertainty") is in principle not distinguishable from variance in a new observation caused by a state change (i.e. "unexpected uncertainty"), the potential implications for evidence accumulation are severe: A change of the underlying state of the environment renders (most) past information practically irrelevant, while high amounts of stochastic noise require smoothing of our estimate by averaging over many samples.

An intuitive approach to reducing "unexpected uncertainty" (i.e. uncertainty due to a potential change-point) is to discount past information and attach more weight to more recent evidence. By balancing the impact of prior information and new incoming information, the decision-maker governs the trade-off between the computational goals of identifying an underlying signal precisely and detecting possible change-points reliably. When the environment can be assumed to be perfectly stable - as in classical decision-making tasks - perfect accumulation (i.e. linear integration) of evidence samples is the optimal accumulation strategy (?). However, when past experiences suggest that the environment tends to change, one should adapt the influence of prior informa-

tion and new evidence on the decision variable based on an estimate of the volatility of the environment (?). Such adaptive balancing of prior information and new evidence has been shown to drive behavior in a range of decision-making tasks that require evidence accumulation in a changing environment (????). These recent normative models of evidence accumulation in changing environments incorporate an update-rule by which new incoming evidence and prior information are integrated non-linearly into the newly updated belief regarding the hidden state of the environment based on the probability that a change-point in the hidden state has occurred (?). For example, following a supposed change-point new evidence should be up-weighted while prior information should be discounted more strongly. This dynamic updating can be formulated in a more intuitive way as modulation effects of statistical measures of surprise (unexpectedness of a new observation) and uncertainty regarding the previous belief on the weighting of new samples (?). The influence of a surprising sample should be up-weighted, as it is indicative of a change-point. Similarly, while increased uncertainty regarding the previous belief prevails, new incoming evidence should be also conceded enhanced impact. A line of research combining imaging and computational modeling is beginning to unravel the neural underpinnings of decision-making in changing environments. Neural correlates encoding these statistical measures have been found especially in the anterior cingulate cortex (ACC) and the parietal cortex (???). Using MEG, Murphy et al. showed elegantly how the neural encoding of prior information, new evidence value, and the decision variable is modulated by surprise and uncertainty (?).

1.3 Hierarchical decision-making

When we infer the current state of the environment surrounding us, we are usually not asked to directly report the outcome of this inference process - as it is the case in most decision-making experiments. Instead, the outcome of this inference process guides our decisions in that environment. What results is another hallmark of goal-directed behavior in natural environments, that is, hierarchically organized decision-making processes (?). Specifically, higher-level decisions or inference processes over long time scales determine abstract strategies or beliefs regarding our environment. Low-level decisions, which depend on these higher-level strategies or beliefs, result in concrete actions and guide our responses to incoming stimuli. The requirement of our complex

world to compute multiple intertwined decisions at varying levels of abstraction and on distinct time scales is reflected in hierarchical principles of organization observed in the neocortex in several recent studies. For example, Murray et al. showed that the temporal structure in electrophysiological data exhibits a hierarchical gradient from early sensory cortices to higher association cortices (?). Furthermore, the representation of rules in lateral frontal cortex has been observed to follow a rostro-caudal gradient regarding the abstractness of rules (??). Specifically, more abstract rules, i.e. rules more distant from concrete action such as rules about the choice of rules, were represented in more anterior parts of the lateral frontal cortex. Finally, patterns of fMRI signal were found to reflect a hierarchy of reward structures in the medial prefrontal cortex, with activity in more anterior and posterior parts reflecting changes in reward structure across task blocks and trials, respectively (?). Recent studies have probed into the dynamics of decision-making in hierarchical environments (????). Several of these studies used task paradigms that entailed a volatile higher-level decision rule determining how to respond in a low-level perceptual decision-making task (??). Specifically, participants were asked to respond to noisy stimuli applying one of two uncertain and volatile response rules. After each sample, participants were given feedback on the correctness of the response. In these task paradigms feedback was ambiguous as errors could result from both a hidden rule switch or from an erroneous stimulus identification. Behavioral analyses revealed that confidence, i.e. the decision-maker's belief in the correctness of her choice, provides a critical signal for distinguishing between these potential sources of errors (?). On the level of neural implementation, the anterior cingulate cortex was found to accumulate evidence for a switch in the underlying environment as the source of errors (?).

As another example of a hierarchical decision-making scenario, our experiment relates to this line of research. However, our task design differs from the above-reviewed studies in several points. First, we do not provide direct feedback regarding the correctness of responses. Second, we do not interrogate the decision-maker's confidence in decisions. Finally and most importantly, in our task, the two (hierarchically intertwined) decision processes operate on two distinct sources of sensory input presented on varying time scales: A regular stream of sampled noisy cues presented over the course of tens of seconds informs the top-level inference process whereas the low-level decision operates on unambiguous grating stimuli interlaced with variable intervals. This

approach enables us to disentangle the two decision processes and facilitates relating the two processes to neural data independently.

1.4 Neuromodulation of decision-making in uncertain environments

Neuromodulators describe specific systems of neurotransmitters that have been implicated influentially over the last decades in modulating cognitive function, behavior, and global brain state in highly specific manners. Anatomically, these neuromodulatory systems comprise small nuclei in the brainstem and the basal forebrain with widespread ascending projections throughout most of the cortex. At the level of targeted synapses, the involved neurotransmitters typically do not directly participate in the signal transmission in the form of eliciting a postsynaptic excitatory or inhibitory potential but rather alter the response properties of the targeted neurons, thereby modulating signal transmission. Neuromodulatory systems have been linked to the regulation of the global brain state, arousal, or attention. More recently, however, studies have also demonstrated more specific and temporally highly precise effects on cognitive operations, including decision-making. The neuromodulatory systems reviewed in the following include the noradrenergic locus coeruleus (LC-NE), the dopaminergic ventral tegmental area (VTA-DA), and the acetylcholinergic basal forebrain (BF-ACh). The serotonergic dorsal raphe nuclei as well as histamine, which have been furthermore identified as neuromodulators, will be omitted in the scope of this thesis as little is known about them in the context of inference and decision-making.

The locus coeruleus (LC) comprises a few thousand neurons in the rostral pons, constitutes the main source of norepinephrine (NE) in the central nervous system, and entertains widespread projections throughout almost the entire brain (?). LC-NE function has been influentially conceptualized in terms of two operation modes that impact decision-making distinctly, namely a fast (phasic) and a slow (tonic) release mode (?). Phasic bursts of LC activity and consecutive NE release have been demonstrated to be closely time-locked to salient stimuli in elementary decision-making tasks (??). This phasic activity has been proposed to facilitate the exploitation of profitable behavior (?) and has been shown to reduce the bias of decision-makers towards one of two alternative choices, thereby significantly enhancing task performance (?). LC-NE activity in the tonic mode, i.e. fluctuations of NE release on a time scale of seconds to minutes, is hypothesized to engender exploration of alternative (task) strategies potentially through

an increase in gain at targeted synapses in a temporally non-specific manner (??). Intriguingly, the activity of the LC-NE neuromodulatory system has been formulated as signaling "unexpected uncertainty", i.e. uncertainty that does not arise from the intrinsic stochasticity of a sensory signal ("expected uncertainty") but from a potential change in the underlying source which generated the sensory signal (?). Increased NE-release due to increased "unexpected uncertainty" could serve the decision-maker to disengage from a previously profitable strategy and reorient attention to other features or the environment or new incoming evidence (???)

The acetylcholine (ACh)-releasing neurons that reside in multiple nuclei in the basal forebrain and entertain widespread projections to the entire cortex (??) constitute another neuromodulatory system relevant for inference problems in settings of uncertainty. On a conceptual level, ACh has been hypothesized to signal "expected uncertainty", that is uncertainty arising due to the stochastic nature of a noisy stimulus or cue (?). Thus, NE and ACh can be regarded as counterparts presumably mediating two distinct signals of uncertainty - "unexpected" and "expected uncertainty", respectively - associated with a particular cue or piece of information. From a Bayesian standpoint of optimal inference, both sources of uncertainty should impact the balancing of prior belief and new information in the form of upweighting new pieces of evidence (??). This notion finds empirical support in the finding that ACh projections to the primary visual cortex increase the signal-to-noise ratio by enhancing bottom-up thalamo-cortical sensory input while simultaneously suppressing top-down cortico-cortical connections (???)

As a third neuromodulatory system that is potentially relevant in the context of our study, dopaminergic neurons in the ventral tegmental area (VTA) entertain strong projections to the prefrontal cortex and have been studied extensively in the context of reward processing. Robust insights stem from classical conditioning experiments, in which monkeys were presented visual cues, which were deterministically followed by a reward. Before the animal learned the cue-reward contingency dopamine (DA) mid-brain neurons fired in response to the (unexpected) reward. After learning, increased firing rates of DA neurons were observed after the cue, not the reward itself, and firing rates decreased when an expected reward was absent (??). These striking findings motivated the established framework of DA neurotransmission reflecting so-called reward-prediction errors (i.e. a computational measure of the discrepancy between ex-

pected and actual reward) used in reinforcement learning models to update internal estimates of values of certain states or actions and thereby driving learning and future behavior (?). Critically, influential theories of prefrontal cortex function (especially regarding adaptive behavior) have assigned a key role to DA signaling (?). The framework posits that the prefrontal cortex (PFC) is involved in guiding adaptive behavior by instantiating and maintaining higher-order sets of goals or rules which modulate the processing of (sensory) input (?). An apt illustration (and experimental paradigm) is the well-established Stroop-task in which participants are asked to denote the color of a word, which states the name of a color, printed in another (conflicting) color (e.g. the word "blue" printed in red ink) - either by attending to the written word or the ink color (??). The same sensory input in this paradigm elicits converse responses based on a predefined task set ("word" or "ink color") - supposedly maintained in PFC. When environments are volatile and tend to change, they warrant updates of goals or task strategies to engender adaptive behavior. In such scenarios, the rule center in the form of the PFC needs to balance the trade-off between i) maintaining a strategy robustly in the face of irrelevant stimuli (distractors) and ii) remaining responsive to salient signals of the environment. It has been hypothesized that the ample DA projections to PFC - signaling discrepancies between anticipated reward and actual reward - constitute a temporal gating mechanism to afford updates of strategies or goals when the environment demands it and stabilize a specific task set in the meantime (????).

The above highlighted theoretical frameworks, which implicate the LC-NE, the BF-ACh, and the VTA-DA neuromodulatory systems in decision-making and inference under uncertainty and in volatile environments, motivated us to particularly probe into the activity of these systems in our study.

1.5 Challenges of brainstem fMRI and pupil diameter as a proxy for neuromodulatory arousal systems

Recording the neural activity in neuromodulatory brainstem centers non-invasively poses several challenges. First, the relatively small size of neuromodulatory nuclei (e.g. the LC; Keren et al., 2009) and the anatomical position in the brainstem, which is prone to physiological nuisance due to adjacent blood vessels and pulsatile cerebrospinal fluid (?), pose substantial challenges for fMRI imaging of these structures (??). Second, inferring the activity of particular neuromodulatory systems (such as the LC-NE) from the

fMRI signal in the identified region (e.g. the LC) is hampered by the fact that most of these nuclei also contain various other functional types of neurons such as populations of inhibiting interneurons (?). Third, the temporal resolution of fMRI data is inherently low and prevents the analysis of rapid fluctuations of neural activity such as phasic trial-related responses in neuromodulatory brainstem nuclei. The first problem can be addressed by adhering to recently advanced imaging techniques comprising brainstem-specific acquisition parameterizations and nuisance regression protocols (?). A potential approach of surpassing the latter limitations relies on supplementing the imaging data with pupillometric data, i.e. the recordings of pupil diameter at a high temporal resolution, which serves as a reliable proxy for the neural activity in neuromodulatory arousal systems (??). The diameter of the pupil is controlled by two intertwined pathways comprising various brainstem nuclei and circuits that regulate the tone of two muscles - the iris dilator muscle and the iris sphincter muscle - through sympathetic and parasympathetic nerve fibers, respectively. The primary function of the pupil can be described analogously to a camera shutter as governing the amount of light that enters the eye and ensuring the optimal illumination of the retina. In this sense, the amount of incoming light constitutes the major influence on pupil diameter, effecting a pupil constriction as a response to increased luminance through the pupillary light reflex (?). However, it is long known that pupil size is not solely determined by the amount of incoming light and that the pupil diameter tends to fluctuate significantly even under constant luminance. Early studies of pupillometry (i.e. the quantification of pupil diameter and reactivity) argued that these non-luminance mediated pupil responses reflect neural activity associated with behavioral correlates such as arousal, attention, or cognitive effort (??). A recent line of research has related non-luminance mediated fluctuations of pupil diameter to transitions in cortical brain state (??) Intriguingly, the noradrenergic LC can be regarded as a prime candidate effector of these non-luminance mediated pupil dilations, as its neurons project to both the sympathetic pathway activating the iris dilator muscle as well as to the Edinger-Westphal nucleus to inhibit the constriction pathway (?). Consistent with this are various decision-making studies that related changes in pupil diameter to behavioral readouts compatible with current theories of LC-NE neuromodulation (????????). Corroborating the hypothesis that non-luminance mediated pupil responses reflect in large parts the activity in the LC-NE neuromodulatory system, recent direct evidence stems from i) imaging studies

that relate fMRI signal strength in the LC to baseline pupil diameter (?), or task-evoked changes in pupil diameter (?), ii) electrophysiological studies reporting pupil dilations directly elicited by microstimulation (??) or optogenetic activation (?) of noradrenergic LC neurons and iii) measurements of activity at cortical NE axons originating in the LC (?). The latter study related activity in cortical noradrenergic axons in particular to fast phasic changes in pupil diameter, which were best reflected in a strong cross-correlation between LC-NE activity and the first derivative of pupil diameter (?). However, non-luminance-mediated fluctuations of pupil diameter have also been related to activity in subcortical structures other than the LC. These comprise in particular the BF-ACh (??) and the VTA-DA neuromodulatory systems (?). Critically, recent analysis approaches (??) and refined knowledge of how cognitive input signals to the pupillary apparatus (temporally) relate to changes in pupil diameter (?) facilitate the inference of cognitive processes from pupillometric data. In sum, fluctuations of pupil diameter constitute a reliable proxy for activity in a network of subcortical arousal systems. Within this network, evidence for a causal connection of pupil diameter and LC-NE activity is particularly strong. Importantly, the high temporal resolution of pupillometric data enables relating this proxy of neural activity to trial-wise behavioral readouts, which would otherwise exceed the sparse temporal resolution of fMRI.

2 Results

In the present study, we were interested in the mechanisms of how the brain configures itself and reroutes information flow during decision-making dynamically when required by an uncertain and changing environment. We, therefore, tailored a task in which a simple sensorimotor decision was coupled to a higher-level belief updating process regarding the current state of the environment. We used computational modeling to gain insights into the slow belief updating process and we then deployed functional MRI and pupillometry to pinpoint neural signatures of the belief updating process and the coupled sensorimotor decisions. The results section will be structured as follows: First, we will describe the approach and the results of the computational modeling of the belief updating process. We will then relate these computational insights to neural data using two approaches, i) general linear modeling and ii) epoch-based analyses of the low-level sensorimotor decisions. Finally, we will relate one candidate mechanism,

namely neuromodulation through brainstem nuclei, to the coupling of the two decision processes, again utilizing general linear modeling as well as model-free analyses.

