Aus der Klinik und Poliklinik für Nuklearmedizin
der Universität zu Köln
Direktor: Universitätsprofessor Dr. med. A. Drzezga

Developing an fMRI paradigm for studying reinforcement learning with gustatory stimuli

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Omer Reiner
aus Haifa Israel

promoviert am 21. November 2022

Dekan: Universitätsprofessor Dr. Med. G. R. Fink

1. Gutachterin oder Gutachter: Universitätsprofessor Dr. med. T. van Eimeren
2. Gutachterin oder Gutachter: Universitätsprofessor Dr. med. Dipl.-Psych. J. Kambeitz

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich keine Unterstützungsleistungen

Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Der dieser Arbeit zugrunde liegende Ethikantrag wurde ohne meine Mitarbeit in der Klinik und Poliklinik für Nuklearmedizin von Herrn Univ.-Prof. Dr. Med. Thilo van Eimeren und Herrn Dr. med. Hendrik Theis gestellt.

Die Rekrutierung der Probanden für diese Studie wurde von Herrn Patrick Weyer durchgeführt.

Das Computerprogramm für die Erstellung der Verhaltensaufgabe wurde zum Großteil von mir geschrieben, geringfügig basierend auf Vorarbeiten der Forschungsgruppe "Translational Neurocircuitry" des Max-Planck-Instituts für Stoffwechselforschung. Die verwendeten Bilddateien stammen dabei aus derselben Forschungsgruppe.

Die Probandenaufklärungen, die Gewichts- und Größenmessungen, die Blutabnahmen und die Geschmackstests sind von mir selbstständig durchgeführt worden. Das Computerprogramm für die Geschmackstests stammte von der Forschungsgruppe "Translational Neurocircuitry" des Max-Plancks-Instituts für Stoffwechselforschung und wurde geringfügig von mir modifiziert.
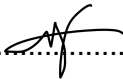
Die dieser Arbeit zugrunde liegenden fMRT-Messungen sind von mir mit Unterstützung der medizinisch-technischen Radiologieassistenten Herr Patrick Weyer, Herr Kurt Wittenberg und Frau Elke Bannemer durchgeführt worden. Die anatomischen MRT-Messungen wurden ohne meine Mitarbeit von den oben genannten Personen durchgeführt.

Die Untersuchungen wurden nach entsprechender Anleitung durch Herrn Univ.-Prof. Dr. Med. Thilo van Eimeren, Frau Dr. Bojana Kuzmanovic und Herrn Dr. Lionel Rigoux von mir selbstständig ausgewertet.

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Köln, den 24.03.2022

Unterschrift: ....

Acknowledgments

Table of Contents

List of Abbreviations

BMI	body mass index
BOLD	blood-oxygen-level-dependent (imaging)
CT	computed tomography
DaT-Scan	dopamine transporter scan
DEXA	dual-energy X-ray absorptiometry
EEG	electroencephalography
EPI	echo-planar imaging
fMRI	functional magnetic resonance imaging
fNIRS	functional near-infrared spectroscopy
FOV	field of view
FSL	FMRIB (Oxford Centre for Functional MRI of the Brain) Software Library
FTO	fat mass and obesity-associated (gene)
GLM	general linear model
HOMA-IR	homoeostatic model assessment of insulin resistance
MEG	magnetoencephalography
MNI	Montreal Neurological Institute
MRI	magnetic resonance imaging
OECD	Organisation for Economic Co-operation and Development
PE	prediction error
PET	positron emission tomography
PST	probabilistic selection task
SAV	state-action value
SNR	signal-to-noise ratio
SPECT	single-photon emission computed tomography
SPM	statistical parametric mapping
TE	echo time
TI	inversion time
TR	repetition time
VAS	visual analogue scale
VBA	variational Bayesian analysis

1. Zusammenfassung / Summary

Zusammenfassung

Eine der größten Herausforderungen für die globale Gesundheit ist die weltweit steigende Prävalenz von Adipositas. Die Untersuchung der Entscheidungs- und Lernprozesse für fehlreguliertes Essverhalten und das konsekutiv bessere Verständnis dieser ist entscheidend für die Bekämpfung der Adipositas-Epidemie.

Das Ziel dieser Studie ist die Entwicklung und Validierung eines Experiments, welches diese Prozesse untersucht.

Zu diesem Zweck untersuchten wir zehn gesunde Probanden mittels einer modifizierten „Probabilistic Selection Task" (PST) – einer wissenschaftlich etablierten Aufgabe zur Bewertung von Lernverhalten. In der ersten Phase müssen die Probanden zwischen zwei visuellen Reizen wählen, die jeweils eine unterschiedliche Gewichtung aufweisen und entsprechend mit unterschiedlicher Wahrscheinlichkeit zu einer Belohnung führen werden. Z.B.: hat der visuelle Reiz A eine intrinsisch höhere Gewichtung als Reiz B und entscheidet sich der Proband richtigerweise für A, so wird eine Belohnung mit höherer Wahrscheinlichkeit ausgegeben; entscheidet er sich für B, entfällt mit höherer Wahrscheinlichkeit die Belohnung. In der zweiten Phase werden die Probanden an weiteren Reiz-Kombinationen getestet, ohne eine Rückmeldung für ihre Entscheidungen zu erhalten.

Die ursprüngliche PST modifizierten wir für diesen Versuch: Statt einer monetären Belohnung wurden gustatorische Stimuli ausgegeben mit dem Ziel, den Effekt von Nahrungsmitteln auf das Lernverhalten zu untersuchen. Anschließend analysierten wir die Ergebnisse mit computergestützten Modellierungsmethoden und kombinierten diese mit den zeitgleich akquirierten Multiband-Sequenz-basierten fMRT-Bildgebungsdaten.

Allen Probanden dieser Studie gelang es, die gustatorischen Stimuli angemessen zu interpretieren und darauf zu reagieren. Für den am höchsten gewichteten visuellen Reiz (A) zeigte sich eine durchschnittliche Genauigkeit von 88%±20.0%; für den am niedrigsten gewichteten Reiz (B) eine Genauigkeit von 73%±18.0%. Diese Ergebnisse stimmen überein mit denen anderer Studien.

Hervorzuheben ist, dass der Erfolg der Aufgabe teils von der individuellen Bewertung der ausgegebenen Belohnung abhing. Verloren Probanden im Laufe der Aufgabe die Motivation zur Belohnung, so zeigten sie eine schlechtere Fähigkeit, höher gewichtete Reize korrekt zu wählen.

Bei der computergestützten Modellierung war das sogenannte asymmetrische Lernmodell überlegen. Bei diesem Lernmodell wird positive Verstärkung und negative Bestrafung unterschiedlich stark gewertet.

Die durchschnittliche Lernrate für positive Verstärkung betrug $\alpha_+ = 0.13 \pm 0.15$, für negative Bestrafung $\alpha_- = 0.08 \pm 0.17$ (Durchschnitt ± Standardabweichung).

Die fMRT-Bildgebungsdaten waren von suboptimaler Qualität. Während wir einige der von uns erwarteten Haupteffekte im visuellen, motorischen und sensorischen Kortex beobachten konnten, fehlte die erwartete neurologische Aktivität im Belohnungssystem – die für unsere wissenschaftliche Frage von zentraler Bedeutung ist.

Zusammenfassend zeigt unsere Studie, dass eine Umsetzung der PST mit gustatorischen Stimuli möglich ist. Um die entsprechende neurologische Aktivität evaluieren zu können, sind jedoch Verbesserungen der fMRT-Einstellungen erforderlich.

Ein optimiertes System könnte in weiteren Studien zu einem besseren Verständnis über neurobiologische Lernprozesse, die zu Adipositas führen beitragen, insbesondere die Rolle von Nahrungsmitteln als besondere Verstärker.

Summary

One of the main challenges for global public health in the modern world is the rising prevalence of obesity. Obtaining a better understanding of the dysregulated feeding behaviour that leads to obesity, by investigating the decision making and learning processes underlying it, could advance our capabilities in battling the obesity epidemic.

Consequently, our aim in this study is to design and validate an experiment that could evaluate these processes.

To do so, we examined ten healthy participants using a modified version of the "probabilistic selection task" (PST) - an established behavioural task for evaluating learning behaviour. In the first phase of this task, the participants are presented with pairs of visual cues and are required to select one cue on each trial. Each cue has a different win probability and participants are tasked with identifying the more rewarding options. For example, if the participants correctly choose the visual cue A, which has a higher win probability, they would likely receive a reward. If the participants choose the less rewarding cue B, the reward would likely be omitted. In the second phase of the task, the participants are tested on their acquired knowledge, without receiving feedback for their choices. We modified the paradigm by delivering gustatory stimuli as a replacement for monetary rewards, to assess the effect of nutritional rewards on the learning behaviour. We subsequently analysed the behavioural results with computational modelling and combined this with imaging data simultaneously acquired with a functional magnetic resonance imaging (fMRI) multiband sequence.

All ten healthy participants in this study succeeded in interpreting and interacting with the gustatory stimuli appropriately and consequently managed to conduct the task as expected. The participants successfully learned the task, presenting an average accuracy of 88%±20.0% for the most rewarding cue (A) and 73%±18.0% for the least rewarding cue (B), resembling performance results from other studies.

Noticeably, the performance on the task was partially affected by the participants' subjective valuation of the reward. Participants whose motivation to drink the reward and liking of its taste decreased during the task presented more difficulties correctly choosing the more rewarding cues.

Computational modelling of the behaviour found that the so-called asymmetric learning model, in which positive and negative reinforcement are differently weighted, best explained the group. For this model, the average learning rate for positive feedback was $\alpha_+ = 0.13 \pm 0.15$, for negative feedback $\alpha_- = 0.08 \pm 0.17$ (average ± standard deviation).

The acquired fMRI data was suboptimal. While we could observe some of the main effects we

expected in the visual, motor and sensory cortices, we did not detect the neurological activity we expected in the reward system, which is central to our scientific question.

Thus, our study shows it is possible to implement the PST with gustatory stimuli. However, to evaluate the corresponding neurological activity, our fMRI configuration requires improvement. An optimised system could be used in further studies to improve our understanding of the neurobiological mechanisms of learning that lead to obesity and elucidate the role of food as a distinctive reinforcer.

2. Introduction

2.1 Obesity and health

2.1.1. The epidemiology of obesity

Obesity is a known major risk factor for various diseases such as cardiovascular disease, type 2 diabetes mellitus, osteoarthritis and some types of cancer.[1] This is a widespread problem and the prevalence of obesity is rising quickly among both adults and children worldwide. The rate of obesity has almost tripled globally between 1975 and 2016 to a mean rate of 39% of all adults being overweight and 13% being obese, with an even higher prevalence in member countries of the Organisation for Economic Co-operation and Development (OECD).[2]

Consequently, obesity is a substantial problem for global public health. In 2017, high body mass index (BMI) has risen to be the fourth leading independent risk factor causing death and disability in the world's population in total and the third highest factor for women. This was surpassed, aside by smoking, only by high systolic blood pressure and high resting plasma glucose, factors which are themselves both components of the metabolic syndrome, for which obesity is a risk factor.[3,4] It is estimated that reducing the prevalence of high BMI has the potential for the greatest reduction of years of lost life among all major risk factors.[4,5]

In addition to the individual health risk, obesity also poses a heavy financial burden on many countries, especially in high-income countries. The ever-increasing rate of obesity and related diseases is projected to constitute 8.4% of health expenditure in average in the OECD until 2050.[6]

Thus, the endeavour to reduce obesity in the world's population is a major goal for the global health community. In order to successfully combat obesity, the health community requires a better understanding of its causes.

2.1.2. Defining obesity

When total energy intake exceeds energy expenditure, the body stores the excess energy as a reserve for future use. For short-term storage, the body stores energy in the form of glycogen, deposited primarily in the liver and skeletal muscle. The typically larger, more long-term reserve is stored mainly in adipose tissue in the form of triglycerides.[7,8] If the excess in energy balance persists over time, adipose tissue proliferates and expands, resulting in increased

body weight.[8] When an individual's weight crosses a predefined threshold, he or she is considered overweight. When crossing an even higher threshold, he or she is then considered obese.[9]

In order to properly investigate obesity, one needs a reliable definition and measurable parameters. Imaging based analyses, such as dual-energy X-ray absorptiometry (DEXA), computed tomography (CT) or magnetic resonance imaging (MRI) allow a precise measurement of the amount and distribution of body fat.[9] However the comparably high cost of imaging methods, their lower availability and, in the former cases, exposure to radiation, make them impractical for many studies.[9]

In place of these more direct measurements, one can attempt using surrogate parameters to indirectly estimate body fat.

One of the most widespread normative parameters for quantifying weight is the BMI, which requires knowing only the individuals' height and weight. For adults, a BMI of $18.5 - 24.9 \frac{kg}{m^2}$ is considered normal, $25 - 29.9 \frac{kg}{m^2}$ is considered overweight and individuals with a BMI over $30 \frac{kg}{m^2}$ are considered obese, categorized further into grades I-III when the BMI is over $30 \frac{kg}{m^2}$, $35 \frac{kg}{m^2}$ or $40 \frac{kg}{m^2}$ respectively.[9] This parameter is useful when investigating large populations and has been shown to correlate with disease and disability. However, this simple parameter provides only a relativity crude assessment of body fat and can often misrepresent the individual's constitution. On one hand, BMI can overestimate overweight in some individuals, such as exceptionally muscular individuals with low body fat percentage.[10] On the other hand, BMI has been found to significantly underestimate body fat percentage in some cohorts, when compared to more precise results from imaging based analyses.[9,11–14]

Alternative proportion parameters such as waist circumference, waist-to-height ratio and waist-to-hip ratio have also been suggested to estimate body fat. These parameters were found to estimate especially intra-abdominal fat (also known as visceral fat) more accurately than BMI and thus correlate better to clinical end-points.[13,14] Another, more technology-based method to measure body fat is a bio-electrical impedance analysis. This method potentially allows an even more precise characterisation of body fat in general and visceral fat in particular, while still avoiding the disadvantages inherit to imaging based analyses.[15–17]

Visceral adipose tissue has a considerably higher endocrine activity than subcutaneous adipose tissue and this is likely the reason why abdominal fat correlates more directly to clinical endpoints. It is thought to be through the secretion and metabolism of various hormones,

adipokines and pro-inflammatory cytokines, that visceral adipose tissue exerts its negative impact on the body and drives disease.[8,18–20] It may therefore also be useful to characterise these mediators more directly. Peripheral levels of adipokines such as leptin, hormones that interact with adipose tissue such as insulin or ghrelin, and metabolic compounds such as fasting glucose, cholesterol and triglyceride can all help in assessing the effect an individual's fat deposits have on his health.[11,21,22] These parameters can also be combined, such as with homoeostatic model assessment of insulin resistance (HOMA-IR) to achieve a more nuanced estimation of the individual's metabolic status.[23,24] Finally, the measurement of clinical parameters such as blood pressure may be useful in gauging the more downstream effects of obesity on the individual's health.[25]

2.2   The causes of obesity

2.2.1.   Obesity and behaviour

In the modern world, where access to energy dense foods is easy and food cues are ubiquitous, rising levels of obesity are commonly attributed to dysregulation of feeding behaviour, i.e. overeating.[26–29] Principally on an individual level, a positive caloric balance can also occur when an individual is maintaining a stable amount of caloric intake, if they reduce caloric expenditure too greatly. However, at least in some populations, excessive caloric intake alone seems to sufficiently explain rising obesity rates, as average caloric expenditure in these populations has remained equal or even risen over the decades. This finding strengthens the importance of overeating as a main factor promoting obesity, regardless of physical activity levels.[28,29] Furthermore, there is some evidence that total energy expenditure in humans is constrained, meaning weight gain through increased intake cannot be entirely impeded by increasing physical activity.[30]

Consequently, it is of particular importance to understand feeding behaviour and overeating - the conscious or unconscious decision-making process that leads to the action of consuming more calories than needed for the body's energy balance. Improving our understanding of the cognitive mechanisms underlying overeating can provide us with valuable insight into how to slow down or even reverse the global trend of rising obesity rates.

2.2.2.   Feeding behaviour and learning

In humans and other higher animals, feeding behaviour is regulated centrally by the brain. Information from the external world, such as the presence of food or anticipation of food

shortage, information from the body, such as status of current energy reserves or expenditure, as well as information from the brain itself, such as subjective memory of the food presented, is collected by the brain, which then performs a decision-making calculation regulating feeding behaviour.[27,31,32]

These neural calculations are not static, but are adaptive and can change with new evidence, attempting to optimize the calculation and take the most appropriate decision for unknown conditions with uncertain results. This adaptation to additional information is understood as the process of learning.[33–36]

Learning itself is not a singular process but takes place in several different forms. For example, learning can be based on positive or negative feedback, be implicit or explicit, goal-directed or habitual. Different types of learning and their degree can depend on the circumstances, but will also vary between individuals in the same circumstances.[37]

In the case of obesity, inappropriate learning is suspected to be the root of dysregulated feeding behaviour. Obese individuals have been previously shown to present impaired learning, especially from negative feedback, when compared to lean age-matched adults. This impaired learning shown not only for food-related behaviour[38,39] but also in non-food-related learning tasks.[38,40,41]

2.2.3. The connection between the brain and obesity

Feeding behaviour is regulated by two nominal brain systems: the homoeostatic system and the non-homoeostatic, or hedonic, system.

The more primitive homoeostatic system functions to maintain and balance the body's energy needs. Signals from this system promote feeding behaviour when energy and nutritional deficits are detected but evoke an anorexic response when those needs are met.

In contrast, the hedonic system allows a person or animal to attribute a subjective positive or negative value to food, beyond its nutritional value. It is within this system that learning mechanisms play a central role in defining behaviour.

The value attribution performed through the hedonic system can override the signals of the homoeostatic system in a top-down fashion and can promote the consumption of nutrition beyond the body's energetic needs, which, if persisted over a sufficient period of time, results in obesity. Conversely, the hedonic signalling can also drive the person or animal to avoid certain substances, such as unpalatable food, even while experiencing strong hunger.

Additionally, an interaction of the two systems exists also in the opposite direction. The homeostatic system modulates hedonic signals in a bottom-up fashion and can, for example, raise the perceived value of a food when hungry or decrease it when full.[27,31,32]
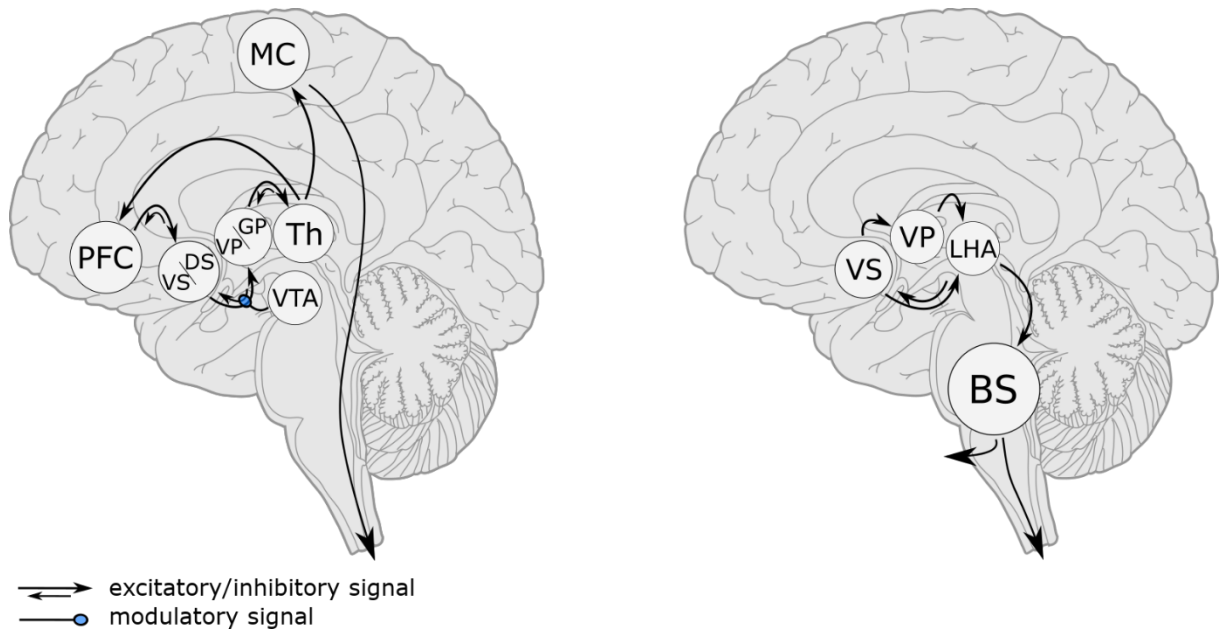
Figure 1: Schematic overview of the main pathways of the hedonic system and hedonic-homoeostatic relationship. The hedonic system is illustrated on the left and the interaction with the homoeostatic system on the right. PFC=prefrontal cortex; VS=ventral striatum; DS=dorsal striatum; VP=ventral pallidum; GP=globus pallidus; Th=thalamus; MC=motor cortex; VTA=ventral tegmental area; LHA=lateral hypothalamic area; BS=brain stem.

2.2.4. The hedonic system

In order to better understand how the hedonic system attributes positive value to stimuli, i.e., perceives them as rewards or reward predictors, and how learning processes influence this mechanism, it is important to understand the structure of this system. The hedonic system, sometimes known as the reward system, is activated not only in food related scenarios, but coordinates learning and decision making also from other primary reinforcers as well as from secondary reinforcers, enabling learning also in more abstract contexts.[42]

While many different brain regions and the interactions between them are involved in the different aspects of reward evaluation and response, the core of the hedonic system is considered to be the cortico-ventral basal ganglia circuit. Put simply, information is processed in the prefrontal cortex and passed to the ventral striatum, from there to the ventral pallidum, then to the thalamus and back to the cortex, with the signals being modified by dopaminergic projections from the midbrain, especially at the ventral striatum. This is however not a one way loop, but rather a complex interconnected network with many signals travelling directly and reciprocally between the mentioned structures.[42]

In this circuit, the forebrain processes input from sensory and internal sources to formulate multiple possible plans, actions and orchestrate behaviour. Here, especially the orbitofrontal cortex, and several regions of the anterior cingulate cortex: the dorsal anterior cingulate cortex,

17

ventromedial prefrontal cortex and the subgenaul area play a major role in reward valuation. These signals are sent to the ventral striatum (which includes the nucleus accumbens and the olfactory tubercle) to be evaluated and perform critical computations on reward prediction and reward prediction errors. This computation allows the selection and acquisition of specific behaviours and is heavily modulated by dopamine, originating mainly from projections rising from the ventral tegmental area in the midbrain in what is known as the mesolimbic pathway. The chosen behaviour is then further regulated by the ventral pallidum and then passed on to the thalamus (especially to its mediodorsal nucleus), which projects back to the cortex which then executes the learned behaviour.[42]

The interaction of this circuit with the basal ganglia motor loop allows the execution of concrete motor actions needed to perform the behaviour. The basal ganglia motor loop has a similar hierarchy to the cortico-ventral basal ganglia and is composed of the motor regions of the cortex, the dorsal striatum, the pallidum and the thalamus and is modified by dopamine projections mainly from the substantia nigra pars compacta.[42]

Interaction of the two circuits also occurs in the midbrain itself, as the ventral tegmental area and the substantia nigra pars compacta do not have clearly defined borders, but are more contiguous and share some overlapping projection fields. This may explain why the motor loop also plays a direct role in reward and reinforcement learning, partially overlapping in function with the cortico-ventral basal ganglia circuit.[43]

Dopamine and reward prediction errors

A major component of the acquisition of new behaviour through the hedonic system is its reaction to reward prediction error (PE), mediated mainly by the dopaminergic neurons of the midbrain.