2.1 Task, participants, descriptive behavior

22 Subjects (median age 27, range 21–44, eight male) performed our task in three sessions, of which the first session was a purely behavioral training session and was not considered in subsequent behavioral modeling results. One participant dropped out after the training session due to incompatibility of his glasses with the eye tracker in the MRI scanner and a second participant was excluded from further analyses as 31.8% of given responses were invalid button presses.

In the main condition of the task (“Inferred rule”, see Figure 1a), participants were asked to respond to visual cues with one of two alternative sensorimotor rules. The two rules mapped two potential responses (left or right button press) onto two potential visual cues (a horizontally or vertically oriented grating). Which rule (i.e. cue-response mapping) was correct changed covertly but noisy information about the correct mapping was presented in the form of horizontally displaced dots. Specifically, the dots appeared with varying horizontal distance from a fixation cross and the position was sampled from one of two possible, overlapping normal distributions. Observation of multiple consecutive dots allowed the observer to infer the currently active distribution, i.e. the distribution dot positions were drawn from. With irregular intervals, a visual stimulus (choice cue) – with vertical or horizontal orientation – was presented and the participant was required to respond with a left or right button-press. The rewarded answer depended on the distribution which had generated the last-seen sample, that is, each distribution was coupled to a cue-response mapping (i.e. rule). For example, when the dots were sampled from the left distribution, participants had to respond with a left button press for vertical gratings and a right button press for horizontal gratings (“Rule 1”), and vice versa, when the dot position was drawn from the right distribution (“Rule 2”). Critically, the active distribution – and thereby the rewarded cue-response mapping – changed with a probability of 1/70 (hazard rate) between consecutive dots. In a reduced version of the task (“Instructed rule”), participants were instructed about the correct cue-response mapping. This version of the task was used to i) train participants on these mappings, which stayed constant across all blocks and sessions, and ii) as a control condition, in which stimuli and responses were matched with the full task

variant, but uncertainty concerning the true rule was eliminated.

In the main task condition (“Inferred rule”), subjects had on average 79.7% correct trials ($\pm 5.7\%$ SD; Figure 1c) and a mean reaction time of 863 ms ± 91.3 ms SD (Figure 1b). Two-sample t-tests showed no significant differences in reaction time and performance between two scanning sessions in the “Inferred rule” condition. Across all sessions and runs, we observed a strong dependence of reaction times on stimulus identity regardless of the applied rule (Supplementary Figure 1). Specifically, subjects responded significantly faster to vertical compared to horizontal choice cues. Importantly, reaction times did not covary significantly with the applied rule (Supplementary Figure 1). In the control condition (“Instructed rule”), where participants were informed about the correct stimulus-response mapping rule, participants reached an average of 98.3% of rewarded trials ($\pm 0.2\%$ SD; Figure 1c) and a mean reaction time of 781 ms ± 17.1 ms SD (Figure 1b). Two-sample t-tests showed that both reward rate and reaction time differed significantly across task conditions, that is participants performed better and faster in the control condition, which was expected as the “Instructed rule” condition did not involve uncertainty regarding the correct rule. In sum, participants performed well in all sessions and results were not biased by a significant learning effect between sessions as deduced from similar performance and reaction time across scanning sessions.

2.2 Normative model of belief updating captures participants’s behavior

Next, we were interested in how well a Bayesian normative model of evidence accumulation explained our participants’ inference about the correct rule. We build on recent normative accounts of decision-making in changing environments, which postulate that human decision-makers solve suchlike problems likely by applying some form of non-linear belief updating rule (??). In principle, these models entertain a belief variable about a hidden state of the environment and an update rule, that governs, how new evidence is integrated into this belief variable. As we intentionally refrained from instructing participants about how to optimally solve the task, we were at first interested to see if the participants’ behavior could indeed be captured by such a belief updating process at all. Therefore, we fitted a specific normative model for evidence accumulation in two-alternative forced choice tasks with a changing hidden state (?) to each participant’s behavioral data, i.e. the participant’s responses and the observed samples (dot positions).

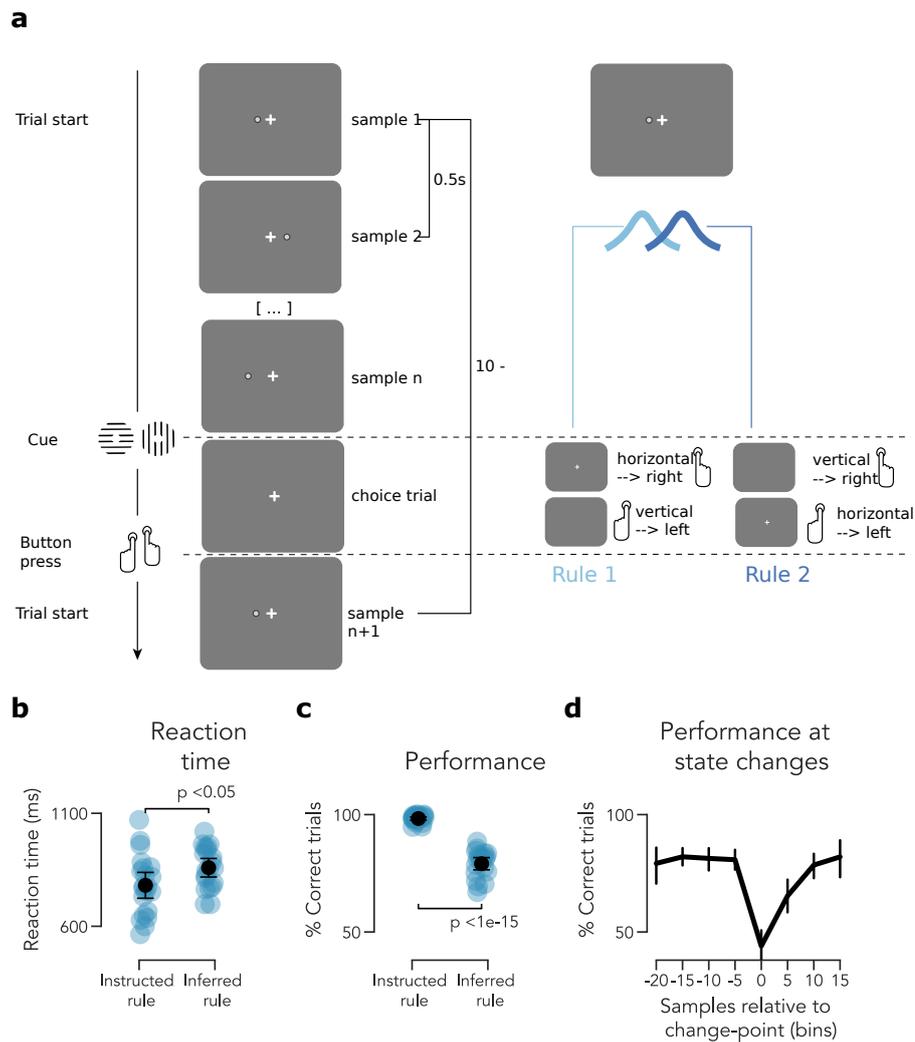


Figure 1: Behavioral task. **a** *Top left:* Participants are presented a continuous stream of small points (“samples”) on the imagined central horizontal axis. *Top right:* The position of these points is sampled from one of two overlapping normal distributions; Critically, the active generating distribution can change at any time with the fixed probability of 1/70. *Bottom left:* From time to time a horizontal cue grating is presented upon which participants have to respond with left or right button press. *Bottom right:* The rewarded cue-response mapping (“Rule 1” or “Rule 2” corresponds with the currently active generating distribution, that is when points were lastly sampled from the left distribution, participants had to respond with left button press to a vertical grating and with right button press to a horizontal grating. **b** and **c** show the percentage of rewarded trials (performance) and reaction times in the “Inferred rule” (full task) and the “Instructed rule” (control) condition. Group average (N = 20); single data points, individual subjects; error bars, 2x s.e.m; statistics, two-sample t-test. **d** Shows percentage of rewarded trials in the “Inferred rule” condition pooled across bins of trials with similar distance to the closest change-point. X-axis denotes lower limit of bins, each bin contains 5 sample positions. Group average across subjects; error bars, 2x s.e.m;

The applied Bayesian model normatively describes a belief updating process in uncertain environments, which are characterized by i) two competing hypotheses about an underlying ground truth and ii) instability regarding the correctness of one of those (i.e. the environment can switch from one state to the other). The normative model by Glaze et al. (2015) entertains an ongoing belief about the current state of the environment, which is updated after each new incoming piece of evidence using a belief update rule. Specifically, at each time step, the belief in one of the two possible states of the environment L_n is updated by discounting the prior belief non-linearly and then adding the evidence value for one of the two possible states (i.e. the log-likelihood ratio) associated with the new incoming sample. In our task setting, the two possible states of the environment comprised the two overlapping normal distributions, from which dot samples were potentially drawn and which were associated with distinct stimulus-response mappings. Importantly, the non-linear discounting of the prior belief in the normative model depends on the subjective estimate of the volatility of the environment, i.e. the probability that a fundamental change in the environment occurs (hazard rate H). When the environment is assumed to be perfectly stable ($H = 0$), the model results in perfect accumulation of samples. When the environment changes with a probability of 50% after each trial ($H = 0.5$), the prior belief is completely discounted such that the new belief equals the current incoming piece of evidence. Generally speaking, the higher one believes the risk of a change point to occur to be - i.e. the higher the subjective estimate of H -, the more is the previous belief discounted, granting higher impact to the current sample (see Figure 2c). In Bayesian terms, the hazard rate governs the natural trade-off between sensitivity to switches and strength of the current belief by balancing prior and likelihood ratio. Figure 2a exemplifies this trade-off by contrasting the course of the belief estimates of two hypothetical observers with different hazard rates given the same evidence: The "alert" observer (red, $H = 0.1$) reacts faster to a change in the generating distribution, however, this happens at the expense of various "false-alarms" (i.e. sign changes of the belief estimate despite the absence of an actual change point). The "conservative" observer (blue, $H = 0.001$) develops robust beliefs during longer streaks of stable environments but reacts sluggishly to changes.

We fitted this model to the participants' responses and the observed cues separately per subject and session using a Bayesian inference approach with two free pa-

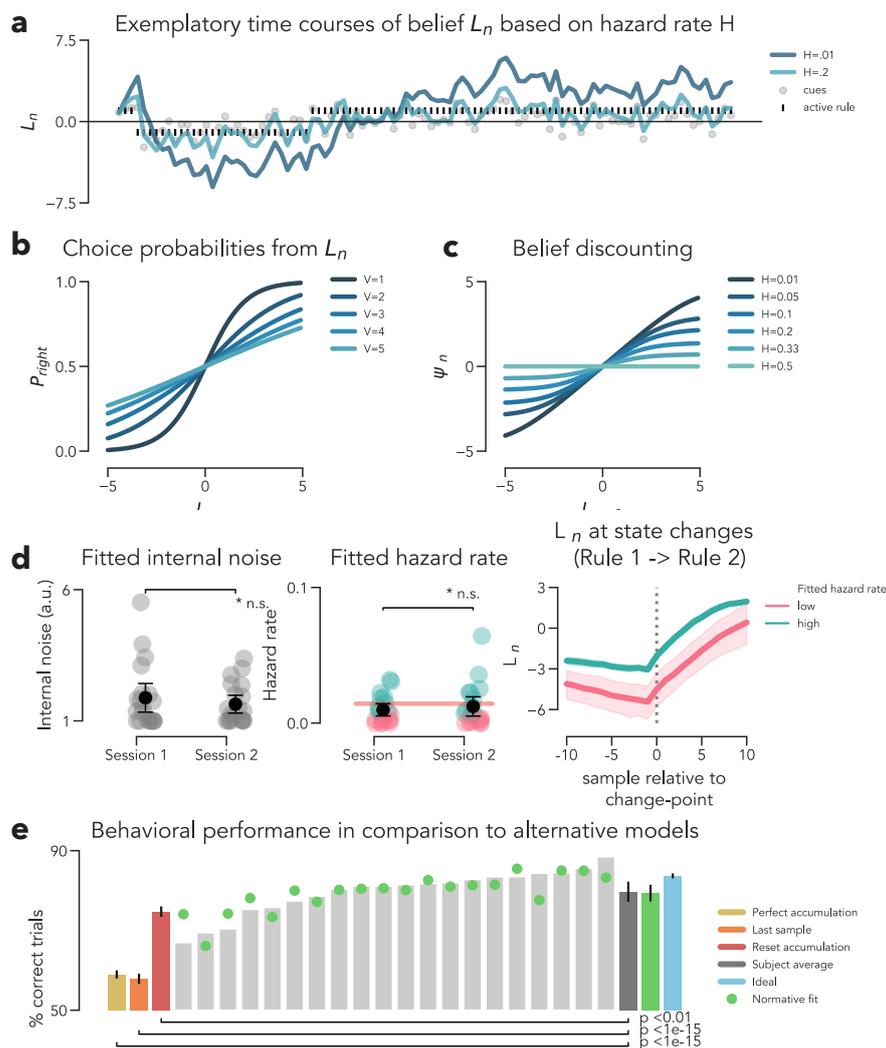


Figure 2: Normative model (?) reflects participants behavior. **a** shows 200 simulated cues and the updated belief variable of two hypothetical observers using different estimates of the hazard rate. **b** depicts how the translation of the posterior belief on the ordinate into the discounted prior belief ψ depends on the hazard rate H . **c** shows how the belief variable translates into choice probabilities, scaled by internal noise V . **d** *left* Modal values of Bayesian posterior distributions of internal noise parameter fitted separately per subject and session. *center* Modal values of Bayesian posterior distributions of hazard rates fitted separately per subject and session. Red horizontal marks true generative hazard rate. Colors distinguish high from low 50% of hazard rate estimates used for conditioning in the *right* panel. *left* and *center*: Group average ($N = 20$); single data points for individual subjects; error bars depict $2x$ s.e.m. *right* shows fitted estimates of L_n around state changes conditioned on fitted hazard rate. **e** Performance (% correct responses) of participants, normative model and alternative heuristics. Light grey, actual participants; dark grey, participants average; green (dots and bar), participants fits of the normative model; blue, ideal normative model equipped with true generative hazard rate and no internal noise; yellow, linear integration of all evidence samples; orange, decision based only on last observed evidence sample; red, linear accumulation of all evidence samples since last decision. Error bars, $2x$ s.e.m; statistics, two-sample t-test.

rameters, the hazard rate H and a parameter representing internal noise. For all participants and all sessions, the posterior distributions of the fitted hazard rates peaked between 0 and 0.1 (Figure 2d). Moreover, the majority of these fitted subjective hazard rates were closely scattered around the true hazard rate used in the generative process ($H = 1/70$), suggesting, that participants made accurate assumptions about the statistical process underlying the presented samples. Two-sample t-tests confirmed that the modal values of fitted posterior distributions of the hazard rate across subjects did not vary significantly between the two scanning sessions ("Session 1" and "Session 2", Figure 2d). Similarly, Bayesian posterior distributions of the second free parameter of the model, internal noise, peaked at reasonable estimates mostly between 1 and 2 and showed no significant difference between the two scanning sessions in a two-sample t-test (Figure 2d).

To investigate whether the model by Glaze et al. (2015) indeed captured the participants' behavior adequately, we compared the participants' performance to three alternative models and heuristics, i) a "last-sample heuristic" (i.e. taking into account only the last observed dot position), ii) perfect accumulation of all observed samples, and iii) a model, which accumulated all samples since the last choice ("reset accumulation"). As depicted in Figure 2e, participants performed significantly better than expected from last-sample heuristic (mean $57.9\% \pm 2.9\%$ SD; $T=15.38$, $p<10^{-15}$), from perfect accumulation (mean $59.0\% \pm 2.3\%$ SD; $T=15.22$, $p<10^{-15}$) as well as from "reset accumulation" (mean $74.7\% \pm 2.9\%$ SD; $T=3.51$, $p<0.01$). While participants performed significantly worse than the ideal observer, i.e. the normative model equipped with the true generative hazard rate and no internal noise (mean $83.7\% \pm 1.4\%$ SD; $T=3.06$, $p<0.01$), there was no significant difference in performance between participants' and the normative model equipped with the individual fits of the participants' hazard rates and internal noise estimates (mean $79.4\% \pm 4.7\%$ SD; $T=0.22$, $p=0.83$). These latter differences might be explained by lacking knowledge about the true hazard rate as well as internal noise and lapses of attention. In sum, the majority of subjects performed the task adequately, reaching a mean of 79.7% rewarded trials. Moreover, comparisons to alternative heuristics suggest that subjects base their choices on several samples of evidence thereby implying some form of belief updating process. Finally, the employed Bayesian model of belief updating - which is able to capture such belief updating processes (?) - could be fitted successfully to the behavioral data. This allowed us to access

computational variables such as the participants' ongoing latent belief about the active distribution or the evidence strength associated with new incoming samples throughout the task and relate them to neural signals in further analyses.

2.3 Cortical correlates of sensorimotor decision

The previous analyses indicated that the majority of participants solved our task through dynamic belief updating and we were able to capture this process with an established computational model (?). We were next interested in neural correlates of the two intertwined decision processes, namely the high-level belief updating captured by the Bayesian model of belief updating and the low-level sensorimotor decision. We, therefore, recorded functional MRI and pupillometry of participants performing our task.

To relate behavioral data to neural activity, we formulated a general linear model (GLM) capturing the relevant components of both decision processes as follows: The low-level decision process comprised in essence three elements, i) the presented stimulus (i.e. trial cue), ii) the given response and iii) the applied rule (i.e. the result of the high-level belief updating process). Each of these three components afforded two possible conditions: The trial cue could be either horizontal or vertical, participants could respond with a left or right button press, and one of two possible mappings of these responses onto stimuli would be rewarded as the correct rule. However, as one of these three components would always carry mutual information with the other two (e.g. if subjects responded with a left button press to a vertical stimulus, the applied rule would per definition be "Rule 1" and vice versa), four possible combinations of choice cue, response and applied rule resulted. As trial cue and response were moreover hardly discernible given the temporal resolution of functional MRI ($TR = 1.9s$), we accounted for the low-level decision in our GLM with a boxcar regressor lasting from onset (i.e. showing of the trial cue) to the response (i.e. button press) dummy-coded for the four possible cue-response combinations. The higher-level belief updating process, on the other hand, could be captured by the following components. First, the continuously updated belief regarding the currently active distribution (corresponding to the rewarded rule) quantified as L_n in the Bayesian model of belief updating; Second, the evidence carried by each new incoming sample quantified as the log-likelihood ratio (LLR) of each new sample. The latent belief variable L , as well as the evidence value LLR, were included both as a signed variable and as the magnitude of these quantities (i.e. belief strength

and evidence strength, respectively). Finally, the recent insight that evidence accumulation in dynamic environments is critically modulated by the change-point probability (CPP) of new evidence - i.e. the probability that the new sample constitutes a change point (?) - motivated us to also include this measure in our general linear model.

We fitted these regressors in a single general linear model to the preprocessed fMRI signal in each voxel as follows:

$$\begin{aligned}
 \mathbf{fMRI\ signal}_{\text{vox}} = & \beta_{0,\text{vox}} + \beta_{1,\text{vox}} \mathbf{L}_{\text{abs}} + \beta_{2,\text{vox}} \mathbf{L}_{\text{signed}} + \beta_{3,\text{vox}} \mathbf{LLR}_{\text{abs}} \\
 & + \beta_{4,\text{vox}} \mathbf{LLR}_{\text{signed}} + \beta_{5,\text{vox}} \mathbf{CPP} + \beta_{6,\text{vox}} \mathbf{trial}_{\text{horiz, left}} \\
 & + \beta_{7,\text{vox}} \mathbf{trial}_{\text{horiz, right}} + \beta_{8,\text{vox}} \mathbf{trial}_{\text{vert, left}} \\
 & + \beta_{9,\text{vox}} \mathbf{trial}_{\text{vert, right}} + \beta_{10,\text{vox}} \mathbf{trial}_{\text{missed}} \tag{1}
 \end{aligned}$$

Where L denotes the latent belief regarding the active distribution, LLR the evidence value of each sample, and CPP the change-point probability and, where boldface regressors and the response variable denote column vectors, in which each element correspond to a repetition time of the fMRI acquisition. To pool the resulting beta coefficients across participants, we averaged them across all vertices from a set of anatomically and functionally defined regions of interest (ROIS) spanning the entire cortical surface (?). We chose this procedure because we treated these ROIS as functional units. This approach enabled us to first delineate cortical regions involved in the two intertwined decision processes. In the remainder, cursive abbreviations in brackets refer to cortical labels according to this parcellation.

We first assessed the beta-weights of our regression model across regions of the cortical surface according to an anatomical and functional parcellation of the human cerebral cortex (?). In the "Inferred rule" task condition the low-level sensorimotor decision - in our general linear model formulated as a boxcar regressor with four possible states compliant with possible cue-response combinations ("trial") - was significantly correlated with a widespread pattern of cortical activity which seemingly reflected the different elements of the sensorimotor decision, namely cue, response and rule application (Fig. 3a). First, among the cortical areas positively correlated with the sensorimotor choice on average - i.e. irrespective of the choice grating orientation and the direction of the button-press - were all stages of the visual cortex. This probably reflected the processing of the choice cue, a screen-spanning high-contrast grating.

Among visual cortices, t-statistics were highest in V1-V4 followed by specific areas of ventral and dorsal stream visual cortex and MT and neighboring areas.

Second, the low-level decision was on average (i.e. irrespective of response direction and cue orientation) correlated with a network of higher cortical areas including premotor, prefrontal, and cingulate cortex as well as areas in the parietal and temporal cortex. Negatively correlated brain areas included parts of the anterior cingulate and medial prefrontal cortex, the posterior cingulate cortex, and the temporal cortex. Third, analysis of contrasts of correlation beta weights between right and left choice revealed for both choice directions a pronounced lateralized pattern of correlated activity in the contralateral somatosensory and motor cortex as well as in adjacent cortical areas (Fig. 3b). Specifically, activity lateralized to the left somatosensory cortex during right button-press and to the right somatosensory cortex during left button-press. This lateralized activity during the participants' response presumably reflected preparatory and executive activity of the motor-response and served as a sanity check for our regression model.

As stated above, the low-level decision comprises the cognitive operations of i) identifying the stimulus, ii) retrieval of the believed rule mapping, and iii) application of this rule, i.e. the response. We were interested to see whether the pattern of activation during the low-level sensorimotor decision reflected in the regression results was deviant in our control condition, in which the correct rule did not have to be inferred but was disclosed to participants. We, therefore, fitted a reduced version of our GLM to the control data and computed contrast maps of the choice regressor beta-weights between the "Inferred rule" condition and the "Instructed rule" control condition. We found several clusters of brain areas on the cortical surface, which were correlated significantly stronger with the low-level sensorimotor decision in the "Inferred rule" condition compared to the "Instructed rule" condition (Fig. 3c). Among these were areas in the frontal and opercular cortex (*AVI*, *FOP5*), in the dorsal cingulate and superior parietal cortex (*POS2*, *PCV*, *7Pm*, *RSC*), the inferior parietal cortex (*IP1*, *IP2*), and in the anterior cingulate cortex (*8BM*, *a32pr*). We controlled for the possibility that the difference in regression beta-weights in these brain regions was caused by the necessarily different formulation of the general linear model (i.e. computational parameters such as belief or log-likelihood of samples were not available in the control condition) by supplementing a model-free analysis based on fMRI signals during choice epochs (-2 to 12s from

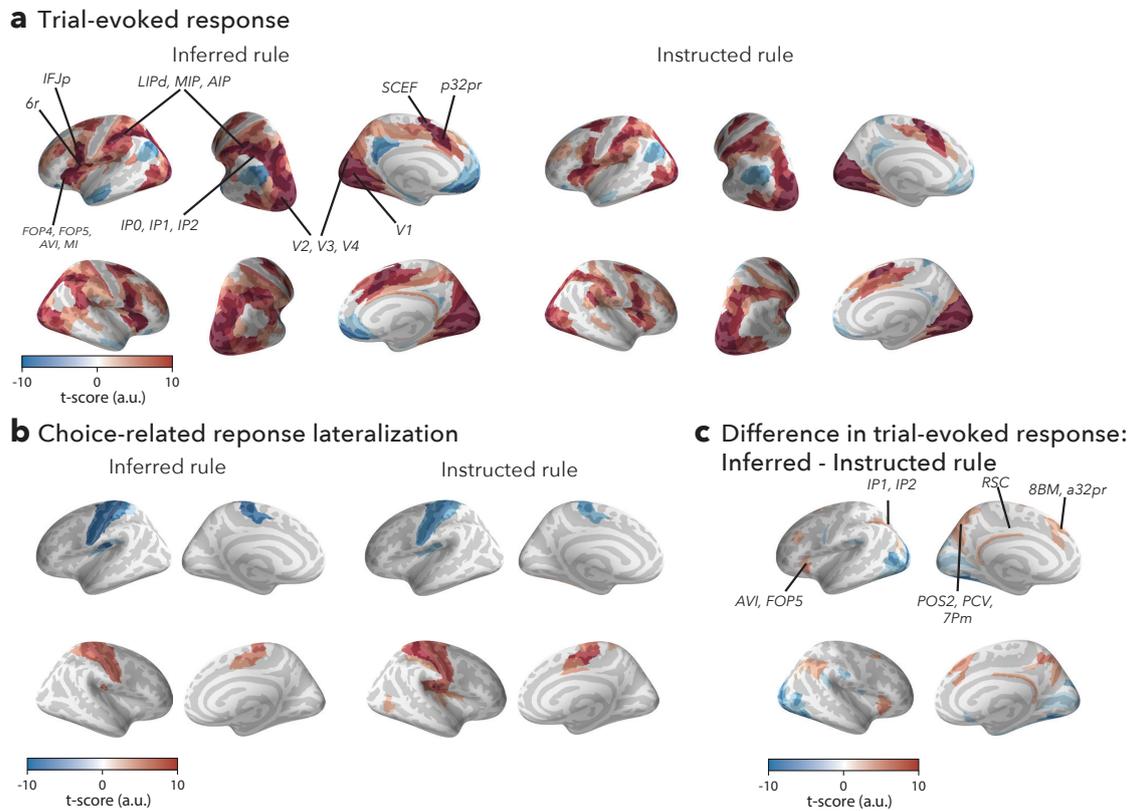


Figure 3: Trial-evoked cortical responses. **a left:** Trial-evoked responses for “Inferred rule” condition. Left hemisphere in the top row, right hemisphere below. Names of areas according to the cortical parcellation (?) with highest correlation across both hemispheres are annotated in the top left row. **right:** Trial-evoked responses for “Instructed rule” condition. **b** Choice-related response lateralization computed as subtraction of regression beta-weights of lateralized responses (i.e. left-right button press) for each condition. Results depicted on both hemispheres. **c** Difference in trial-evoked responses between “Inferred rule” and “Instructed rule” condition. Names of cortical areas significantly stronger activated in “Inferred rule” condition across both hemispheres are denoted in the top row. All panels: FDR-corrected t-statistic across subjects of regression beta-weight.

the onset of the choice) in these candidate regions (Supplementary Fig. 2). Indeed, baseline-corrected fMRI signals were significantly higher during choice epochs in the “Inferred rule” condition compared to the “Instructed rule” control condition, thereby confirming our initial finding in the general linear model (Figure 3c). We, thus, identified a set of cortical regions in the parietal and frontal cortex, which were significantly more active during choices that relied upon an inferred stimulus-response mapping compared to a directly instructed one.

We, next, conditioned the trial-evoked time courses of fMRI signal in these identified cortical regions on the participants’ belief strength regarding the active distribution. We therefore separately visualized the trial-evoked fMRI signal epochs in trials with the lowest and highest 40% of the belief magnitude according to the normative Bayesian model (fitted individually to each subject’s responses) directly before the trial (Figure 4a). This analysis revealed a significant inverse scaling in a subset of these cortical areas. Specifically, in areas in the inferior parietal (*IP1*, *IP2*), the anterior cingulate (*8BM*) and the opercular cortex (*AVI*) neural responses during choice formation were significantly higher when the participants’ belief strength was relatively weak. In other words, responses in these brain regions were larger during low-level sensorimotor decisions, in which the participants were more uncertain about the correct rule. As a behavioral correlate of this finding, reaction times of decisions were inversely correlated with bins of belief strength before a decision (Figure 4b).

To sum up, we found a correlation between the low-level decision and various cortical brain areas, including a pronounced visual component and motor lateralization. Moreover, several higher cortical areas were correlated with the sensorimotor choice, potentially reflecting the retrieval of the correct rule inferred in the higher-level decision process. Importantly, a subset of these higher cortical areas was significantly more active during inference trials compared to instructed rule trials and scaled with uncertainty regarding the correct rule, i.e. the outcome of the higher-level decision process.