These neurons constantly release a tonic amount of dopamine to the synapses of their respective projections in the striatum. When receiving input about an unexpected stimulus, these neurons react with a phasic change in the dopamine levels they secret. When unexpected reward arrives the levels rise as a burst, which activates the "Go" pathway over D1-like receptors. After an unexpected omission of reward there is a sharp dip in this neurotransmitter's levels in the synapses, which activates the "NoGo" pathway over D2-like receptors. The magnitude of this phasic change has been shown to be linear in size with respect to the level of "unexpectedness" of the reinforcer, i.e., with the PE amplitude. This means, that as learning progresses and the reward (or the omission thereof) becomes more expected, the magnitude of phasic dopamine change reduces. If the reward or its omission is completely expected, no change in dopamine level occurs. In parallel, as the reward becomes

more expected, this dynamic of the PE gradually shifts to the context associated with the reward. For example, the action taken or the cue that was perceived just before the salient reward was received would now assume the PE dynamic for the release of dopamine. This process can also repeat and chain more distant actions or stimuli to the primary reward.[44–48]

The basal ganglia govern this associative learning in several forms. Firstly, the association between an action and its outcome, which is central for goal-directed behaviour. Secondly, between a stimulus and the person's response, an association central in habit formation. The basal ganglia can also shift an association and turn a previously goal-directed action into a more habit-based behaviour.[48,49]

In addition to phasic changes, the tonic level of dopamine secretion also affects the pattern of learning. Tonic level and rewards seem to have a linear correlation, with higher tonic levels of dopamine secretion promoting learning from positive reinforcement. The relationship with negative reinforcement has an inverted U-shape form, so both too low and too high tonic levels of dopamine impede learning from reward omission.[50,51]

It should be noted that the dopamine neurons of the ventral tegmental area and the substantia nigra pars compacta present a rather homogeneous electrical activity, likely because they are interconnected with gap junctions, which allow for a faster electrical coupling than chemical synapses .[52–55] This makes it plausible to model their output as a single scalar value, such as the scalar prediction error described below in section "Computational modelling: reinforcement learning".[55,56]

In addition, while the dopaminergic neurons of the midbrain are especially reactive to unexpected reward or it's omission, they also react to physical salience, risk and weakly also to punishment.[57]

Food and dopamine

Food intake has a dual effect on dopamine signals - first through taste and then through nutritional digestion. Both these responses elicit dopamine release in the brain but do so over segregated pathways and with different results. The immediate orosensory response releases dopamine in key regions for reward valuation and motivation, reflecting the desire to eat and assigning value to the food as reward. The delayed post-ingestive response induces dopamine release mainly in the inhibitory pathway of the putamen, suggesting the intake of nutrition may dampen the desire to feed in a response taking place about 15 minutes after the orosensory

input of the same food.[58] This dynamic gives food rewards another characteristic that distinguishes it from other reinforcers.

2.2.5.   The homoeostatic system

Abstract rewards such as money or psychostimulants such as cocaine and methamphetamine play out their effects mainly in the hedonic system. In contrast, food rewards also affect the more basal homeostatic system, which in turn affects the hedonic system in a bottom-up fashion. For example, food deprivation has been shown to increase hedonic response to the consumption of food and even enhances the response of the hedonic system to non-food rewards.[59–61]

One of the main neuronal circuits managing this interaction between the hedonic and homoeostatic systems is the neuronal loop between the nucleus accumbens and the lateral hypothalamic area.

The lateral hypothalamic area is the primary orexigenic nucleus in the brain and receives feeding-related homeostatic and circadian signals. The lateral hypothalamic area can send excitatory signals to the nucleus accumbens, which itself can send disinhibiting signals back to the lateral hypothalamic area, over both a direct and an indirect pathway (over the ventral pallidum). The lateral hypothalamic area can then influence feeding behaviour by sending signals to brain stem regions involved in feeding-related motor control as well as sympathetic and parasympathetic nuclei controlling feeding-related functions such as salivation and gastric acid secretion. This reciprocity between centres of both the hedonic and homoeostatic systems forms a self-reinforcing connection between the nutritional and motivational value of food, strengthening feeding behaviour.[31,62] It is therefore possible that dysregulation of this interaction also influences obesity.

2.3   How to research learning behaviour

After establishing that dysregulated learning can lead to obesity and describing which brain systems play a central role in this mechanism, it should be discussed which methods and models can be implemented to further investigate these mechanisms in a human population.

2.3.1.   Modelling learning behaviour

In order to properly interpret data from an experiment and infer upon the learning progress of the participants, a good model of the learning process is needed.

In modern learning theory, there are three established theories describing the acquisition and adaptation of new decision-making-related behaviour: classical Pavlovian conditioning (the association of a natural unconditioned stimulus and a learned conditioned stimulus), operant conditioning (the individual's or animal's behaviour influences the stimulus and is negatively or positively reinforced by the outcome) and social learning through observation and imitation. In addition, there are crossovers between these types of learning, such as general and specific Pavlovian-instrumental transfer.[63,64]

Another important hypothesis in learning behaviour is the dual-system theory of instrumental conditioning. According to this theory, an individual's decision-making process is determined through a balance between two separate types of instrumental learning: goal-directed and habitual learning.[49,65] In obesity, especially habitual learning has been implicated in the determination of behaviour. Feeding behaviour learned this way can override suppressive homoeostatic signals and, unlike goal-directed decision making, ignores the possible long term negative ramifications of overeating.[39,66]

Measuring learning

In order to evaluate if and to what extent learning has been accomplished, we need to be able to measure it. The success of learning is however not a universally defined parameter and may be depended on the behaviour being learned and the observer's expectation of the wanted result.[67,68]

In the context of operant conditioning and instrumental tasks, for tasks containing binary choices with one choice being considered correct (e.g., rewarding or avoiding punishment) and the other incorrect, one can divide the experimental paradigm into a learning phase, during which new knowledge is acquired and a testing phase, to measure the acquisition.

If the learning process was successful, the participant should be able to perform in the testing phase with an adequate success rate (e.g. better than a predefined threshold).[69]

If the task contains multiple possible choices with an intrinsic hierarchy on an ordinal, interval or ratio scale, one could evaluate the participant's ability to successfully order the choices on their place in the hierarchy.

In addition to measuring learning by observing its final state as measured in the testing phase, one could also attempt to evaluate the learning process with a more mechanistic approach by analysing the learning phase itself. By fitting the participant's conduct during the acquisition phase to a learning curve, one could try assessing the speed and efficiency of his or her learning process.[67,68,70]

However, as the participant's strategy at each step of the task is hidden from the observer, such a learning curve can only be approximated, and a robust model is required to derive a sufficiently accurate result.

This type of model, which approximates the participant's behaviour based on an assumption of its underlying mechanism, is known as a generative model and often needs to be calculated in a computational modelling framework.

Importantly, for the generative model to be useful and give a good estimation of the true computation performed by the brain, it must also be generalisable to new participants or tests.[68,70–72]

The value of such model-derived learning curves and associated parameters comes into play, especially when validating them with external measures. Correlating the learning curve parameters with imaging data greatly enhances the interpretation of imaging signals by enabling us to localize and quantify activity in the brain related to the concrete learning process.[72,73]

Computational modelling: reinforcement learning

While each of the models of learning theory tries to explain how new behaviour is acquired in somewhat a different way, and they may truly describe biologically separate mechanisms, they have all been mathematically modelled quite reliably by implementing the concept of (reward) PE in appropriate algorithms.[71,73]

According to the concept of PE, it is not the intrinsic value of the reward itself that drives learning, but the discrepancy between the actor's expectation and the actual outcome. A surprising reward is much more salient than a predictable one, and an unexpected punishment is much more penalizing than one which was seen coming. Stimuli that are entirely expected by the actor have no effect on the learning process for that behaviour.[45,46,71,73]

One of the most successful algorithms for generative models of learning behaviour is that of reinforcement learning, when based on PE. This class of algorithms, originating from computer science and artificial intelligence, have been used quite successfully in the fields of neuroscience and behavioural science. Interpreting data in a reinforcement learning framework has significantly improved our ability to investigate the function of specific brain structures during learning.[73,74]

Especially model-free reinforcement learning has been shown to correspond well with dopamine-based and habit learning, as opposed to model-based reinforcement learning which corresponds more with goal-directed learning.[74]
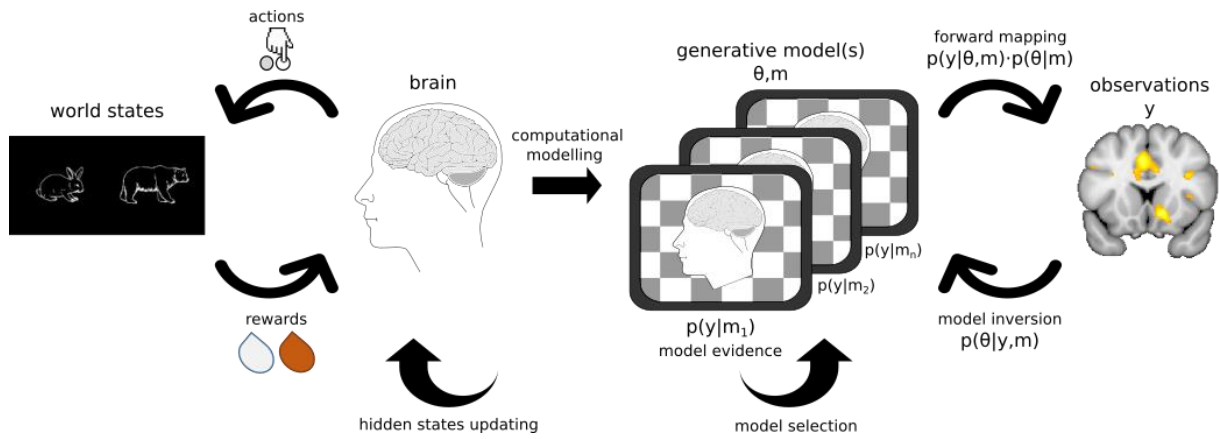
Figure 2: The Interaction between the brain and the world according to the Bayesian brain hypothesis mirroring the relationship between generative models and experimental observations in the context of Bayesian statistics

$$\underbrace{p(\theta|y,m)}_{posterior} = \frac{\overbrace{p(y|\theta,m)}^{likelihood} \cdot \overbrace{p(\theta|m)}^{prior}}{\underbrace{p(y|m)}_{model\ evidence}}$$

Equation 1: Bayes' theorem. p()=probability; θ=model parameters; y=observation data; m=model.