2.4 Cortical correlates of the belief updating process

After having identified cortical brain areas active during the formation of the low-level sensorimotor decision, we were next interested in cortical signatures of the higher-level belief updating process. We, therefore, inspected the regression beta-weights in cortical surface regions of three regressors provided by the same general linear

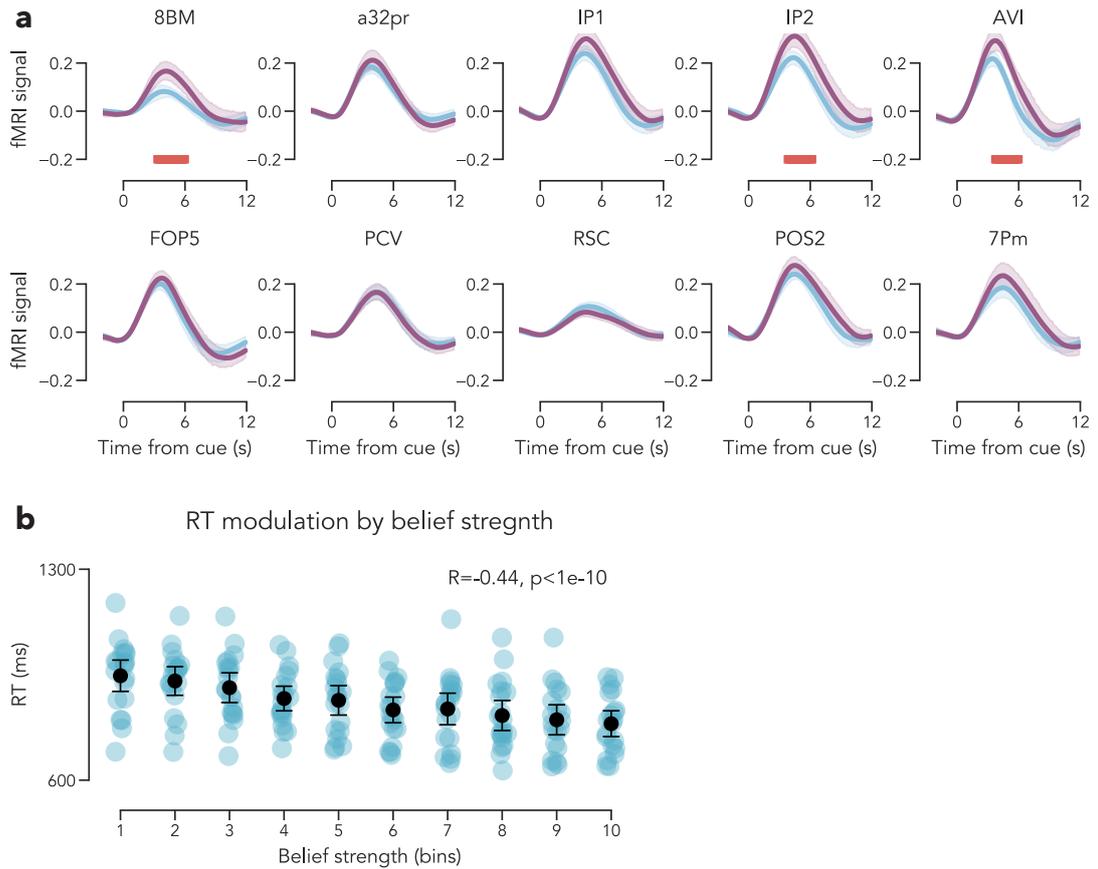


Figure 4: Cortical signature of inference coupling. **a** Shows time courses of trial-evoked activity in areas significantly more active during trials of the “Inferred rule” condition (compare Fig. 3c) conditioned on belief strength (lowest 40% vs. highest 40%) before trial onset. Shading, bootstrapped 95%-CI; statistics, cluster-based permutations test at $p < 0,05$. **b** Shows reaction time binned by belief strength before trial onset. Error bars, 2x s.e.m.; statistics, linear regression.

model formulated above (Equation 1). The regressors that we postulated to capture the higher-level belief-updating process included i) the participants' continuous latent belief strength regarding the currently active distribution (i.e. L_n in the normative model), ii) the magnitude of the evidence value associated with each new sample (i.e. the log-likelihood ratio in the normative model) and iii) the change-point probability associated with each new sample. The first two regressors were directly acquired from fitting the normative Bayesian model of belief updating (?) to the participants' data, while the change-point probability was computed according to a recent study, which found this measure to govern the weighting of new incoming evidence in the model (?).

The magnitude of the log-likelihood ratio, i.e. the magnitude of the evidence value associated with each sample, was positively correlated with activity in the anterior part of the intraparietal sulcus (*AIP*) while exhibiting a whole-cortex spanning pattern of significantly negative correlated areas including early and higher visual cortices, somatosensory and motor cortices as well as regions in temporal, parietal and frontal cortices (Figure 5a). We furthermore assessed correlates of the signed log-likelihood ratio, i.e. not the magnitude but the actual evidence for a sample to be drawn from one of the two distributions in comparison to the other (Figure 5d). Two clusters of cortical areas showed a significant lateralized response, that is, a positive correlation in one hemisphere and a negative correlation in the other. These clusters comprised areas in the medial temporal cortex (*MT* and *MST*) as well as areas in the dorsal visual stream (*V7* and *V3B*). The medial temporal cortex is a well-studied part of the brain and has been canonically associated with the encoding of motion direction (?), but also perceived spatial location of objects (?). The dorsal visual stream has been classically framed as the "where"-pathway of visual processing, endowed with extracting and integrating spatial visual information (?). The positive correlation of the signed log-likelihood ratio with these areas in the left hemisphere and the negative correlation with the same areas in the right hemisphere indicates that these areas are activated when samples appear on the contralateral side of the fixation cross. Belief magnitude was positively correlated with visual areas (*V2*, *V6*) and an adjacent posterior cingulate cortex area (*ProS*) as well as auditory association cortex across both hemispheres. Additionally, belief strength was positively correlated with activity in areas in the frontal (*47m*), the opercular (*FOP*), and the insular cortex (*Ig*) of the left hemisphere. Moreover, belief magnitude was negatively correlated with a cluster of cortical regions spanning parts of inferior and superior

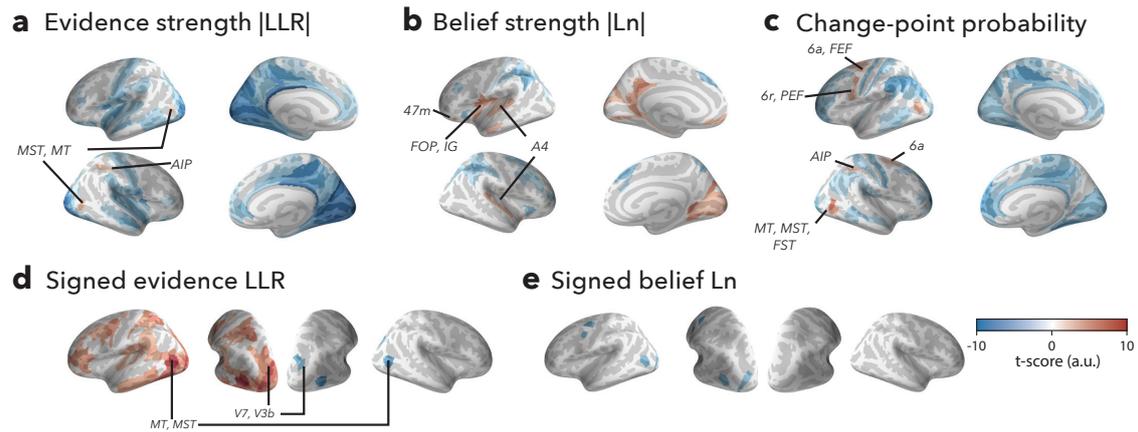


Figure 5: Cortical signature of belief updating. **a** Regression coefficients of magnitude of log-likelihood ratio of new samples. **b** Regression coefficients of magnitude of updated belief variable. **c** Regression coefficients of change-point probability associated with new evidence samples. **d** Regression coefficients of signed LLR of new samples. **e** Regression coefficients of signed belief variable. All panels: FDR-corrected t-statistic across subjects of regression beta-weights.

parietal cortex (*IP1*, *IP2*, *AIP*, *PFm*) as well as orbitofrontal cortex (*AVI*), anterior cingulate cortex (*8BM*), and sensorimotor related area *6ma* (Figure 5b). The signed belief magnitude correlated negatively with two clusters of cortical areas in the medial temporal and the prefrontal cortex of the left hemisphere without a respective correlate of significant correlation in the right hemisphere (Figure 5e). The change-point probability associated with new incoming samples was positively correlated with activation in the medial temporal cortex and surrounding areas (*MT*, *MST*, *FST*) as well as premotor areas (*6a*, *FEF*, *6r*, *PEF*). Moreover, the change-point probability was negatively correlated with visual areas (especially cortical areas constituting the ventral stream of the higher visual cortices), primary motor and somatosensory cortices, areas in the temporal and parietal cortex as well as anterior cingulate and medial prefrontal cortical areas (Figure 5c).

By employing this analysis we tried to disentangle cortical signatures of the slow belief updating process regarding the current state of the environment (i.e. the currently “active” distribution and the associated stimulus-response rule). We, therefore, evaluated the correlation of cortical fMRI signal with several computational regressors,

namely the belief regarding the active distribution (belief strength and signed belief), the evidence value of new samples (again, evidence strength and signed evidence value), and the change-point probability associated with new samples. In sum, we found i) a correlate of the signed evidence value of samples in the medial temporal cortex and the dorsal visual stream ii) increased activation of the medial temporal cortex and the prefrontal area 6a following samples with a high change-point probability, iii) a correlate of the evidence strength with an area in anterior intraparietal sulcus and iv) a signature of belief strength regarding the rule in areas surrounding visual cortex.

2.5 Brainstem nuclei reflect sensorimotor decision in "Inferred rule" trials

We considered neuromodulatory brainstem centers such as the noradrenergic locus coeruleus (LC-NE) as candidate regions to be involved in the interplay between the low-level sensorimotor decision and the high-level belief updating. This was motivated by several insights: First, neuromodulatory brainstem nuclei entertain a widespread pattern of ascending projections which makes them anatomically well-suited structures for modulating cortical activity (??). Second, they are capable of shaping neural parameters such as synaptic gain (?) and the excitation/inhibition-ratio (??) – two mechanisms on the circuit-level, which are intriguing as they have been postulated to critically modulate cortical processing in decision-making (??). And third, brainstem arousal networks have been implicated in various decision-making studies including such that investigate decision-making under uncertainty in changing environments (??).

To test this hypothesis, we first investigated if activity in neuromodulatory brainstem nuclei was correlated with the behavioral regressors representing the two decision-processes, namely the trial-regressor for the sensorimotor-decision and belief strength ($|L_n|$), evidence strength ($|LLR|$), and change-point probability for the higher-level belief updating process. Specifically, we delineated the following subcortical nuclei in the MRI data using probabilistic atlases: Noradrenergic locus coeruleus (LC), acetylcholinergic basal forebrain (BF), serotonergic dorsal raphe nuclei (DRN), and dopaminergic substantia nigra (SNc) and ventral tegmental area (VTA). None of these brainstem nuclei was correlated positively with any of the computational regressors reflecting the higher-level belief updating process, namely belief strength, evidence strength, or change-point probability. Only evidence strength was negatively correlated with activity in sublenticular BF to a significant extent ($p < .05$, FDR-corrected t-test; Supple-

mentary Figure 3). However, several of the delineated subcortical nuclei exhibited robust correlations with the sensorimotor decision ("Choice"). Specifically, these nuclei included LC ($p < 0.001$, FDR-corrected t-test), DRN ($p < 0.001$, FDR-corrected t-test) and SNc ($p < 0.001$, FDR-corrected t-test). Moreover, septal BF was negatively correlated with the low-level sensorimotor decision ($p < 0.01$, FDR-corrected t-test; Supplementary Figure 3). To pinpoint the nature of this association between neuromodulatory brainstem nuclei and the low-level decision, we next extracted epochs of fMRI signal in these delineated nuclei in the time window of -2 to 12s from the onset of the choice grating cue. Again, we found robust responses following choice onset in LC, DR, and SNc ($p < .05$, cluster-based permutations test), corroborating our findings from our general linear model. Paralleling our approach for the cortical regions of interest, we then separated the choice epochs during the "Inferred rule" task condition from those in the "Instructed rule" control condition and observed significant differences in responses in LC, SNc, DR, and sublenticular BF ($p < .05$, cluster-based permutation tests; Figure 6a). This finding indicated that activity in these subcortical regions of interest was stronger when the applied rule was the result of a higher-level inference process compared to control trials in which the correct rule was instructed.

To conclude, while brainstem activity was only slightly negatively correlated with one of our computational regressors reflecting the higher-level belief updating process, we observed a robust relation between a network of brainstem centers and the low-level sensorimotor choice in our general linear model. Moreover, a second model-agnostic analysis approach revealed that this increase in brainstem activity was especially pronounced when the low-level sensorimotor decision was coupled to a higher-level belief updating process.

2.6 Pupil responses reflect change-point probability of evidence samples

Up to this point, we were able to delineate cortical areas involved in both the sensorimotor decision and the higher-level belief updating. Moreover, we showed that both cortical regions, as well as neuromodulatory brainstem centers, were active during the low-level sensorimotor choice and, critically, even more so when the sensorimotor choice was coupled to the outcome of a higher-level decision. Using pupillometry, we were now interested in establishing a connection between the dynamics of the high-level belief updating process and neuromodulatory brainstem systems. Two features of

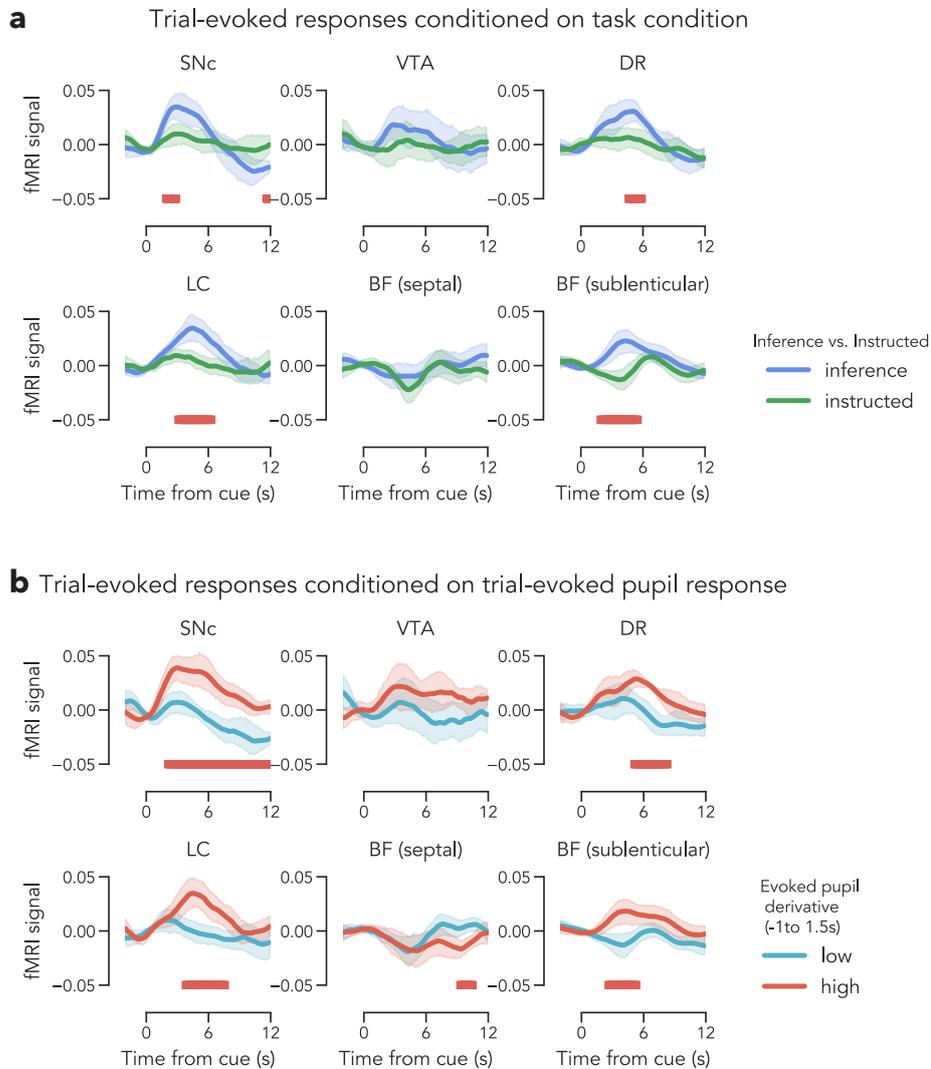


Figure 6: Brainstem activation reflects sensorimotor decision. **a** Trial-evoked fMRI signal responses in subcortical nuclei conditioned on task condition. **b** Trial-evoked fMRI signal responses in subcortical nuclei with single trials conditioned on measure of trial-evoked phasic pupil response (i.e. mean of the first derivative of the pupil diameter in the time window of -1 to 1.5s from sample onset; highest 40% vs. lowest 40% of trial-evoked pupil response). All panels: Shading, bootstrapped 95%-CI; statistics, cluster-based permutations test.

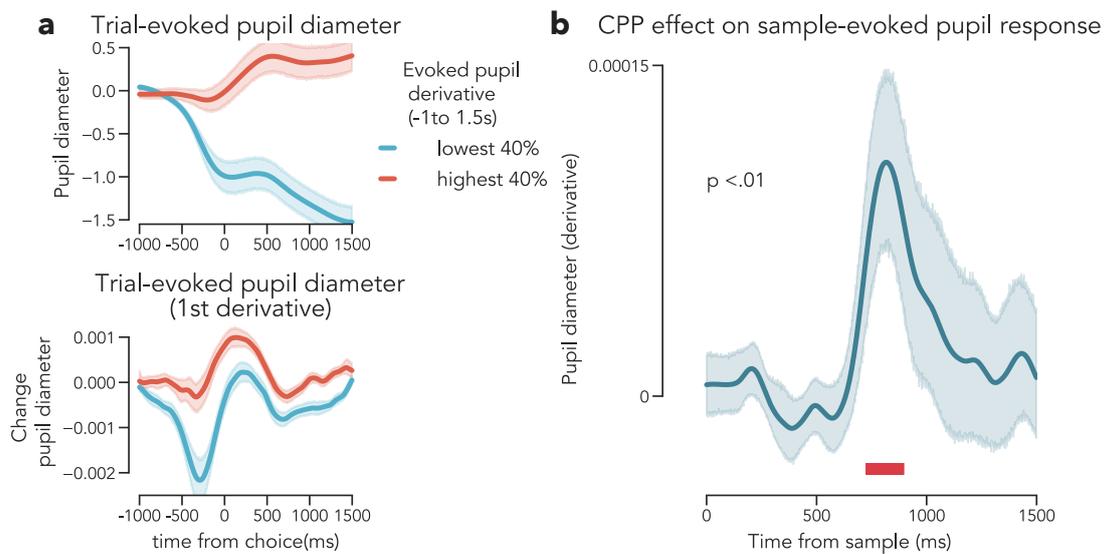


Figure 7: Pupil reflects surprise about new samples. **a top:** Mean pupil diameter upon presentation of new samples averaged across subject and split by measure of phasic pupil reaction, i.e. mean of the first derivative of the pupil diameter in the time window of -1 to 1.5s from sample onset. **bottom** Mean of first derivative of pupil diameter in the same time window again split by phasic pupil measure. Shading, bootstrapped 95% CI. **b** Difference between mean time course of first derivative of pupil epoch of highest 30% of phasic pupil measure and mean first derivative of pupil epoch across all phasic pupil measure bins. Shading, bootstrapped 95%-CI; statistics, cluster-based permutations test.

the tracking of pupil diameter allowed for establishing this link. First, non-luminance-mediated responses of pupil diameter have been shown to reflect activity of various neuromodulatory systems including LC-NE (????), BF-ACh (??) and VTA-DA (?). Second, pupillometry acquires data of pupil diameter at a vastly higher temporal resolution compared to fMRI. The latter is a key advantage of pupillometry over direct imaging of fMRI signal in the brainstem and enabled us to relate the fluctuations in pupil diameter to the fast presentation of single evidence samples. We thereby aimed to establish a connection between properties of the higher-level belief updating process and brainstem arousal systems.

Previous studies have shown that neuromodulatory activity in the noradrenergic locus coeruleus is reflected especially in the first derivative of the pupil diameter (?). We, therefore, chose in line with previous studies (??) the first temporal derivative of the

pupil diameter as a measure of pupil-linked phasic arousal (Figure 7a). We first established the relation between the first derivative of pupil diameter and neuromodulatory brainstem nuclei in our data by conditioning the before extracted choice epochs of fMRI signal in brainstem nuclei on the mean change rate in pupil diameter (i.e. first derivative) during -1 to 1.5s from the response. This comparison yielded a robustly stronger fMRI signal during choices with high evoked pupil response compared to trials with low evoked pupil response in LC, SNc, DR, and subthalamic BF ($p < .05$, cluster-based permutations test; Figure 6b). Next, we extracted epochs of pupil diameter time courses around each evidence sample shown during the high-level belief updating process. We conditioned these pupil time series on the change-point probability associated with each sample and then compared for each subject sample epochs with high change-point probability (highest 30%) with the mean pupil response for all samples. We observed a highly significant elevation of the first derivative of the pupil diameter around 700ms following the onset of the sample in surprising epochs compared to the average, i.e. the pupil dilated faster following surprising samples of evidence in comparison to the general pupil reaction upon presentation of a sample (Figure 7b). This finding indicated a connection between a network of brainstem arousal systems reflected in the first temporal derivative of pupil diameter and samples of evidence associated with a high change-point probability presented during the high-level belief updating process.

3 Discussion

Decision-making, understood as the deliberate selection of an action plan or a belief among several alternatives based on the accumulation of evidence regarding the possible outcomes, is one of the fundamental modules which constitute cognition (?). Decisions are made at various levels of abstraction and timescales, ranging from fast perceptual decisions to high-level value judgments. Increasing insights into the neural mechanisms of the most basic perceptual decision making in the last decades now afford the interrogation of the neural mechanisms that underlie decision making in more ecological conditions - such as environments that are uncertain and undergo sudden changes (??). Here we probed into the dynamics of hierarchical decision-making, necessary in a setting in which the state of an uncertain and unstable environment deter-

mines the best behavioral strategy for more basic sensorimotor decisions. We showed that the participants' behavior was well captured by a dynamic belief updating process (?). Using general linear modeling of fMRI data, we identified cortical regions that were involved in the low-level sensorimotor choice and the application of the inferred rule and areas that reflected computational quantities of the slow belief updating process. Finally, using pupillometry and brainstem fMRI, we established specific patterns of activity of neuromodulatory brainstem regions in both the slow belief updating process as well as the fast sensorimotor decisions.

3.1 Evidence accumulation in changing environments

First of all, our findings replicate important previous insights from studies of decision making in changing environments and show that critical dynamics also hold when the decision is decoupled from the direct motor response. Inferring the generative distribution and the rewarded response rule in our task required evidence accumulation in a noisy and volatile environment. The Bayesian normative framework, which we harnessed in our behavioral analyses, posits that the optimal observer combines a prior belief - non-linearly discounted dependent on the assumed hazard rate - with new incoming evidence to form the new belief about the state of an uncertain environment (?). We were able to show that this computational framework captures our participants' behavior regarding the belief updating process adequately and outperforms a variety of alternative heuristics (Figure 2). In the normative model, through the non-linear dynamics of the discounting, evidence accumulation in unstable environments exhibits a particular sensitivity to the computational variables of change-point probability and uncertainty (?). In particular, the influence of evidence samples associated with a high change-point probability on the final choice is upweighted in comparison to samples in-line with the current belief (i.e. with a low change-point probability). Unraveling neural underpinnings of these computational insights, Murphy et al. indicates that the upweighting of sensory evidence might be mediated by the phasic activity of brainstem arousal systems (?). Again, our finding that pupil-linked brainstem arousal systems are activated following evidence samples indicative of change-points (Figure 7) corroborates this account and is in line with other studies regarding the role of pupil-linked arousal systems in changing environments (??). Importantly, our results provide evidence for these computational and neurobiological accounts of evidence accumu-

lation in volatile environments in novel task conditions. Whereas other tasks statically mapped inferred environmental context onto responses (????), our task design decouples these two. That is, in our task the inferred context is not directly interrogated but conveys information about the rewarded task rule in a coupled sensorimotor decision. Our results thereby indicate that the non-linear dynamics of evidence accumulation as outlined by previous research (??) hold also for more abstract decisions that are decoupled from a motor response. Our finding raises the question of where the evidence integration takes place. When the possible decision outcomes are directly mapped onto motor responses, a neural signature of accumulated evidence is found most prominently in the gradual build-up of choice predictive activity in the motor cortex (?????). Theoretical (??) as well as empirical research (???) suggests that evidence accumulation might be realized in a distributed fashion across multiple nodes of the cortical hierarchy. The decision variable in our task can be behaviorally tracked as the belief in the current state of the environment. However, we do not find plausible cortical areas in which activity is significantly correlated with belief strength. One possible explanation might be the discrepancy between the timescale of belief accumulation, which is determined by the frequency of evidence sample presentation of approximately 2/s, and the temporal resolution of fMRI, defined by the repetition time of 1.9 s. As another possibility, the sensitivity of our general linear modeling approach might not be sufficiently high to detect neural signatures of more subtle computational variables such as the inferred belief about the state of the environment. For example, competing hypotheses (e.g. beliefs) could be represented in the same population of neurons in a distributed fashion. Univariate approaches (such as the standard general linear model) evaluate changes in voxel-wise neural activity in response to certain task manipulations (such as a change in belief). In contrast, multivariate approaches (such as multivariate pattern analysis) are able to detect patterns of fMRI signal across a set of voxels. Multivariate approaches could thus be able to discern the distributed coding of different hypotheses in a cortical region, which is not paralleled by a change in the mean activity of the involved voxels and is, thus, not necessarily detectable in a univariate analysis. Future studies could, therefore, complement our findings by employing neuroimaging techniques with superior temporal resolution such as MEG as well as more sensitive analysis approaches such as multivariate pattern analysis.

3.2 Correlates of sensorimotor choice and task rule

Using general linear modeling we found a network of cortical brain areas significantly correlated with the low-level sensorimotor decision (Figure 3). Among those cortical brain regions with the strongest correlation were visual areas - plausibly reflecting the high-contrast choice cue - but also various higher association cortical regions including the inferior parietal (*IP0, IP2*), the superior parietal (*LIPd, AIP*), the anterior cingulate (*p32pr*) and adjacent (*SCEF*), the premotor (*6r*) and the frontal (*FOP4, FOP5*) cortex. We present two possible explanations for these observed activity patterns, namely i) the activation of attention networks and ii) the representation of inferred task rule sets. First, the pattern of neural correlates of the sensorimotor decision in the cortex might reflect the activation of canonical attention networks (??). The presentation of the choice cue constitutes a highly salient yet temporally unexpected stimulus that interrupts ongoing evidence accumulation in the slower belief updating process and warrants reorientation of attention to the low-level sensorimotor choice. Seminal studies of attention have identified two distinctive functional networks of attention at both task and rest (??): Regions including the intraparietal sulcus and the frontal eye field in both hemispheres constitute the dorsal attention network, while cortical regions including the temporoparietal junction and the inferior frontal gyrus of the right hemisphere form the ventral attention network. The dorsal attention network has been classically associated with top-down directed attention, while the ventral attention network is recruited when unexpected salient sensory signals require a shift of attention (?). Reorienting attention to new demands - such as processing an unexpected stimulus - probably depends upon both attention system, the ventral attention system might act as an interrupt signal inducing the dorsal attention system to shift attention to the novel stimulus (?). A correlation of activity in these functional networks with the onset of the choice cue is, therefore, neurophysiologically highly plausible and, indeed, the pattern of activation observed during sensorimotor decisions matches the key nodes of both attention systems.

Second, the pattern of neural correlates of the sensorimotor decision could also reflect the retrieval of a task rule. Intriguingly, the pattern of cortical activation during sensorimotor choice in our task maps with a host of empirical studies that have used experimental settings, in which different task rules had to be applied to identical stimuli in order to identify neural signatures of such task rules (?). For example, several

fMRI-studies in humans found especially the lateral anterior prefrontal cortex (anatomically overlapping with the inferior frontal sulcus) but also areas in the parietal cortex to contain rule-specific information (????????). Some of these fMRI studies were able to identify rule-specific patterns of activation in these regions using multivariate decoding approaches (????), others found rule-specific patterns of interactions between different regions in the prefrontal cortex (???). In line with this, studies conducted in non-human primates have found rule-selectivity in a frontoparietal network of neural populations (???). Moreover, the high temporal resolution of neural recording studies afforded insights into how task rules are established and updated in the face of changing contexts. Johnston et al. shows that neurons in the anterior cingulate cortex are especially predictive of task rule shortly after a new rule has been established, while the dorsolateral prefrontal cortex is more constantly predictive of task rule suggesting that the ACC might be involved in updating task rules and dorsolateral PFC in maintaining them (?). Another extensive study examining temporal evolution of task information across brain regions shows that information about the task rule (dissociated from the cue itself) first transiently evolves in the inferior temporal cortex and the visual area V4 to then evolve in sustained activation in dorsolateral the PFC and the lateral intraparietal cortex, from where it spreads across the sensorimotor pathway (?). In the light of this line of research, significant activation in prefrontal and parietal regions during the low-level sensorimotor decisions in our task could be interpreted as retrieval and application of task rule sets. We extended these results by comparing correlated cortical activity during sensorimotor choices with our control condition, in which participants were directly instructed on the rewarded rule ("Instructed rule"). We, thereby, delineated cortical regions that were significantly more active when the task rule was inferred compared to when it was instructed (Figure 4). In particular, the ACC was among these cortical areas, which were significantly more active during trials of the "Inferred rule" condition. ACC has been hypothesized to reflect contextual information that is relevant for the outcome of actions and to be involved in mapping these contexts to behavioral strategies (???). As discussed above, when such action-reward contingencies change, ACC has been found to show context-selectivity especially shortly after the change - thereby presumably inducing the update of a formerly maintained action strategy (?). Thus, our finding that ACC is more active during inferred rule trials indicates that the task rule is updated when the onset of the choice grating prompts the formation of a decision and

the commitment to a response. We, moreover, found that in a subset of the regions more active during choices in the “Inferred rule” condition compared to the instructed condition, activity during the sensorimotor choice covaried with uncertainty regarding the inferred rule (Figure 4). An intriguing hypothesis comprises that these differential responses could reflect the encoding of decision uncertainty in these brain regions (?). In an economic decision-making paradigm, fMRI signal responses in posterior midline cortical regions expressed so-called second-order uncertainty, i.e. uncertainty about the currently active action-outcome contingency (?). Another study showed a negative correlation between confidence (i.e. the complement of uncertainty) and fMRI signal responses in the ACC and the inferior frontal gyrus among others which also aligns with our findings (?). Prominently, work from rats performing an odor categorization task showed that a large proportion of neurons in the orbitofrontal cortex encode decision uncertainty caused by noisy task cues (?). However, it is difficult to distinguish encodings of (variable types of) uncertainty from secondary effects of top-down attention discussed above, as decisions with higher uncertainty are naturally more demanding. Consistent with this alternative explanation, various of these regions - in particular the intraparietal sulcus and the anterior-medial PFC - are also involved in cognitive control and frontoparietal attention networks described above (???). Task rule representation, reorientation of attention, and effects of variable types of uncertainty are not mutually exclusive explanations for our observed patterns of cortical activation during sensorimotor choice but could be regarded as complementary. Further analyses should aim at isolating the effects of rule representation through decoding rule information from different cortical brain areas or by measuring the temporal evolution of connectivity between stimulus- and response-specific cortical areas as a marker of instantiation of a rule.