Computational modelling: approach learning vs. avoidance learning

As discussed in section 2.2.4, individuals likely differ in their learning dynamic from positive feedback, also known as approach learning, and their learning dynamic from negative feedback, also known as avoidance learning[40,48,50,51,57,69] Moreover, learning from negative feedback would likely differ if the stimulus is an explicit punishment or an omission of reward.[57]

When designing a reinforcement learning-based generative model, one can model each different learning dynamic as a separate learning rate. However, adding more parameters to a model can potentially allow more overfitting and thus weaken the model evidence. This creates a trade-off between generalisability and precision of the model for real-world data.[72]

Computational modelling: model inversion

After gathering data and constructing a generative model to interpret that data, we need to fit the data to the model and assess its accuracy in a process known as model inversion. Additionally, we can set up alternative generative models and test which one best explains the observed data.[72]

Different statistical techniques exist to perform model inversion of reinforcement learning models.

In the past, maximum likelihood estimation was a popular technique for fitting data, as it is often easy to compute and relatively straightforward to interpret. However, maximum likelihood estimation gives a point estimate without a representation of uncertainty and is susceptible to overfitting.[72]

Advances in computational capabilities have promoted Bayesian estimation as an often-superior approach. The Bayesian estimator is based on Bayesian statistics and therefore requires calculating full posterior densities, which can in turn be approximated with various algorithms.

Two main algorithms groups are currently being used in the research of learning behaviour: Monte-Carlo sampling methods and variational Bayes. Monte Carlo sampling methods such as Markov chain Monte Carlo are based on repeated random sampling. These methods tend to be slow but principally (according to the law of large numbers) converge to the real posterior density and thus can generate relatively exact estimations of the posterior. In contrast, variational Bayes methods tend to be faster to compute while still generating a good approximation of the posterior densities. An additional advantage is their capability to directly measure model evidence, useful for comparing models.[72,75,76]

Interestingly, according to a popular theory known as "Bayesian brain hypothesis", the brain itself may be inferring on the outside world (or on the inner states of the body) by performing Bayesian estimations. Correspondingly, the brain is assumed to have a model or an expectation of the world (or of the body) used as the prior. Once the brain receives sensory input, it performs model inversion and calculates a posterior probability. If the prior and posterior diverge, the brain strives then to resolve this prediction error by mobilizing the body to change sensory input, by updating its prior beliefs, i.e. learning, or a mixture of both solutions.[36] This concept is schematically shown in figure 2.

In this framework, the dysregulated learning driving obesity could be understood as a systematically inadequate resolution of prediction errors.[40,41]

Computational modelling: model selection

As all models are simplifications of the real world, we can take an extra step in improving the accuracy of our chosen model by suggesting alternatives and calculating which model best explains our observations. This comparison can be performed on an individual level, assessing which model best explains the data obtained from an individual participant. However, for an experiment assessing a group of participants or comparing between groups, a calculation must be carried out to select one or more models which are most appropriate across all participants examined.

Assuming one model best explains all participants is known as a fixed-effects assumption. This approach is very susceptible to outliers. Allowing the calculation to fit different models to different participants is known as a random-effects model selection approach.[77]

As with model inversion, one can perform the computation of log model evidence needed for the above-mentioned methods with variational Bayes using the Laplace approximation or by using a sampling method such as Monte Carlo sampling methods. Due to variational Bayes' superior speed, it is currently the more commonly used approach in the field of behavioural neuroscience.[72]

However, it is important to note that while these comparisons can find the best model among the suggested ones, they are not capable of assessing how the models fit the real world. The best model from the comparison still needs to be assessed in an independent manner for its actual accuracy.[77]

2.3.2. A behavioural task to assess reward-based learning

The behavioural paradigm and the related components of an experiment must be carefully constructed, depending on the aspects one wishes to investigate. For researching dysregulated learning in feeding behaviour, it would be advantageous if the paradigm gathers data that would be appropriate to model as model-free learning and could ideally differentiate between approach learning and aversion learning. Although a Pavlovian task can also be used, an instrumental task would likely be more engaging and hence improve participant participation. Finally, one could design a completely novel task or use a previously established task. Reusing or modifying an established task has the advantage of building on existing literature and simplifies the comparison of results, especially if the task has already been examined with the additional investigatory tools (e.g., imaging methods) one would also like to use.

Probabilistic selection task

The probabilistic selection task (PST) is an established paradigm for evaluating reinforcement learning, which is also capable of differentiating between learning from positive and negative reinforcement.[67] The original study examined a cohort of patients with Parkinson's disease, comparing different medication statuses as well as against a control group. It has since been expanded to study other populations, as well as adding imaging data with functional magnetic resonance imaging (fMRI) [78–82] and electroencephalography (EEG).[78]

The original PST used a performance-based criterion to determine how the individual participant advances through the phases of the task.[69] However, as an adaptation to the constraints of fMRI analysis, the studies acquiring fMRI data choose to set a fixed amount of trials for each phase of the task.[78–82]

Previous studies implementing the PST to study reinforcement learning could show that variation of dopamine responses in the brain can substantially affect learning. This was shown in participants with abnormal dopamine levels resulting either from pharmacological manipulation[81], medical conditions such as parkinsonism[67,79,83] or even genetic polymorphisms[50,84], indicating this task is appropriate for investigating dopamine-based learning.

Choosing the visual cue set

Choosing the visual cue set for a paradigm is not a trivial task, as it can affect the results of both behavioural performance and imaging data. Choosing the cue set must therefore be done with careful consideration. On a behavioural level, certain cues may affect the participant's prior valuation of the cue, e.g. the participant's associations or affinity towards the chosen objects or colours can change his or her expectations, and therefore performance, when choosing between the pairs.[85] Stimulus discriminability can influence the difficulty level of learning and inconsistency between cues within the same task can create an uneven playing field, thus biasing learning. This has been specifically shown to happen with the set of hiragana symbols used in the original PST.[86]

Differences between cues in visual factors like intensity, brightness, and abstractness can all affect the signals acquired by the fMRI, potentially introducing unwanted variability. It is therefore important to minimize such differences between the presented cues as much as possible. As a result, a popular choice for fMRI tasks is using black and white images with similar properties.[85]

The original PST and many following studies used a set of hiragana symbols, as mentioned above.[69,78,79,83,84,86–88] However, some later studies [81], including some which have also studied obesity,[80] have used a cue set containing drawings of more easily identifiable animals or other simple symbols.[89]

Choosing the reward modality

For either Pavlovian or instrumental learning tasks, reinforcers are needed to induce learning. The reinforcer should have a value, either positive or negative in nature, before the beginning of the experiment. In a Pavlovian context, the reinforcer would be called the unconditioned stimulus and in an operant conditioning context, the reinforcer is a reward or punishment. By process of conditioning, a previously neutral stimulus (the Pavlovian conditioned stimulus, or the operant conditioned response) is then associated with the value of reinforcer. This can be an auditory signal, a visual cue, an odour, etc.[37,64]

The reinforcer can belong to one of two types: primary or secondary. Primary reinforcers are of inherent biological value to the body, instigate reflexes in the body and the more primitive parts of the brain and do not need to be actively learned by more advanced areas of the brain. Examples of primary reinforcers are food, drink, warmth, sex, sleep or relief from pain.[37,64,90] Secondary reinforcers are stimuli that have been associated with a primary reinforcer and thus gain value, but this association can principally also be unlearned. In an experimental setting, this value should be present before and outside the context of the experiment. For example, sight or smell of food can be used to proxy the reinforcement value of actual food.[37,64,90,91] Money has also found widespread use as the reward in human experiments, because of its close association as a token for more primary reinforcers.[37,90,92] Money also has some practical advantages, as it likely keeps a relatively steady value as a reinforcer during the experiment and simply informing the participant about his win or loss is likely enough to capitalize on its value, without needing to physically give to or take away money from the participant on a trial-to-trial basis. This reward modality has been used in many studies investigating behaviour and learning, also in the context of obesity, as the condition is assumed to be a result of more general inadequate learning processes.[38,40,41,80]

However, while money may be closely associated with reward in general, it is likely that primary and secondary reinforcers still differ in the way they are evaluated by the brain. Food also triggers other brain systems, which can, in turn, interact with the reward system.[58,90] Food rewards may therefore have a specific effect on the learning process, which may differ between obese and normal-weight adults in a greater manner than non-food reward-based learning. This makes it worthwhile investigating learning in obese adults while using gustatory stimuli.

Optimizing the task paradigm for imaging acquisition

As discussed below, imaging data is noisy and certain adaptations of the task paradigm can be implemented a priori to improve the signal-to-noise ratio (SNR).

For fMRI data, temporal resolution can be improved by "jittering" the interstimulus intervals, i.e., using (preferably varying) delays between stimulus onset and sampling of brain images with delays that are not multiples of the repetition time (TR) of the MRI sequence. This enables the acquisition of responses to stimuli at different time points, improving the reconstruction of brain activity from the observed hemodynamic response.[93,94]

Another method is the inclusion of "null events", events that are similar to the intertrial interval and should not be detectable by the participants. These events aim to create an approximation of a baseline neuronal activity that can assist in the analysis of main effects.[94]

2.3.3. Imaging data acquisition

Different techniques exist to investigate functional activity in the brain. While in animal research some invasive tools exist, human research is usually constrained to non-invasive or minimal-invasive methods.

Popular methods used for such studies include EEG, magnetoencephalography (MEG), functional near-infrared spectroscopy (functional near-infrared spectroscopy), positron emission tomography (PET), single-positron emission computed tomography (SPECT) and fMRI.

These methods vary in the directness of the measure of neuronal activity, temporal resolution, spatial resolution, mobility as well as related costs and accessibility.

EEG which measures electrical potentials over the brain and MEG which measures the neuromagnetic field, both have the most direct measure for the actual electrical and therefore functional activity of neurons with excellent temporal resolution (under $1\,ms$). However, EEG has the worst spatial resolution of the above-mentioned methods (ca. $10\,mm$) and can only indicate the broad area from which the activity arises but is the cheapest and most accessible method. MEG's spatial resolution is moderate (ca. $5\,mm$) but this can be improved by simultaneous use with EEG. However, the relatively high cost of this dedicated equipment minimises the extent of its use. fNIRS also has a slightly higher spatial resolution than EEG ($5 - 10\,mm$) and a similar temporal resolution but only measures the delayed hemodynamic response and not the direct electrical activity. EEG, MEG and fNIRS are all limited in measurement-depth mostly to measuring in cortical regions. In contrast, PET, SPECT and fMRI can also accurately examine deeper regions of the brain. fNIRS, PET, SPECT and fMRI

all measure neuronal activity indirectly by measuring the haemodynamic response, while the radioactive tracer used in PET and SPECT can also be used to measure more differentiated aspects of the brain, such as the distribution of dopamine receptors captured with a dopamine transporter scan (DaT-Scan). However, these techniques also mean exposing the participants to radiation. The spatial resolution of PET and SPECT is similar to MEG (ca. $4\,mm$ and $6\,mm$, respectively) but the temporal resolution is much poorer ($1-2\,min$ and $5-9\,min$, respectively) with PET also being the costliest of all the above-mentioned methods. Finally, fMRI shows the best spatial resolution (ca. $2\,mm$) of all methods, while showing a reasonable temporal resolution ($4-5\,s$, primarily due to the delay of the hemodynamic response). It is also possible to combine fMRI with EEG or fNIRS in an attempt to combine the respective advantages in spatial and temporal resolution. While fMRI is relatively costly and an indirect measurement for neuronal activity, this method presents a good balance between resolutions, safety and cost, which has promoted its popularity in the field.[95–97]

fMRI and BOLD

MRI is an imaging technique, that uses a strong static magnetic field, gradient magnetic fields and radiofrequency pulses to create images that can display the insides of the human body while differentiating various tissue types.