3.3 Encoding of task rule and coupling of decision processes

Thus far, we aimed at pinpointing mechanisms of both constituent processes, i.e. i) the slow belief updating process resulting in a belief regarding the state of the environment and ii) the fast sensorimotor decision, in which the information flow relies on the presumed context. Next, we were interested in the coupling mechanism, i.e. how the dynamic and uncertain belief about the environment shapes the information flow in the sensorimotor decisions. Theoretically, the coupling could be instantiated in vari-

able ways. Two intuitive instantiations comprise i) the explicit encoding of belief in a rule area and ii) the implicit encoding of the rule in the form of (functional) connectivity of associated sensory and motor areas. First, evidence accumulation regarding the state of the environment could result in a latent belief state which is encoded explicitly in some form of rule area – for example in the prefrontal cortex. Upon presentation of the cue signaling the sensorimotor choice, this latent variable could be read out and translated into the according rule which in turn shapes the information flow in the fast decision from sensory to motor cortical areas. Broadly in line with this hypothesis is a recent reinforcement learning study in which errors could be either due to the noisy nature of the stimulus or a covert state change that manifested as a reversed rewarded rule similar to our task (?). The authors showed that ACC in monkeys performing the task encoded accumulated evidence for a rule switch and was moreover predictive of a behavioral switch in the upcoming choice (?). Second, instead of culminating in an abstract representation of the decision outcome in the form of an explicit rule encoding, the belief updating process could also result in the affordance of an action plan (???). Precisely, the ongoing belief accumulation could instantiate the associated task rules “on the fly”, for example in the form of continuously manipulating the functional linkage of prefrontal regions with target-specific and response-specific regions in the visual and motor cortex, respectively. In the case of perceptual decision-making, a large host of research indicates such a more direct processing hierarchy: Instead of computing the decision variable abstractly and then converting the readout into a motor response, perceptual decisions rather entail information integration from sensory to motor areas, whereby sensory evidence is transformed directly into a motor plan. Therefore, characteristic signatures of a decision variable are typically in premotor and motor regions associated with the modality of the response report (Donner et al., 2009; Gold and Shadlen, 2007 reviews seminal findings in non-human primates). Critically, in our task, there is no direct mapping of decision outcome onto motor response. Hence, in analogy to perceptual decision making, the outcome of the slow belief accumulation process in our task could consist of competing provisional action intentions – as in the form of fluctuating instantiations of the task rules. In our data, we did not find a clear signature of an explicit belief encoding area (see Fig. 5b). On the one hand, this could be interpreted as indicative of an implicit encoding of the rule in the form of connectivity between sensory and motor cortical areas. On the other hand, the lack of this finding

could also be due to the above-discussed limitations of our general linear modeling approach. Indeed, most studies that found representations of task rules in higher cortical areas used decoding approaches such as multivariate pattern analysis to estimate information about the task rule contained in specific brain regions (?????). One study specifically showed that different behavioral contexts could be reliably decoded from PFC even when overall population activity was at the baseline level (?). In sum, our analyses can not distinguish reliably between the two putative coupling mechanisms. To probe deeper into these possible mechanistic instantiations of the coupling between a higher-order belief updating process and fast sensorimotor decisions is a prospect of further research. A possible approach could comprise examining the functional connectivity between cue-specific sensory and response-specific motor cortical areas and relating this ongoing measure of “rule instantiation” to higher cortical regions.

3.4 Role of neuromodulatory brainstem centers

Finally, we found neuromodulatory brainstem nuclei to be involved in the process of hierarchical decision-making at several stages in our experiment. First, various subcortical structures - in particular serotonergic DRN, dopaminergic SNc, and noradrenergic LC - were significantly active during sensorimotor choices in the inference condition, as indicated by both general linear modeling and epoch based analyses (Figure 6). Second, we linked phasic responses of pupil-linked arousal systems to the presentation of surprising evidence samples (Figure 7).

Our first observation, that various brainstem regions are significantly active during sensorimotor responses, is in line with influential accounts of neuromodulation - especially in the case of the LC-NE system. It has been shown that LC neurons respond phasically to salient target stimuli (?). A seminal account of LC function states that this phasic response contributes to optimal behavior by increasing the gain of neurons at multiple processing stages in a modulatory fashion (?). LC thereby enhances the processing of the relevant stimulus by acting as a temporal filter (?). This framework can account for our finding of LC responses to the choice cue qualitatively observed in both task conditions. Another influential theory posits noradrenergic LC neurons to specifically fire in response to (temporally) unpredictable task-relevant stimuli instantiating a “network reset” signal (??). In particular, in the inference condition, this might interrupt the ongoing belief accumulation regarding the state of the environment and shift the

attention to the target stimulus. This notion of “network reset” signal also resonates with the function of the ventral attention network mentioned above, which is activated when salient but unexpected stimuli warrant a reorientation of attention, and aligns with some of the cortical regions active during sensorimotor choice (??). In fact, it has been hypothesized that LC responses to unexpected salient stimuli drive activity in the ventral attention network (??). In comparison to the inference task condition, in the “Instructed rule” condition the choice cue is temporally much more predictable and warrants less of an attentional shift. This complies with our observation that LC activity is significantly higher during sensorimotor decisions that rely on inferred rules than during instructed sensorimotor choices.

Besides LC, we also observed robust responses of subcortical dopaminergic structures - most pronounced in SNc, but qualitatively also present in VTA. According to established theories of DA function dopaminergic midbrain structures signal reward prediction errors, i.e. the difference between predicted reward and actual reward (??). In accordance with this framework, the firing rate of dopaminergic midbrain neurons has been found to increase in response to unexpected rewards, mirroring a positive difference between the expected and the received reward. Although in our task participants do not receive a direct reward, the choice cues may be interpreted as conditioned stimuli that have been learned to relate to the amount of the remuneration distributed at the end of a run. Indeed, influential studies show, that after learning stimulus-reward contingencies dopamine neurons respond to the conditioned stimulus instead of the reward itself (?). Again, the unexpectedness of the choice cue in the “Inferred rule” condition might explain the significantly higher responses of dopaminergic neurons in comparison to the “Instructed rule” condition. From a functional perspective, activation of dopaminergic midbrain structures could promote the allocation of limited resources to further engage in profitable behavior - in our case in attending to the presented choice cue and responding with button-press (?).

Our second result regarding pupil dilations relies on the insight that non-luminance-mediated responses of pupil diameter reflect activity in various neuromodulatory brain-stem centers including BF-ACh (??), VTA-DA (de Gee et al., 2017), and in particular LC-NE (?????). Our finding that pupil-linked arousal systems are activated upon presentation of evidence samples associated with a high change-point probability is consistent with canonical theories about LC as well as recent mechanistic insights into deci-

sion making in changing environments (????): A high change-point probability of evidence samples is naturally accompanied by an increase of “unexpected uncertainty”, that is, uncertainty not due to the probabilistic nature of a stimulus itself (“expected uncertainty”), but which arises due to a possible change in the environment. Influential computational accounts relate phasic LC-NE activity to such “unexpected uncertainty” (??). The effect of phasic LC activity is an upweighting of bottom-up processing relative to top-down feedback in order to learn efficiently about the environment and reduce this form of uncertainty (??). Thus, in the context of evidence accumulation in changing environments, samples associated with a high change-point probability elicit phasic activation of pupil-linked arousal systems, which in turn induce an upweighting of the presented evidence (??). This mechanism gives rise to the non-linear dynamics observed in belief updating in changing environments (??). In sum, our results indicate that neuromodulatory brainstem centers – in particular, the LC-NE system – play a critical role in hierarchical decision making by i) upweighting the impact of evidence samples with a high change-point probability on evidence accumulation and ii) redirecting attentional resources when required.

With a rich methodological approach comprising fMRI, pupillometry, and computational modeling of behavioral and neural data, our study affords important insights into decision making in more ecological conditions as constituted by multiple decision levels and a volatile and uncertain environment. Nevertheless, the reader should acknowledge several limitations of this study. First, the fMRI signal – which relies primarily on detecting oxygenated (and deoxygenated) hemoglobin – is only a proxy for neural mass activity, which gives rise to inevitable limitations in relating one to the other (Logothetis, 2008). Second, we performed analyses at the level of anatomical labels (?), which acts as a form of (anatomically informed) spatial smoothing. Future studies should aim at identifying effects on a voxel level. Third, despite huge advances over the last decades, elements of the complex interplay of cortical and subcortical influences on pupil diameter remain unclear (?). Thus, inferring neural activity (e.g. in the LC) from changes in pupil diameter, therefore, entails an irreducible level of uncertainty. Finally, we found a strong correlation between participants’ reaction times and choice cue identity, indicating that participants were able to process vertical choice stimuli faster than horizontal stimuli. This makes intuitive sense, as for vertical stimuli the requested response side corresponds with the side of the generative distribution, irrespective of

the applied rule. Although this should not affect our further analyses – as the rules itself did not bias reaction times – future studies should aim to prevent this bias by orthogonalizing generative distributions and response directions. Fruitful prospects for future research could comprise harnessing imaging techniques with higher temporal resolution such as MEG especially supplemented with state-of-the-art source reconstruction techniques (??). Moreover, different analysis approaches including functional connectivity analysis and decoding based approaches such as multivariate pattern analysis could promisingly complement our study and probe into the questions regarding the mechanism of coupling, that our study raises.

3.5 Outlook

In the here described study we deployed a novel decision-making task that coupled a higher-level inference process regarding the uncertain state of a volatile environment to a simple visual orientation discrimination task without uncertainty. Our task was constructed as such that a latent change in the environment resulted in a reversed mapping of stimulus and rewarded response in the simple visual discrimination task. We fitted a normative model of belief updating in changing environments to the data using a Bayesian posterior sampling approach, which allowed us to assess an ongoing estimate of the participants belief regarding the state of the uncertain environment. Subsequently, we related neural data in the form of fMRI signal to both ongoing computational quantities of the higher-level inference process (i.e. belief updating) as well as the low-level visual discrimination decisions. We, therefore, combined a general linear modeling approach with model-free analyses of fMRI time courses. We focused in these analyses on fMRI signals in neuromodulatory brainstem centers and complemented our analyses with pupillometric data, which constitute an established proxy for activity in brainstem arousal systems. A futile approach for deepening the insights into hierarchical decision-making in volatile environments should evaluate our data from the perspective of functional networks of neuronal populations. For example, multivariate pattern classification could be harnessed to define voxels in retinotopically organized visual cortical areas that encode the orientation of the grating stimulus as well as areas in the parietal and frontal cortex that are selective for the motor response. Spontaneous co-fluctuations between the resulting grating orientation specific activity patterns in visual cortex and the response specific activity patterns in motor corti-

cal areas could then be evaluated as functional instantiations of the inferred rule. This measure could be related to the continuous estimate of the current belief regarding the state of the environment in the top-level inference process, which is accessible to the experimenter through the model-fitting implemented in the present study. Future studies should further relate changes in these patterns of co-fluctuation to neural activity in neuromodulatory brainstem areas as well as to pupillometric data as a proxy for such neuromodulatory systems. Furthermore, future studies should complement our findings by acquiring neural mass data of participants performing our task with using recording modalities with higher temporal resolution such as MEG. Thereby, analytic limitations caused by the relatively poor temporal resolution of functional MRI could be surpassed. For example, neural mass data recorded with MEG could be related to the computational quantities that change at a much higher rate than the fMRI acquisition time such as the participants' belief or the change-point probability of evidence samples.

4 Methods

4.1 Participants

A total of 22 participants (median age 27, range 21 - 44, eight male) took part in our experiment, out of which all but one completed three sessions: one behavioral and two MRI sessions. Another participant was excluded from analyses for continuously poor performance such that in total 20 out of 22 tested participants were considered for further analyses. All participants were healthy individuals with normal or corrected to normal vision recruited via the recruitment pool of the Institute for Neurophysiology and Pathophysiology of the University Hospital Hamburg-Eppendorf. Exclusion criteria included a current or past diagnosis of mental or neurological illness, use of illegal substances, above-average consumption of alcohol as well as non-compatibility with the MRI scanner. Participants gave written informed consent and the study was approved by the Ethics committee of the Hamburg Medical Council. The experiment comprised three sessions, one behavioral training session and two sessions in the MRI-scanner. Participants were remunerated with 10 Euros per hour, fixed 20 Euros for a blood sample, moreover 10 Euros for completing all three sessions and an additional flexible amount, which depended on task performance during all three sessions and

could result in a maximum bonus of 30 Euros.

4.2 Task

Our main task condition comprised a simple two-alternative forced choice (2AFC) task coupled to an inference problem in which participants had to infer the correct rule to be rewarded in the 2AFC task. Specifically, participants were shown a stream of small iso-luminant dots (samples) appearing at a frequency of 2 Hz and with a duration of 100 ms on the imaginary central horizontal of the screen. The distance of samples to the central fixation cross was variable and was sampled from one of two possible overlapping Gaussian normal distributions with equal variance ($\sigma_{\text{left}} = \sigma_{\text{right}} = 1$; arbitrary units) and means symmetric to the fixation cross ($\mu_{\text{left}} = -0.5$, $\mu_{\text{right}} = 0.5$; arbitrary units). Sample positions were drawn from the same distribution (generative distribution) until a change point occurred, which could happen at any time point with the probability of 1/70 (hazard rate). As distributions were overlapping participants were not able to judge with certainty from a single sample, which distribution was the generative one. However, by taking several consecutive samples into account, participants could infer the generative distribution. From time to time (on average every 16.5 s; range 6.8 – 29.6 s), a horizontal or vertical grating stimulus appeared upon which participants were required to respond with a left or right button press. Which response was to be rewarded, depended on the currently generative distribution. That is, the left distribution corresponded to “Rule 1” (vertical grating - left button press/ horizontal grating - right button press) and the right distribution with opposite “Rule 2” (vertical grating-right button press/ horizontal grating-left button press). The response rules stayed constant across all three sessions and participants were instructed at the beginning of each block, which distribution corresponded with which response rule. In a reduced version of the task, participants were instructed the currently active distribution and only had to apply it to the choice stimuli (horizontal or vertical). This task condition served as a control condition for some analyses and for a training purpose, such that participants memorized the correspondence of active distribution and rewarded response rule. We refer to the full and the reduced task conditions as “Inferred rule” and “Instructed rule”.

4.3 Stimuli

All stimuli were created using Matlab and the Psychophysics Toolbox Version 3 (?) and presented on a medium grey background. The evidence samples consisted of a light grey circle with a radius of 0.1 d.v.a. and a surrounding dark grey ring with a total radius of 0.14 d.v.a. The sample points were located on the central horizontal with the distance from the fixation cross sampled from one of two distributions as described above. The choice stimulus was a circular Gabor patch with full black and white contrast and sharp edges with either vertical or horizontal orientation and a spatial frequency of 1.2 cycles per d.v.a. The stimulus had a radius of 13.85 d.v.a. and spanned across the full height of the screen. The Gabor patch was interrupted by a circle of 2.5 d.v.a. in the grey background color, which surrounded the fixation cross in the middle of the screen. The fixation mark was a white symmetric cross in the center of the screen with a length of 0.51 d.v.a. and a thickness of 0.05 d.v.a. In the MRI scanner stimuli were presented on an MRI-compatible LCD screen with a resolution of 1920x1080 pixels at a refresh rate of 60 Hz. The screen was positioned at an approximate distance of 60 cm and viewed through a surface mirror that was mounted on top of the head coil. In the training session in the psychophysics lab stimuli were presented on a VIEWPixx monitor (VPixx Technologies, Saint-Bruno, Quebec, Canada) with the same resolution and refresh rate as the monitor in the MRI scanner. In the scanner, participants reported left or right choices with their left or right hand, respectively, using two MRI-compatible button interfaces (Current Designs, Philadelphia, Pennsylvania, USA). At the end of each run, participants received feedback regarding their performance in the form of percentage correct choices, added monetary reward, and accumulated total monetary reward.

4.4 Procedure

The experiment consisted of three sessions, one training session and two sessions in the MRI scanner. The MRI sessions took place on the same testing day with a break of 105 minutes in between sessions, the training session took place 1-2 days before the MRI sessions. During the training session, participants first performed a run, in which the rule was continuously instructed and had to be applied to memorize the rules. Meanwhile, a retinotopic wedge stimulus was shown to familiarize participants with retinotopic mapping runs in the scanner. Next, participants performed one instructed rule

run, in which the participants were still informed about the correct rule, but had to memorize it during the following trials. Afterward, participants performed five runs of the full task condition in which the correct rule had to be inferred (“Inferred rule” condition). Before each run, participants were shown a visualization of the mappings of rules onto generative distributions to avoid error trials due to false rule association. In each MRI session, we first ran three blocks of retinotopic mapping, after which participants performed three blocks of the full inference task condition on average lasting 609 s (SD = 0.54s) and including 36.0 choices (SD = 1.94) and, thereafter, two blocks of the “Instructed rule” condition on average lasting 604 s (SD = 6.40s) and including 56.5 choices (SD = 1.44). As during the training session, participants were reminded of the correct rule-distribution mapping before each run.

4.5 Behavioral model

We used a recently established normative model for inference problems in noisy and volatile environments to capture the high-level belief updating process necessary in our task to infer the currently generative distribution and the associated rewarded response rule (?). In principle, the model entertains a belief variable regarding the state of the environment with two mutually exclusive possible states and governs through an update rule how new evidence and prior belief are integrated into a newly updated belief. Specifically, at each time step n (defined by a new arriving piece of evidence), the belief L_n is updated by adding the log-likelihood ratio LLR_n of sample X_n to the discounted prior belief $\psi(L_{n-1}, H)$:

$$L_n = \psi(L_{n-1}, H) + LLR_n \quad (2)$$

LLR_n is computed as the ratio of the probabilities of the sample X_n to be observed under each of the two states of the environment, on the logarithmic scale. In our task, the states represent the two normal distributions, from which the dot position is potentially sampled. Accordingly,

$$LLR_n = \log(\text{normal}(X_n | \mu_1, \sigma_1^2)) - \log(\text{normal}(X_n | \mu_2, \sigma_2^2)) \quad (3)$$

Critically, the prior belief $\psi(L_{n-1}, H)$ depends on the posterior belief after the previous sample L_{n-1} and the subject’s estimate of the hazard rate H , i.e. the risk of the

currently "active" distribution to switch.

$$\psi(L_{n-1}, H) = L_{n-1} + \log \left[\frac{1-H}{H} + \exp(-L_{n-1}) \right] - \log \left[\frac{1-H}{H} + \exp(L_{n-1}) \right] \quad (4)$$

This discounting is a non-linear transformation and constitutes the key difference of this model in comparison to canonical evidence accumulation models such as perfect accumulation (?). From equation 4 follows, that the hazard rate H scales the translation of the previous posterior belief L_{n-1} into the current prior belief $\psi(L_{n-1}, H)$. When $H = 0$, ψ equals L_{n-1} , resulting in perfect accumulation of samples. On the other extreme, when $H = 0.5$, the two terms in equation (4) cancel out and L_n equals LLR_n . Thus, H balances the impact of new evidence and prior belief on the current and thereby governs the tradeoff between the precision of the belief and the sensitivity to change-points (Figure 2a-c).

Choice probabilities were computed from the posterior belief L_n following a logistic function scaled by a free parameter V , representing internal noise in the neural circuitry:

$$p_{n,\text{left}} = 1/(1 + \exp(-L_n/V)) \quad (5)$$

4.6 Model fitting

We fitted the above described normative model of belief updating to our data using a Bayesian modeling approach. Bayesian statistical approaches seek to provide the probability distribution of a parameter over the possible parameter space (i.e. the posterior distribution) given certain observed data and prior knowledge about the parameter. Specifically, Bayesian techniques compute posterior distributions of parameters according to Bayes' theorem, which states

$$P(\text{parameter} | \text{data}) = \frac{P(\text{data} | \text{parameter})P(\text{parameter})}{P(\text{data})} \quad (6)$$

Due to the denominator term in this equation, the posterior distribution cannot be solved by calculus, giving rise to the necessity of sampling algorithms. The rationale behind these is to first sample (random) parameter values and then decide on whether to accept or reject these parameter values based on the log-posterior (i.e. the result of the numerator of the equation). By repeating these steps many times, one can approx-

imate the exact posterior distribution. To keep the sampling process efficient, a proper algorithm strives to keep both rejection rates and autocorrelation between consecutive proposed parameter values as low as possible. Rejections are counterproductive as they imply the use of computing power (proposing a parameter value and calculating its log-posterior) without gaining insight into the posterior distribution. Autocorrelation impedes the sampler in exploring the whole parameter space, thus decreasing effective sampling size and making the sampler less efficient. However, modern sampling algorithms deal with these constraints, making Bayesian methods efficient and feasible. In our analyses, we used the Bayesian modeling language *Stan* (*PyStan: the Python interface to Stan*, Version 2.17.1.0) which uses a Hamiltonian Monte Carlo algorithm for sampling (?). Hamiltonian Monte Carlo, a specific type of Markov Chain Monte Carlo algorithms, harnesses laws from Hamiltonian mechanics in order to propose parameter values in the first place that are more likely to be accepted. We fitted the model with the hazard rate H and the internal noise V as the only free parameters for each subject and each session independently. As priors we used a uniform distribution over the whole possible parameter space between 0 and 1 for H and a broad Gaussian normal distribution centered on 1 ($\sigma = 50$) with a lower bound of 1 for V . The choice of these highly uninformative priors was motivated by our lack of solid a priori knowledge about the subjects' behavior. We used four sampling chains with 5000 sampling iterations each, of which the first 50% were considered as warm-up samples and were not used to determine the posterior distribution. This Bayesian fitting procedure yielded a posterior distribution for both parameters for each subject and session. For further analyses such as computing the belief of the normative model L_n across time, we used the modal values of the posterior distributions, i.e. the most probable parameter values.

4.7 Alternative models

We compared the normative Bayesian model of belief updating to three alternative models or heuristics: i) perfect accumulation of all samples seen throughout the run (equals $H = 0$), ii) a heuristic of choosing only on basis of the last-seen sample (equals $H = 0.5$) and iii) perfect accumulation between choices and a reset after each choice ("reset accumulation"). We ran these models on the actual data of all participants and then compared performance, i.e. the percentage of rewarded trials, between these alternative models, the normative model (equipped with the true generative $H = 1/70$

as well as with the fitted H of participants) and the actual performance of participants.

4.8 Computational parameters of the belief updating process

We used the model fits to extract relevant computational measures that capture the belief updating process and could in later steps be correlated with neural data. These measures were belief, LLR and change-point probability. We computed time courses of the updated belief variable L_n according to the normative model and used the subject and session-specific fits of the hazard rate H and we computed time courses of LLR according to sample positions and Equation 3. Moreover we computed a measure of change-point probability (CPP), derived from the normative model, at each time point n as follows:

$$\text{CPP}_n = \frac{H[\text{N}(X_n | S_1)p(S_{2,n-1}) + (\text{N}(X_n | S_2)p(S_{1,n-1}))]}{H[\text{N}(X_n | S_1)p(S_{2,n-1}) + (\text{N}(X_n | S_2)p(S_{1,n-1}))] + (1 - H)[\text{N}(X_n | S_2)p(S_{2,n-1}) + (\text{N}(X_n | S_1)p(S_{1,n-1}))]} \quad (7)$$

(Equation 8)

where X_n denoted the value (in our case dot position) of a new evidence sample and S_1 and S_2 the two possible generative distributions with respective mean and variance μ_1, σ_1^2 and μ_2, σ_2^2 . Two reasons motivated our choice of this computational quantity (?): First, the dynamics of this measure are easy to grasp in the context of the normative model. For example, in the case that the LLR_n of a new sample n contradicts the previous belief L_{n-1} this measure will naturally yield a high change-point probability CPP. On the other hand, if a new sample is neutral in regards to the two possible generative distributions ($\text{LLR} = 0$), this measure of CPP equals the model's hazard rate H , which intuitively makes sense as it constitutes the general probability of a change-point to occur. Second, CPP has been found to upweight the impact of a sample on the final choice (?).

4.9 FMRI data acquisition

We used a 3T Siemens Prisma scanner with a 32-channel head coil for data acquisition. The cardiac cycle and respiratory rate were recorded at a frequency of 1000 Hz. We recorded echo-planar imaging (EPI)-images for two types of runs, main experimental runs, and retinotopic mapping runs. EPI-images for main experimental runs were acquired with the following parameters: 62 slices of 2 mm thickness, $\text{TR} = 1.9\text{s}$,

TE = 28ms, FA = 70°, in-plane resolution = 2x2mm. EPI-images for the retinotopic mapping runs were acquired with the following parameters: 26 slices of 2 mm thickness, TR = 0.88s, TE = 28ms, FA = 50°, in-plane resolution = 2x2mm. Moreover, a T1-weighted anatomical scan was acquired in a MPRAGE sequence and with the following parameters: voxel size = 1x1x1mm, TR = 2.3s, TE = 2.98ms, FA = 9°. Finally, we obtained B0-fieldmaps with the following parameter: voxel size = 2x2x2mm, TR = 0.678s, TE1 = 5.42ms, TE2 = 7.88ms, FA = 40°.

4.10 FMRI preprocessing

The functional MRI data were preprocessed using the recently introduced preprocessing pipeline *fMRIPrep*(?). *fMRIPrep* is based on *Nipype*, a python software package used to construct workflows for fMRI analysis, and exploits functionality from several well-established programs, including ANTs, FSL, AFNI, and FreeSurfer. The *fMRIPrep* pipeline consists of two major workflows, for pre-processing anatomical images and functional EPI-images respectively. Specifically, T1-weighted (T1w) anatomical images were corrected for intensity non-uniformity and skull stripped using *N4BiasFieldCorrection* (Tustison et al. 2010) and *antsBrainExtraction.sh* (both ANTs v2.1.0) respectively. The cortical surface was reconstructed using FreeSurfer's *recon-all* functionality (FreeSurfer v6.0.1, Dale et al. 1999). In a customized procedure similar to *Mind-Boggle* (?), the resulting brain matter segmentations of ANTs and FreeSurfer were reconciliated. The T1w image was then registered to MNI 152 *Nonlinear Asymmetrical template version 2009c* through nonlinear registration (?) using *antsRegistration* (ANTs v2.1.0). Finally, *fast* (FSL v5.0.9) was used for the segmentation of white matter and cerebrospinal fluid (?). Functional EPIs were corrected for slice timing and head motion using *3dTshift* (AFNI v16.2.07, Cox and Hyde 1997) and *mcflirt* (FSL v5.0.9, Jenkinson et al. 2002). Furthermore, B0-fieldmaps were used to correct for susceptibility distortion via *fugue* (FSL v5.0.9, Jenkinson 2003), and the images were co-registered to T1w-space, MNI152-space, and FreeSurfer Surface space using boundary-based registration (?) *bbregister* (FreeSurfer v6.0.1). The transformations that resulted from these different corrections and co-registration procedures were applied together in one single step using *antsApplyTransforms* (ANTs v2.1.0) using Lanczos interpolation.

4.11 General linear model

We constructed a multiple linear regression model to predict time courses of single-voxel fMRI signal by a set of behavioral parameters. In detail, this set of predictors comprised i) the sensorimotor choice, quantified as a boxcar regressor lasting from the onset of the choice cue until response reported by button press, ii) the log-likelihood ratio of each incoming sample of evidence computed according to Equation 3, iii) posterior belief L_n according to Equation 2 and iv) change-point probability as in Equation 8. The categorical choice regressor was dummy-coded for the four possible stimulus-response combinations as well as for non-responses, the other three continuous regressors were included in the signed form as well as in absolute form (magnitude). All behavioral regressors were convolved with a canonical hemodynamic impulse response function, resampled to the EPI acquisition frequency ($TR = 1.9s$) and normalized per session and subject by z-scoring.

$$\begin{aligned} \text{fMRI signal}_{\text{vox}} = & \beta_{0,\text{vox}} + \beta_{1,\text{vox}} \mathbf{L}_{\text{abs}} + \beta_{2,\text{vox}} \mathbf{L}_{\text{signed}} + \beta_{3,\text{vox}} \mathbf{LLR}_{\text{abs}} \\ & + \beta_{4,\text{vox}} \mathbf{LLR}_{\text{signed}} + \beta_{5,\text{vox}} \mathbf{CPP} + \beta_{6,\text{vox}} \mathbf{trial}_{\text{horiz, left}} \\ & + \beta_{7,\text{vox}} \mathbf{trial}_{\text{horiz, right}} + \beta_{8,\text{vox}} \mathbf{trial}_{\text{vert, left}} \\ & + \beta_{9,\text{vox}} \mathbf{trial}_{\text{vert, right}} + \beta_{10,\text{vox}} \mathbf{trial}_{\text{missed}} \end{aligned} \quad (1 \text{ revisited})$$

4.12 Cortical regression results

For cortical regression results, we extracted data on FreeSurfer's cortical surface from the 3-dimensional regression results in subject-specific T1w-space using functionality from the python package *nilearn* (*nilearn*, version 0.6.0; Abraham et al., 2014). We averaged across vertices within each cortical region according to a surface-based cortical parcellation (?) to compare regression results across subjects. We further averaged across the two sessions per subject, t-tested across subjects, and averaged the resulting t-statistic across hemispheres for visualization. For all cortical figures, t-statistics were corrected for multiple comparisons using FDR-correction (?).