Paramagnetic materials, i.e., atoms with unpaired electrons, most abundantly found in the body as protons in water molecules, align themselves mostly parallel but also anti-parallel to external magnetic fields and precess with a Larmor precession frequency dependent on the field strength. Using additional gradient magnetic fields, one can force the protons to precess in various frequencies according to their location, thus encoding their position. Such atoms can be tilted out of alignment by applying energy - in the form of radio waves, which also makes them precess together (in-phase) in the transversal plane to the main magnetic field. The in-phase precession also emits a signal in radio frequency which can be picked up by a receiver coil.

Once the radio frequency is turned off, the atoms start realigning themselves to the main magnetic field with a rate defined by the time constant T1. At the same time, they also begin de-phasing at a rate defined by the constant T2, which also decays the emitted signal. However, this dephasing is susceptible to inhomogeneities in the magnetic field and the observed decay rate constant, without accounting for inhomogeneities, is called T2*.

Put together, these properties can be combined and manipulated to create a map of signals from protons and other paramagnetic atoms, differentiated by location and quantity, which are then reconstructed into an image.[98]

Functional imaging with MRI was made possible by combining two further discoveries. Firstly, changes in blood oxygenation from the diamagnetic oxyhaemoglobin to paramagnetic deoxyhaemoglobin, give a measurable signal, now known as blood-oxygen-level-dependent contrast.[99] Secondly, changes in local neural activity are coupled with a hemodynamic response. Increased neuronal activity, which requires more glucose and oxygen, is responded to by increasing arterial blood flow. This delivers more oxyhaemoglobin and, as the neurons show an only slightly higher oxidative metabolism, results in a higher ratio of oxyhaemoglobin to deoxyhaemoglobin, a decrease in the deoxyhaemoglobin induced T2* shortening effect and thus a stronger BOLD signal.[100–102]

The most common sequences used to detect BOLD signals are echo-planar imaging (EPI) sequences. Their design allows acquiring images very quickly which enables the temporal resolution needed for functional imaging, but their reliance on the artefact prone T2* weighting lowers the SNR. This requires efficient preprocessing of the images to remove noise, before performing any statistical analysis.[103]

Simultaneous Multislice Imaging

Simultaneous multislice imaging techniques, also known as multiband sequences, are an advancement in imaging techniques. These protocols can be used to acquire multiple slices at once with a reduced TR, enabling a higher temporal resolution while showing minimal to no reduction of SNR - all properties which are especially useful in fMRI.[104]

The advantages of this technique are especially pronounced in resting-state fMRI and when analysing task-based fMRI with Multi-Voxel Pattern analysis. However, the benefits of this technique are not as definite when analysing task-based data with a general linear model (GLM), compared to more classical echo-planar imaging protocols.[105] The effects on SNR seem also to be dependent on the region in the brain, and the exact protocol parameters should therefore be carefully weighed according to the goals of the task paradigm and the specific brain structures under investigation.[106]

Analysis of fMRI data

In order to interpret and draw conclusions from the fMRI data, one must first preprocess the raw data. This includes correcting for noise and artefacts which can arise from head motion, signal drifts, slice timing offset, spatial distortions, etc., as well as preparing individual data-sets for group-wise analysis by normalizing and smoothing the brain structure.[107]

Once the preprocessing is complete, statistical analysis can then be performed on the corrected data sets. For task-based fMRI, the most common analysis is using a GLM. The

results of the GLM can then be used as parameter estimates for statistical parametric mapping (SPM), which is a mathematical technique (and matching software package) that enables the construction and assessment of statistical maps, showing activation levels across different regions of the brain under different conditions of the task paradigm. These maps can then be used to infer on the study hypotheses.[103,108]

Reinforcement learning in fMRI

Although fMRI cannot by itself directly quantify levels of dopamine release, change in hemodynamic activity in brain structures known to be part of the dopaminergic pathways indicate a respective change in dopamine levels. Studies observing activity in these structures during learning tasks strengthened the findings of other methods in making the connection between dopamine and learning behaviour.

Multiple fMRI studies have shown increased activity in response to reward PE in the striatum, particularly in the ventral putamen and caudate, especially in the left hemisphere, as well as increased activity in the left frontal operculum. Several studies have also shown increased activity in the thalamus.[92]

In addition to PE response, another parameter of interest in reward-based learning behaviour is the expected value or state-action value (SAV) of the cue signalling reward. This value seems to be processed by the ventromedial prefrontal cortex, and especially in the subgenual cingulate cortex.[92]

It is important to note, that while some of these effects are found in all instrumental learning tasks regardless of reward modality, some modalities have additional specific effects. PE from liquid rewards, which were designed to quench thirst, showed additional clusters in the lateral putamen and amygdala.[92] Food rewards showed additional clusters in the anterior insulae but showed less of a response in the amygdala as elicited from erotic stimuli, a different type of primary reinforcer.[90]

In addition to reward modality and experiment design, the computational model and the specific parameters calculated for the model can also affect the fMRI analysis. For example, choosing between individually fitted learning rates and a fixed value for the entire group can result in different structures of the reward system showing activity in response to PE.[92]

Consequently, the methods used in analysing task results must be chosen after careful consideration. This includes not only the fitting of the learning rate but also choosing the number of learning rates (a single learning rate or separate ones for different stimulus types), fitting of additional parameters (e.g., temperature) and the setting of priors.

2.4   The neurobiological mechanisms of learning in obesity

In summation, we could gain valuable information about the learning process which is influencing feeding behaviour and driving obesity by studying its underlying neurobiological mechanisms. This can be achieved by coupling an appropriate task paradigm such as the PST with an imaging technique such as fMRI. By comparing individuals with such a method based on obesity-related parameters, we may be able to draw useful conclusions on learning patterns that promote obesity and gain insight into how these patterns are formed.

A growing number of studies suggest that obese and obesity-predisposed individuals performing on reinforcement learning tasks, including the PST, encounter difficulties specifically in learning from negative feedback, but can generally learn appropriately from positive rewards.[38,40,41,80] This raises the question if this behavioural difference also originates from a variance in dopamine responses compared to non-obese individuals, and would suggest especially the D2R-mediated NoGo Pathway is affected.

Imaging studies could indeed show a link between obesity and reduced D2-receptors.[109] There is also some genetic evidence for this link, such as with the fat mass and obesity-associated gene (FTO), which is associated with a predisposition towards obesity, with impaired learning behaviour from negative feedback and with altered D2-dependent midbrain responses.[80]

Most studies investigating reinforcement learning in obesity with fMRI used monetary rewards. One of these studies observed a diverging reaction to monetary loss in the medial prefrontal cortex, with increased activity in obese adults and a decrease in lean adults. This study observed no difference in the dopamine-dependent PE response between the groups.[40]

fMRI studies using food as the reward modality showed obese participants had an increased gustatory response towards cues signalling food reward, but a decreased response in the striatum to the reward itself. This dynamic also predicted later weight gain.[110,111] Variability of response to food reward in the nucleus accumbens was also found to correlate with weight gain.[112] A decrease of activity in the caudate nucleus in response to food reward was also found to correlate with BMI.[113]

To our knowledge, only a few studies have combined all the elements mentioned above and investigated the learning processes predisposing to obesity with fMRI in a reinforcement learning framework using gustatory rewards. None of these studies used paradigms that can explicitly examine avoidance learning, which is of particular interest in obesity

2.5 Aims and Objectives

The research question for this scientific program was to increase our understanding of the neurobiological mechanisms of learning, especially in the context of nutrition as a primary reinforcer.

The aim of this study was to develop a method for the investigation of the above mentioned. This was chosen to be an implementation of the PST, using nutritional liquid responses and simultaneously acquiring fMRI data.

In order to achieve our study aims, following objectives were defined.

1. Write a computer program which runs a customized PST paradigm, which can interact with participants who are laying in an MRI machine. The program should receive the participant's input and activate the gustometer in response.
2. Customize a setup for a gustometer - a mechanical system that can selectively deliver gustatory stimuli.
3. Validate the timing of all steps of the paradigm as well as the synchronization of the paradigm with the MRI.
4. Analyse the acquired behavioural data to determine if the participants successfully learned during the paradigm. This includes investigating the learning curves, response speeds and computational modelling of the data and comparing these to results from previous studies which implemented the PST with other types of stimuli.
5. Evaluate the acquired imaging data to assess the capability of the multiband sequence to show main effect activation in expected systems, i.e., the reward system, visual cortex, motor cortex, somatosensory and gustatory cortex.
6. If main effect acquisition is successful, analyse the imaging data for associations between behaviour and neural activity, based on parameters derived from the computational modelling.
7. Assess if imaging data from the fMRI has a significant added value to data from the behavioural task when studying and comparing reinforcement learning behaviour between participants in the context of gustatory based rewards.
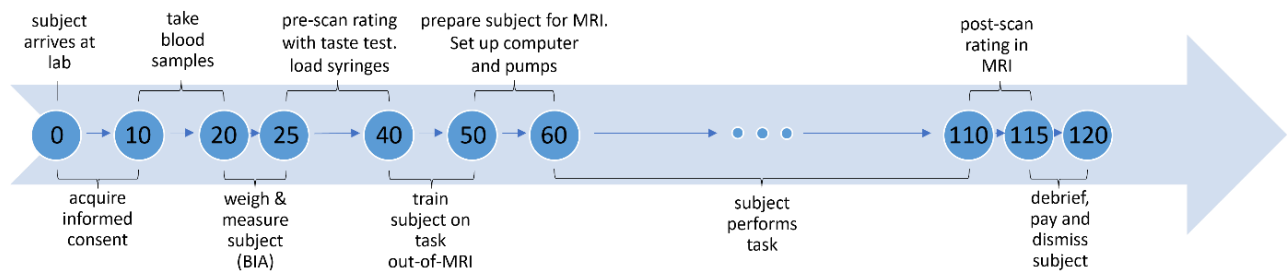
3.   Materials and Methods

Figure 3: Participant examination workflow. Time from arrival is denoted in minutes.

3.1   Participants

10 healthy male volunteers were recruited for the study.

The participants were recruited via the Max Planck Institute for Metabolism Research. We excluded female participants to avoid the effects of the menstrual cycle on the dopaminergic system and reward-related neural function.[114–117] In addition, potential volunteers were screened for age (18-40), having normal BMI, being non-smokers and being free from neurological, psychiatric, metabolic or gastrointestinal diseases including lactose intolerance as well as any other severe chronic or acute illnesses. Finally, participants were asked to express a general liking for milkshakes, as they were expected to perceive the flavoured fluid as a reward.

The workflow for the examination of an individual participant is schematically shown in figure 3.

All participants provided informed consent prior to participating in the study, which had been approved by the local ethics committee of the Medical Faculty of the University of Cologne (Cologne, Germany).

3.1.1.   Metabolic parameters - blood tests and BIA

Participants were instructed to fast for a minimum of six hours before arriving at the laboratory but at the same time to drink a sufficient amount of water as to not arrive thirsty for the examination.

Each participant was weighed and measured, and their respective BMI was calculated. An automatic bio-electrical impedance analysis was performed with a medical body composition analyzer (Seca GmbH, Hamburg, Germany) to assess visceral fat mass.[15–17]

Blood samples were taken and analysed for glucose and insulin levels. From these values the HOMA-IR[23,24] was calculated.

3.1.2. Gustatory stimuli and taste test

As we wished to provide the participants with gustatory stimuli while they are lying supine in the MRI, we opted for using liquids rather than solid food, as they are easier to quickly deliver in small amounts to the participants during the examination. We chose to use milkshakes as the positive feedback instead of juice or other sugary drinks, as their composition from fat in addition to carbohydrates was shown to have a stronger effect with increased reward response when compared to purely carbohydrate-based drinks, even independent of the participants' subjective liking of the taste.[118]

In order to make the negative feedback more comparable to the positive feedback in non-reward aspects and thus reduce unwanted variability in the fMRI analysis, we choose to deliver a neutral-tasting fluid as the negative feedback. This would require the participant to sense a fluid and swallow it, thus mimicking the sensory and motoric brain activity observed when receiving a positive outcome, while the lack of taste and caloric value would represent a lack of reward.

In order to create a setting where the milkshake fluid would indeed be perceived as a positive reward and the tasteless solution as a lack of reward, taste tests were carried out on the scan day shortly before performing the experiment. First, participants were given samples of four different crystalloid solutions and were asked to choose the one most "neutral-tasting" by pipetting a small amount of the fluid into their mouths, while reclining their heads backwards to simulate the application of the stimuli in the MRI machine.

Similarly, participants were given samples from four different flavours of milkshake and were additionally asked to fill an on-screen questionnaire which included the following questions: firstly, general levels of hunger, satiety and thirst, quantified with a visual analogue scale (VAS) and subsequently for each flavour their perception of its sweetness and intensity, quantified with a general labelled magnitude scale[119,120], their liking of the taste, their perception of the liquid's fattiness, creaminess and oiliness and finally how much they would want to drink the milkshake, all of which were quantified with a visual analogue scale.

Subsequently, the milkshake flavour with the highest "liking" and "wanting" scores was chosen for the experiment.

Participants repeated the questionnaire directly after performing the experiment, answering for the flavour they chose for the experiment.

See appendix 7.4.2 for the recipes used.

Table 1: Participant characteristics. Data are given as mean ± standard deviation.

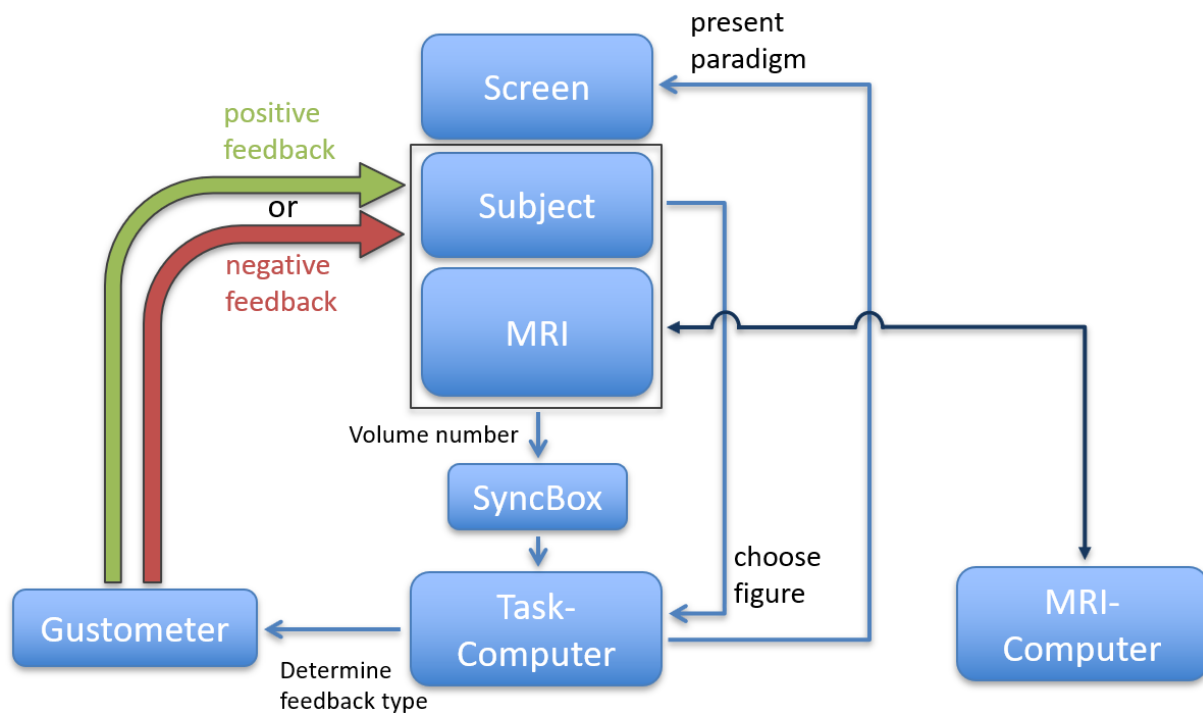| Parameter	Healthy Volunteers (N=10)
Age	26.2±4.2 (years)
Handedness right/left	10/0
BMI	22.3±1.9 ($\frac{kg}{m^2}$)
Fat mass index	3.85±1.48 ($\frac{kg}{m^2}$)
Visceral fat	1.03±0.61 (l)
Glucose	87.5±6.1 ($\frac{mg}{l}$)
Insulin	6.58±2.92 ($\frac{mU}{l}$)
HOMA-IR Index	1.50±0.70 ($\frac{mU*mmol}{l^2}$)

3.2   Experiment setup

Figure 4: Overview of the experiment setup

3.2.1.   Coordination of the experiment components

Synchronisation of stimulus presentation and fMRI data acquisition was accomplished with a SyncBox (NordicNeuroLab, Bergen, Norway), which sent a signal to the task-computer each time a volume was being acquired by the MRI.

The paradigm was presented on a computer screen placed behind the MRI and the participant viewed it through a mirror mounted on the head coil. During each trial, the participant chose a cue by clicking on either the left or right button of a computer mouse placed in his dominant hand and the output was sent to the computer running the task, which then calculated the expected feedback. A control signal was subsequently sent to the gustometer pumps, which then delivered one of the fluids.

The scripts managing the visual presentation of the paradigm, the participants' interaction with it and the activation of the gustometer were implemented in MATLAB (version R2014b, The Mathworks Inc.) using the Psychtoolbox function library (version 3.0.13.).[121–123] For the scripts used in this study, refer to appendix 7.4.1.

3.2.2. Gustometer setup

The Gustometer is a mechanical system, which allows the delivery of fluids to the test participant in the MRI with exact control of timing and quantity, enabling gustatory-stimuli-based feedback for each individual trial. Our Gustometer setup was a modified version of the setup described by Veldhuizen et al.,[124] and consisted of four programmable syringe pumps (model LA-100, HLL Landgraf Laborsysteme, Langenhagen, Germany), each loaded with a $50\,ml$ syringe (Braun, Melsungen, Germany). Two Syringes were filled with a milkshake individually pre-chosen by the participant and two with a pre-chosen neutral-tasting solution. The use of a second syringe for each fluid was used as a reserve to assure sufficient total fluid amount for the entire experiment. From the syringes, the fluids arrived at the participant's mouth over four silicon beverage tubes (Lindemann GmbH, Helmstedt, Germany) with an inside diameter of $2\,mm$. The tubes were held in the participant's mouth by fixing them to a custom-made adapter attached to the head coil and the length was individually adjusted so that the tubes' end rested comfortably on the participant's teeth row.

Each activation of a pump delivered ca. $0.45\,ml$ of fluid to the participant's mouth. Participants were instructed to swallow the delivered amount immediately and to not accumulate it between trials.

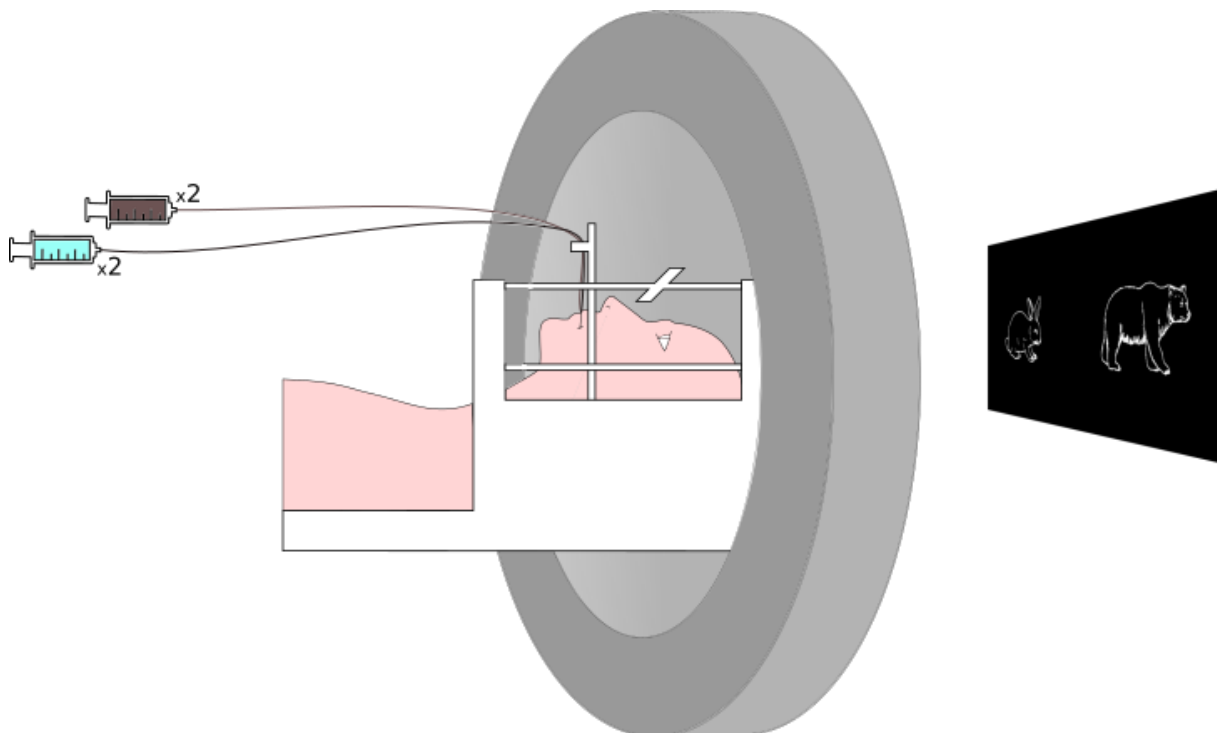For the settings we applied for the gustometer, refer to appendix 7.4.5.

Figure 5: Illustration of a participant laying in the MRI machine, viewing the computer screen through a mirror mounted on the head coil and receiving gustatory stimuli directly to his mouth.

3.3　Behavioural task

3.3.1.　　Paradigm design - probabilistic selection task

For the behavioural task, we implemented a modified version of the previously established PST which compromised two phases: a learning phase and a testing phase.[69,80,81]

The learning phase was designed as a "multi-armed bandit" task – each trial consisted of a choice between two visual cues, each with a different probability to receive a reward.[125] In total, six unique visual cues were presented (marked here as A-F) each with a certain predetermined win probability (A: 80%, B: 20%, C: 70%, D: 30%, E: 60%, F: 40%) which was hidden from the participant and remained consistent throughout the entire phase. In this phase, only three combinations of cues were presented: A-B, C-D and E-F, with each pair recurring 92 times. Pair order was randomized and individual cues were pseudo-randomized to appear on the left side of the screen in exactly half of their respective trials and on the right side in the other half. In order to allow a better comparison of results to similar previous studies we used the same set of visual cues as previously used by Jocham et al.[81] and Sevgi et al.[80]. This simple cue set had the advantage of reducing reliance on working memory, supporting our focus on habitual learning and would avoid concerns of discernibility with the original cue-set used by Frank et al.[69], which led to inconsistent task performance.[86]

Each participant received a different randomly picked set of six cues from a larger pool of possible images. See figure 6 for an example pairing structure.

Feedback outcomes were pseudo-randomized in consecutive bins of ten for each cue (with the last bin consisting of only two trial outcomes), each bin with a set amount of positive and negative outcomes according to the contingency of the cue, with the internal order of the bin randomized irrespective of the pair partner. e.g., for cue A with 80% win probability, each of the bins (except the last) consisted of precisely eight positive and two negative feedbacks in random order.

In addition, 27 null events were inserted randomly between the trials, each $4.3\,s$ long. In these trials, participants were shown only the fixation crosshairs and were neither required to make a decision nor received any gustatory stimulus.

For a schematic description of the structure and timings of an example trial, refer to figure 7.

If the participant did not click any mouse-button in the allotted $1.7\,s$, he was presented with a message informing him he was too slow, which remained on-screen for $2\,s$ before moving on to the next trial.

In the analysis, we calculated accuracy as the percentage of correct choices for all trials in total as well as for each individual pair. We also calculated the cumulative accuracy for each pair as the task progressed.

In the testing phase, participants were presented all possible pair permutations of the same six visual cues they were presented in the learning phase, with each combination repeating eight times. During this phase, participants no longer received feedback and were instructed during their pre-scan training to choose the "better" cue, i.e., the cue associated with more positive feedback, as deduced in the learning phase. 12 null events were added to the set of trials and were each $4\,s$ long.

For a schematic description refer to figure 7.

In the analysis, we computed the accuracy for all trials, trials containing the original pairs from the learning phase A-B, C-D and E-F as well as trials with the 12 novel combinations. Moreover, we calculated the accuracy for "choose A" (A-C, A-D, A-E, A-F) and "avoid B" (B-C, B-D, B-E, B-F) trials. Finally, we calculated the accuracy for high conflict trials: win-win (A-C, A-E, C-E) and lose-lose (B-D, B-F, D-F) as well as for the remaining 9 low conflict trials.

In both phases, the intertrial intervals were randomised, or "jittered", and varied between $0.6\,s$ and $2\,s$, determined by randomly sampling from a truncated exponential distribution as described in equation 2, chosen after performing the design efficiency analysis as described in section 7.4.4.

$$J\ =\ m\ +\ X\,,\quad \{\,J \in \mathbb{R}_+^l\,, m = 0.6s\ ;\ l\ =\ 276\ or\ l\ =\ 120\,\}$$
$$X \sim Exp(\mu)\,,\quad \mu = 0.3$$
$$f_x(x|\mu) = \mu e^{-\mu x}\,,\quad \{x > 0\}$$

Equation 2: Jitter length distribution

where J is a vector whose components are the jitter time length for each individual trial, m is a scalar that denotes the minimal jitter time and X is a vector whose elements donate to the jitter time and have an exponential distribution with the rate parameter μ. $f_x(x|\mu)$ is the probability density function for $Exp(\mu)$.
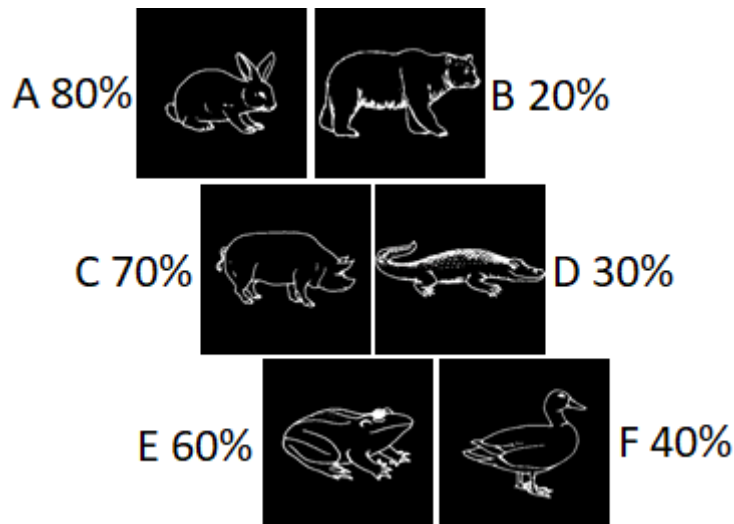
Figure 6: Design of the PST visual cue-pairs in the learning phase. Six cues were chosen at random for each participant from a larger pool of possible images.
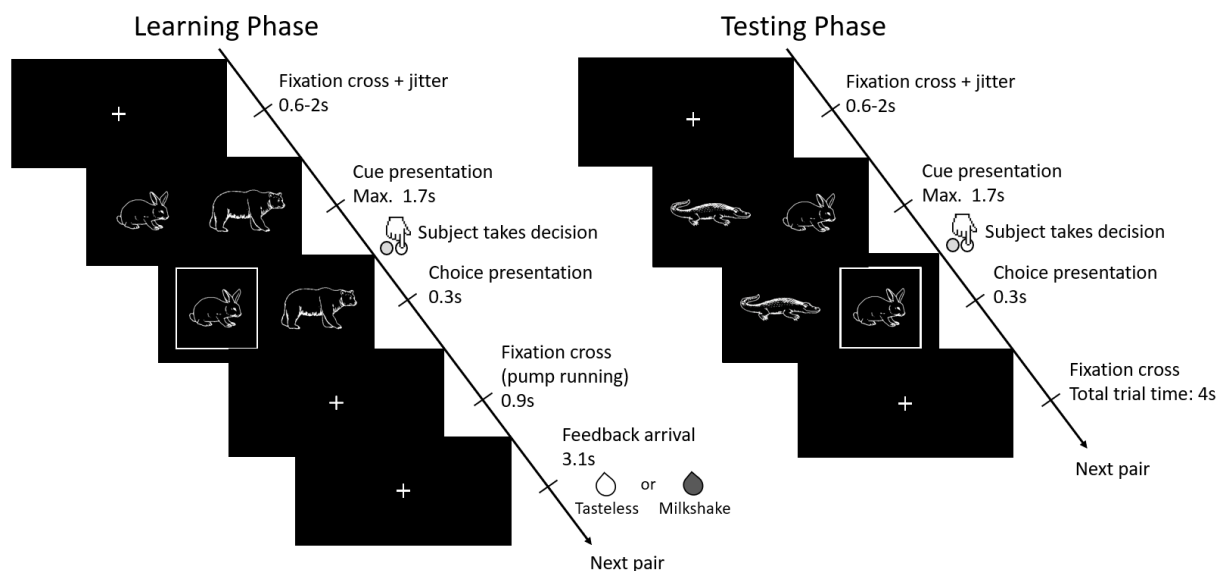

Figure 7: Trial structure. The learning phase trial design is shown on the left and the testing phase trial design on the right.

3.3.2.   Pre-scan training

Before entering the MRI, participants were familiarized with the task by receiving an explanation and performing a short simulation on a computer. In the simulation, participants were presented three pairs of cues and one null event for each phase of the experiment and were also instructed not to choose during one trial in order to simulate a "missed event". During the simulation, the outcome of each trial was given on-screen as direct visual feedback in place of the gustatory stimuli. Participants were notified that the cues in the experiment may be different from the ones presented during the simulation.

After confirming that the task was understood, participants were prepared for the MRI to perform the task.

3.3.3. Computational modelling

Data from each individual participant were fitted to the Q-Learning model as described by the following formula[126]:

$$Q_{cue}(t+1) = Q_{cue}(t) + \alpha \cdot \delta(t+1)$$

Equation 3: Q-learning rule

where $Q_{cue}(t)$ is the SAV (or Q-Value) attached to the cue at time point (trial number) t, α is the learning rate parameter of the actor and $\delta(t)$ is the PE at time t.

In addition to the model with a uniform learning rate for all trials, we evaluated models with separate learning rates: α_+ for trials which resulted in positive feedback and α_- for trials which resulted in negative feedback.

PE was defined as:

$$\delta(t+1) = r(t+1) - Q_{cue}(t)$$

Equation 4: Prediction error rule

where positive feedback at trial t is defined as $r(t) = 1$ and negative feedback as $r(t) = 0$.

For the decision-making function, which gives the model's probability P of choosing a cue X over the alternative cue Y at trial t, we used the SoftMax decision rule:

$$P(t)_x = \frac{e^{\frac{Q_x(t)}{\beta}}}{e^{\frac{Q_x(t)}{\beta}} + e^{\frac{Q_y(t)}{\beta}}}$$

Equation 5: SoftMax decision rule

where β is the temperature parameter. The SAV, learning rate and temperature were calculated for each participant using a modified version of the "VBA toolbox" software package for MATLAB.[75]

8 different models were compared by combining different constraints for the Q-Learning algorithm:

Learning rate α: uniform vs. separate for positive and negative feedback,

Temperature β: individually fitted vs. a fixed value for the entire group,

Data input: only from the learning phase vs. combined from both phases.

Prior state-action values were set to 0.5 for all cues.

In order to find the model from the above that best describes the participants, we performed a random-effects Bayesian model selection with the free energy approximation as the selection criterion, similar to the method described by Rigoux et al.[77]

3.4 fMRI

3.4.1. Jitter design efficiency analysis

In order to choose a design for the paradigm which will also be efficient for the imaging analysis, a custom script was written in MATLAB which simulated data with different intertrial and interstimulus intervals distributions.

The efficiency of each design was compared for each of the contrasts of interest using a method similar to the one described by Henson et al.[94] using the formula in equation 1.

In addition, Laplace Chernoff risk was computed in order to estimate general efficiency over the various contrasts for each design, in a method similar to that described by Daunizeau et at.[127]

$$eff(c) = \frac{1}{c^t * pinv(X^t * X) * c}$$

Equation 6: Efficiency score for a single contrast

$$LC = -trace(C^t * pinv(X^t * X) * C)$$

Equation 7: Laplace-Chernoff risk for all contrasts

Where eff(c) is the efficiency for the contrast c, which is a vector consisting of values of 1 for regressors of interest, -1 for regressors of interest in the negative weight and 0 for regressors of non-interest. X is the design matrix of the GLM. LC is the Laplace-Chernoff risk and C denotes the matrix of all contrasts of interest.

The pinv function gives the Moore-Penrose Pseudoinverse and the trace function is the sum of the elements in the diagonal of the matrix.

Initially, the test was performed with choice data acquired from a pilot participant, who experienced feedback in the form of visual outcomes - smiley-faces and frowny-faces instead of gustatory stimuli and performed a slightly shorter task consisting of a learning phase with only 80 repetitions per pair in the learning phase instead of 92 repetitions as later performed by the study cohort. The analysis of the data from the pilot participant was only performed for a single contrast of positive PE over negative PE. 500 iterations were run for each jitter design. A jitter design was chosen by balancing the differences in efficiency scores against additional scan time.

The results of this analysis, which was based on a different paradigm design and only had a narrow scope of contrasts was limited in its value. In order to investigate concerns that the chosen paradigm design was suboptimal and weakened the differentiation of contrasts in SPM, a post hoc analysis was performed by repeating the test using the data acquired from all 10 participants and comparing further jitter designs and contrasts. Each design was run with 50 iterations for each choice set acquired from the participants. For the full set of tested distributions refer to appendix 7.4.4.

3.4.2. fMRI and anatomical MRI data acquisition

The imaging data were acquired at the Max Planck Institute for Metabolism Research at the University Hospital of Cologne using a 3T MRI Scanner (Siemens Magnetom Prisma, Erlangen, Germany) equipped with a 64-channel head coil.

fMRI data were acquired using a gradient echo-planar imaging technique. 60 axial slices were sampled in a multiband sequence. Phase encoding direction was anterior-posterior. TR was $1220\ ms$, echo time (TE) $30\ ms$, flip angle 80°, echo spacing 0.93 ms, echo-planar imaging factor 140, voxel size $1.5x1.5x1.5\ mm^3$, field of view (FOV) $210\ mm$. The FOV was set parallel to the commissural line and then tilted ca. 30° dorsally, using the ventral edge of the frontal lobe as the bottom edge of the FOV.

After acquiring the functional data from the task, two additional brief echo-planar imaging-scans were acquired, each with three volumes, with the above-mentioned parameters but each sequence with opposite phase encoding direction (one sequence in anterior-posterior direction, the second in posterior-anterior). These were used for the estimation of susceptibility-induced off-resonance fields.

Anatomical imaging was performed in a separate session previous to the fMRI data acquisition. High-resolution T1-weighted and T2-weighted were acquired with whole brain coverage using a 64-channel array head coil. The T1 images were acquired using the following parameters: 192 sagittal slices, TR=$2300\ ms$, TE=$2.32\ ms$, inversion time (TI)=$900\ ms$, flip angle 8°, voxel size $0.9x0.9x0.9\ mm^3$, FOV=$230\ mm$.

The T2 images were acquired using the following parameters: 224 sagittal slices, TR=$3200\ ms$, TE=$460\ ms$, TI=$900\ ms$, voxel size $0.8x0.8x0.8\ mm^3$, FOV=$256\ mm$. For both sequences the FOV was set parallel to the commissural line.

3.4.3. Preprocessing

Preprocessing of individual imaging data sets was performed using with the FMRIB (Oxford Centre for Functional MRI of the Brain) Software Library (FSL) (version 5.0.11, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki)[72,128,129] and SPM (version 12, Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm) software.[108,130]

First, the initial 10 echo-planar imaging volumes ("dummy scans") were discarded to allow for T1-equilibrium effects. The volumes were then corrected for head motions by realignment with the MCFLIRT tool[131] and corrected for susceptibility distortions using the topup tool with a similar method to that described by Andersson et al.[132]

The functional volumes were then co-registered to the anatomical T1 images using the anatomical T2 images as an intermediary. We then used SPM to perform tissue segmentation of the images (Bias regularization=0.001, Bias FWHM=60 mm cutoff), normalization to the Montreal Neurological Institute (MNI)-152 standard brain and Gaussian smoothing ($8\,mm$ FWHM kernel).

In addition, following nuisance signals were calculated using FSL and added to the realignment parameters as multiple regressors to account for further artefacts: motion outliers, calculated as "Derivative or root mean square variance over voxels" (DVARS)[133–135], as well as mean signal derived from spatial masks from tissue compartments of non-interest (white matter and cerebrospinal fluid)[135] eroded using a spherical kernel with a $9\,mm$ radius for the white matter mask and $2\,mm$ radius for the cerebrospinal fluid mask.

3.4.4.   First-level statistical analysis

On the first-level analysis, data from individual participants and phases were fitted to a GLM using SPM to investigate neurological activity related to the reward system as well as to investigate the main effects of vision, motor control, tasting and sensing while performing the paradigm. The canonical hemodynamic response function with temporal and dispersion partial derivatives was used to model the data.

For the learning phase, the following regressors (with their respective parametric modulators in parenthesis) were chosen:

onsets of cue presentation (SAV of the chosen cue in the trial), positive outcome arrival (positive PE), negative outcome arrival (negative PE), missed events, motion parameters, motion outliers and nuisance signals from white matter and cerebrospinal fluid. PE and SAV were derived from the superior model as chosen in the model comparison.

For the testing phase, the following regressors were chosen: onsets of events choosing cue A, avoiding cue B (i.e. choosing the alternative cue in trials that also present cue B) and other choice events. The GLM also included regressors for onsets of missed events, motion parameters, motion outliers and nuisance signals from white matter and cerebrospinal fluid.

3.4.5.   Second-level statistical analysis

Group-level analysis was performed on the first-level results using a flexible fractal design. For the learning phase, we analysed the main effects of cue presentation, outcome arrival, SAV and PE. We also analysed the contrasts between the different outcome and PE types.

For the testing phase, we analysed the main effect of any cue presentation, as well as the contrasts between choose A and avoid B events.

4. Results

4.1 Participants

One participant has been excluded from the analyses of both behavioural and fMRI data of his testing phase due to a technical issue with the MRI machine preventing him from completing the phase.

Another participant was analysed only for his behavioural data and excluded completely from the fMRI analysis, due to a technical failure with the acquisition of the imaging data.

For a characterization of the participants, including blood test and bio-electrical impedance analysis results, refer to table 1.

4.1.1. Taste test - milkshake ratings

One participant did not perform the entire post-scan rating due to technical issues and therefore has no results for that test. A second participant had technical issues with a few of the post-scan subtests (hunger, thirst and satiety) so these results were excluded from the analysis.

All the participants showed reduced post-scan scores for "liking" and "wanting", but for some participants the decrease was more pronounced than for others. Refer to figure 8.

When correlating the change in rating scores of each participant to their corresponding performance in the testing phase, a moderately strong negative monotonic correlation was found between the amount of reduction in the "wanting" visual analogue scale rating from pre- to post-scan (Δ"wanting") and accuracy in "choose A" trials in the testing phase ($r_s = -0.6$, Spearman's rank correlation coefficient). In addition, a fairly strong negative monotonic correlation was found between the Δ"liking" score and win-win accuracy ($r_s = -0.49$), as well as a fairly strong positive monotonic correlation with "avoid B" accuracy ($r_s = 0.47$). For a more detailed overview, refer to table 2.

Table 2: Correlation matrix between participants' rating scores and accuracy performance in the testing phase. Results are given in Spearman's rank correlation coefficients.

	Δ"liking"	Δ"wanting"	Δ"hunger"	Δ"satiety"	Δ"thirst"	accuracy total	accuracy choose A	accuracy avoid B	accuracy win-win
Δ"wanting"	0.08								
Δ"hunger"	-0.71	0.21							
Δ"satiety"	0.47	-0.19	-0.76						
Δ"thirst"	-0.73	-0.10	0.60	-0.17					
accuracy total	0.32	-0.10	-0.14	0.33	-0.14				
accuracy choose A	-0.05	-0.60	0.01	0.17	0.34	0.41			
accuracy avoid B	0.47	0.07	-0.30	0.33	0.26	-0.05	0.31		
accuracy win-win	-0.49	-0.38	0.22	-0.11	0.26	0.43	0.69	-0.36	
accuracy lose-lose	0.36	-0.11	-0.11	0.38	0.45	0.33	0.50	0.85	-0.14
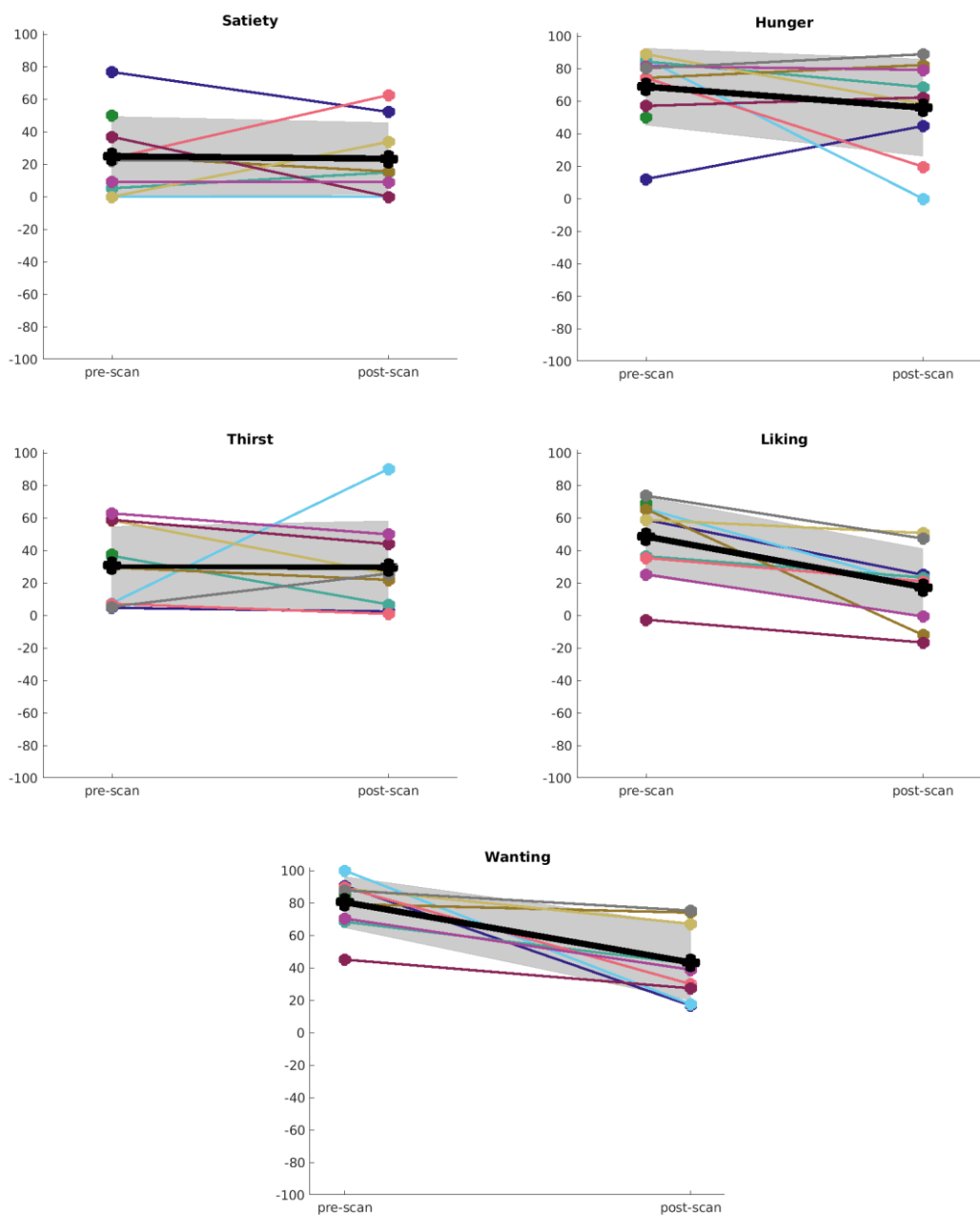
Figure 8: Milkshake taste-test rating scores. Scores are given on a visual analogue scale, mean rating score change is shown as a bold line ± standard deviation as a grey band.

47

4.2 Behavioural results

4.2.1.   Task performance

Learning phase

For the least difficult pair A-B, all participants were able to successfully learn the task, consistently achieving over 75% accuracy after an average of only 7 trials out of a maximal 92 trials with this pair, and achieving a mean accuracy of 91.3%±8.4% at the end of the learning phase. For the more difficult pairs C-D and E-F, not all participants were able to learn the task correctly or have managed to do so only very late in this phase, with two participants performing around or below chance level (i.e., 50% accuracy at the end of the phase) for each pair. Nonetheless, on average, the group managed to learn the tasks also for these pairs, achieving an accuracy of 76.3%±18.7% for the C-D pair, and 73.0%±25.3% for E-F at the end of the phase. The difference between the accuracies of the pairs was not statistically significant (repeated measures ANOVA $F_{(2,18)}=3.52$, $p=0.051$). Total mean accuracy at the end of the phase was 80.2%±13.2% (Data are given as mean±SD).