4.13 Delineation of brainstem ROIs

We used probabilistic atlases to delineate the brainstem regions of locus coeruleus (?), dorsal raphe (?), ventral tegmental area, substantia nigra (Pauli et al. 2018) and the basal forebrain (Zaborszky et al. 2008). The latter was furthermore divided in a sub-lenticular part and a septal part (Zaborszky et al. 2008). The atlases were transformed from MNI152-space to subject-specific T1w-space using the *antsApplyTransforms* functionality from the Advanced Normalization Tools (ANTs v2.1.0) toolbox. To compute ROI-level time series we extracted voxels above a threshold of 0.01 and calculated a weighted average of these voxels over time. To obtain ROI-specific regression results we also calculated the weighted average of the product of the probabilistic atlas and the three-dimensional beta maps resulting from the regression analyses. In analogy to the cortical results, we averaged across sessions, t-tested across subjects and corrected for multiple comparisons using FDR-correction (?).

4.14 Pupil data acquisition and preprocessing

Eye data were acquired using an EyeLink 1000 Long Range Mount (SR Research, Osgoode, Ontario, Canada) positioned outside the MRI scanner and redirected to the participants' pupil via a mirror, which was attached to the head coil. Eye data was recorded at a frequency of 1000 Hz, the tracker was calibrated once at the beginning of each session and EyeLink's in-house software was used to keep track of the pupil. Blinks and other periods, during which the participant's lids were closer, were identified by thresholding the pupil diameter. The detected periods were interpolated linearly. Detection of blinks and interpolation was controlled and corrected manually. A third-order Butterworth filter (0.01-6 Hz) was applied and the time series was z-scored.

4.15 Epoch based analyses

Throughout our analyses, we deployed several epoch-based analysis approaches combining pupil data as well as fMRI data with behavioral parameters.

Choice epochs. We extracted choice epochs from cortical areas (Figure 4 and Supplementary Figure 2) as well as from brainstem ROIs (Figure 6) in both "Inferred rule" runs as well as "Instructed rule" control runs. Cortical ROIs were chosen based on the results from our general linear model and extracted using the HCP MMP 1.0 atlas, an

in-vivo parcellation of the human cortex with 180 areas per hemisphere (?). Specifically, from the time courses of these cortical ROIs we extracted choice epochs as the fMRI signal from -2 to 12 s in reference to the onset of the choice stimulus (vertical or horizontal). We discarded all epochs containing blinks in the concurrent pupil data and considered only subjects with at least 30 remaining blink-free trials. Applying this conservative approach to the epoch-based analysis, 17 subjects remained for the inference condition and 14 subjects remained for the "Instructed rule" condition. In analyses, in which both conditions were compared, only data of the remaining 14 subjects were used. Choice epochs were then binned by either task condition (Figure 6a), belief strength (i.e. the belief magnitude after the last sample seen before the decision; highest 40% vs. lowest 40%; Figure 4a) or our measure of pupil modulation (mean first derivative of pupil from -1 to 1.5s from choice; highest 40% vs. lowest 40%; Figure 6b) averaged per subject and bin and then averaged across subjects per bin. Differences in resulting time courses between bins were statistically tested using cluster-based permutations tests.

Sample Epochs Moreover, we extracted pupil epoch data around each sample shown during the belief updating period. Specifically, we again used the first derivative of pupil diameter (Figure 7a), which has been shown to especially covary with noradrenergic neuromodulatory systems (?). Per subject, we conditioned these pupil epochs based on the surprise (i.e. change-point probability) associated with the shown sample and subtracted the average across all surprise bins from the 30% most surprising samples. Afterward, we averaged across subjects and used cluster-based permutations tests to test against zero (Figure 7b).

5 Summary

In the present study we probed into the dynamics of hierarchical decision-making, necessary for adaptive behavior in a setting in which the state of an uncertain and unstable environment determines the best behavioral strategy for more basic sensorimotor decisions. We combined a novel decision-making task with fMRI and pupillometry. Through sophisticated Bayesian computational modeling of behavioral data we showed that the participants' behavior was well captured by a dynamic belief updating process regarding the state of the uncertain environment and we gained access to continuous estimates of computational quantities of this updating process. By relating these computational readouts to fMRI signal through general linear modeling and model-free analyses, we identified i) brain regions in which activity reflected important components of the belief updating process and the basic sensorimotor decisions as well as ii) cortical brain areas that were significantly more active in trials that required a coupling of the basic decisions to the belief updating process. Critically, brainstem imaging revealed that activity in certain neuromodulatory brainstem centers was also enhanced during such trials that required a coupling. Finally, pupillometry allowed us to relate pupil diameter - an established proxy for brainstem arousal systems - to change-point probability, a measure for the probability that the underlying environment has undergone a state change.

Die vorliegende Studie beschäftigt sich mit hierarchischen Entscheidungsfindungsprozessen. Diese sind für adaptives Verhalten notwendig, wenn eine unsichere und sich verändernde Umwelt die optimale Reaktion auf Umweltreize und damit die Strategie in basalen Entscheidungen bestimmt. Wir kombinierten hierfür ein neuartiges Entscheidungsfindungsexperiment mit funktionaler MRT und Pupillometrie. Durch Bayesianische rechengestützte Modellierung von Verhaltensdaten konnten wir zeigen, dass sich das Entscheidungsverhalten der Proband:innen durch ein normatives Modell von Glaubensaktualisierung beschreiben lässt. Wir setzten Parameter dieses Modells in Relation zu fMRT-Signalen und konnten so zum einen kortikale Hirnareale identifizieren, die sowohl Komponenten des Prozesses der Glaubensaktualisierung als auch der basalen Entscheidungen reflektierten, sowie zum anderen Areale darstellen, die signifikant aktiver waren, wenn die basalen Entscheidungen hierarchisch von dem Glauben bezüglich des Zus-

tandes der volatilen Umwelt abhingen. Ferner konnten wir mittels fMRT-Bildgebung des Hirnstamms zeigen, dass auch die Aktivität in bestimmten neuromodulatorischen Hirnstammkernen signifikant höher war, wenn basale Entscheidungen hierarchisch an einen solchen übergeordneten Glauben gekoppelt waren. Zuletzt konnten wir mittels Pupillometrie die Größe des Pupillendurchmessers - einen etablierten Proxy für im Hirnstamm lokalisierte Erregungsnetzwerke - mit einem Maß für die Wahrscheinlichkeit eines grundlegenden Zustandswechsels der volatilen Umwelt korrelieren.

6 Abbreviations

ACC anterior cingulate cortex

ACh acetylcholine

BF-ACh acetylcholinergic basal forebrain

CCP change-point probability

DA dopamine

DRN dorsal raphe nuclei

EPI echo-planar imaging

fMRI functional magnetic resonance imaging

GLM general linear model

LC locus coeruleus

LC-NE noradrenergic locus coeruleus

LLR log-likelihood ratio

MEG magnetoencephalography

MT middle temporal area

NE norepinephrine

PFC prefrontal cortex

ROIS regions of interest

SNc substantia nigra

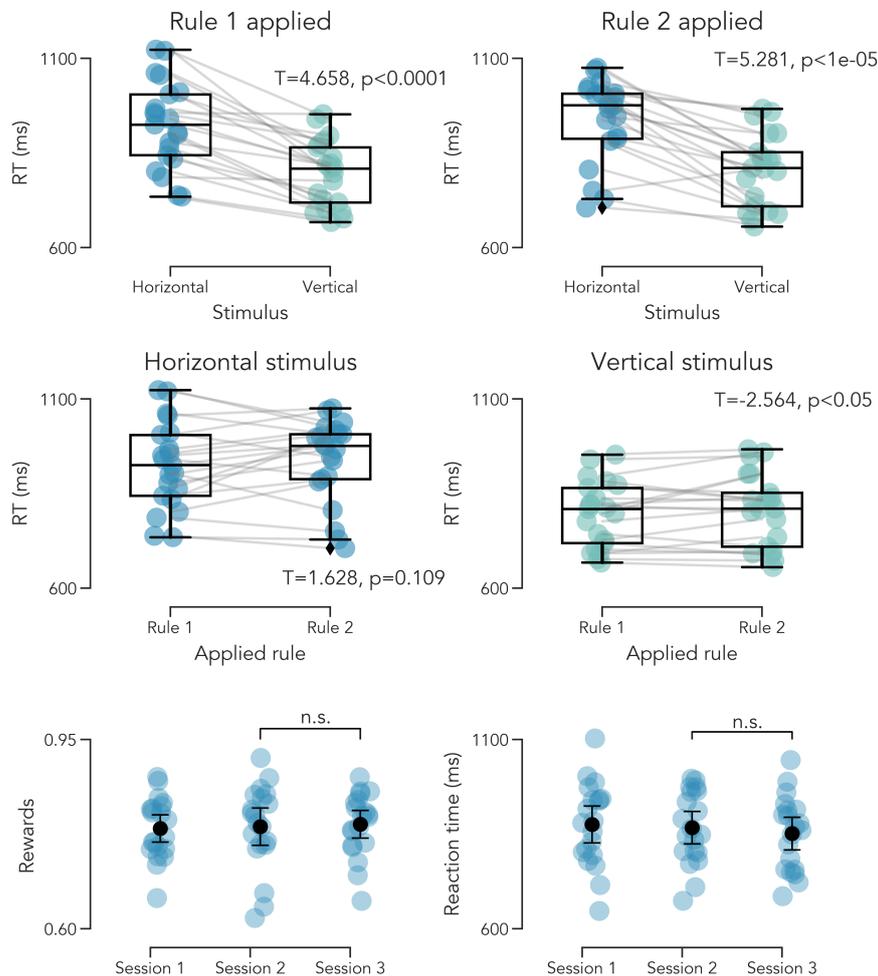
TR repetition time

VTA ventral tegmental area

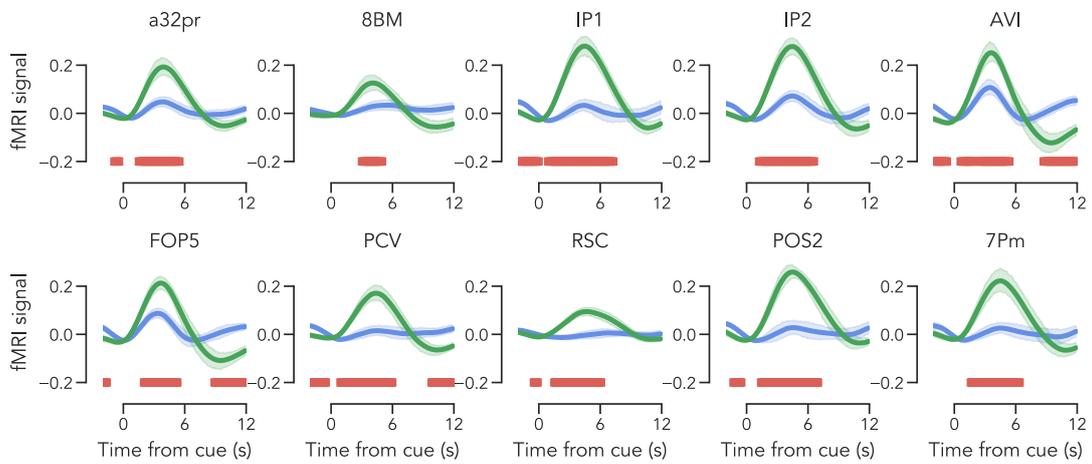
VTA-DA dopaminergic ventral tegmental area

2AFC two-alternative forced choice

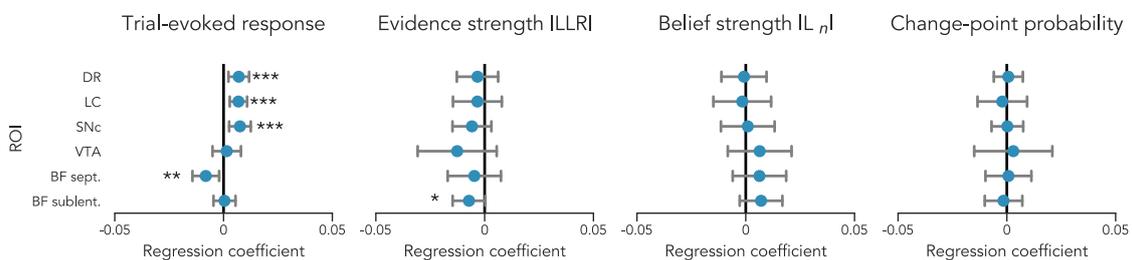
7 Supplementary Materials



Supplementary Figure 1: *top left:* Reaction times of participants in trials in which participants applied "Rule 1" split by the identity of the choice cue (i.e. "horizontal" or "vertical"). *top right:* Reaction times of participants in trials in which participants applied "Rule 2" split by the identity of the choice cue. *center left:* Reaction times of participants in trials with a horizontally grated choice stimulus split by applied rule (i.e. "Rule 1" or "Rule 2"). *center right:* Reaction times of participants in trials with a vertically grated choice stimulus split by applied rule. *top and center:* boxplot depicts median and quartiles across participants; single data points depict single participants' means; statistics two-sample t-test. *bottom left:* % correct trials across sessions. *bottom right:* Reaction times across sessions. *bottom row:* mean and 2x s.e.m depicted; single data points show single participants; statistics across session 2 and 3: two-sample t-test.



Supplementary Figure 2: Time courses of trial-evoked activity in areas significantly more active during trials of the “Inferred rule” condition (compare Fig. 3c) conditioned on task condition (“Inferred rule” vs. “Instructed rule”). Shading, bootstrapped 95%-CI; statistics, cluster-based permutations test at $p < 0,05$.



Supplementary Figure 3: Regression coefficients of selected regressors from the general linear model (see Eq. 1) for neuromodulatory brainstem centers. Mean and 2xs.e.m. across participants depicted. T-tests for significance against zero. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

8 Acknowledgements

Mein großer Dank gebührt Niklas Wilming für das herausragende Mentoring, die Geduld und die vielen Denkanstöße auf dem Weg.

Darüber hinaus möchte ich meinem Doktorvater Tobias Donner danken, für die Mengen wertvollen Feedbacks, den ansteckenden Enthusiasmus sowie die familiäre Lab-Kultur, die du über die Jahre geschaffen hast.

Ferner möchte ich Andreas Engel und dem gesamten Institut für Neurophysiologie für das breitere wissenschaftliche Umfeld danken. Dem SFB 936 sowie der Studienstiftung des deutschen Volkes bin ich zudem zu Dank verpflichtet für die finanzielle und nicht-materielle Unterstützung meines Dissertationsvorhabens.

Die Arbeit an meiner Dissertation hätte mir nicht so viel Spaß bereitet ohne all die fantastischen Arbeitskolleg:innen im Donner Lab. Danke für all die Hilfestellungen, die inhaltlichen sowie die weniger inhaltlichen Diskussionen, die Tischtennis-Duelle und die Kaffee-Pausen.

Zu guter Letzt möchte ich mich bei meinen Eltern für ihre bedingungslose Unterstützung bedanken sowie bei meinen Freund:innen (- im Besonderen Jonas, Charlotte und Ricarda) für das Teilen aller Hochs und Tiefs zwischendrin.

Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

9 Eidesstattliche Versicherung (Declaration of academic integrity)

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: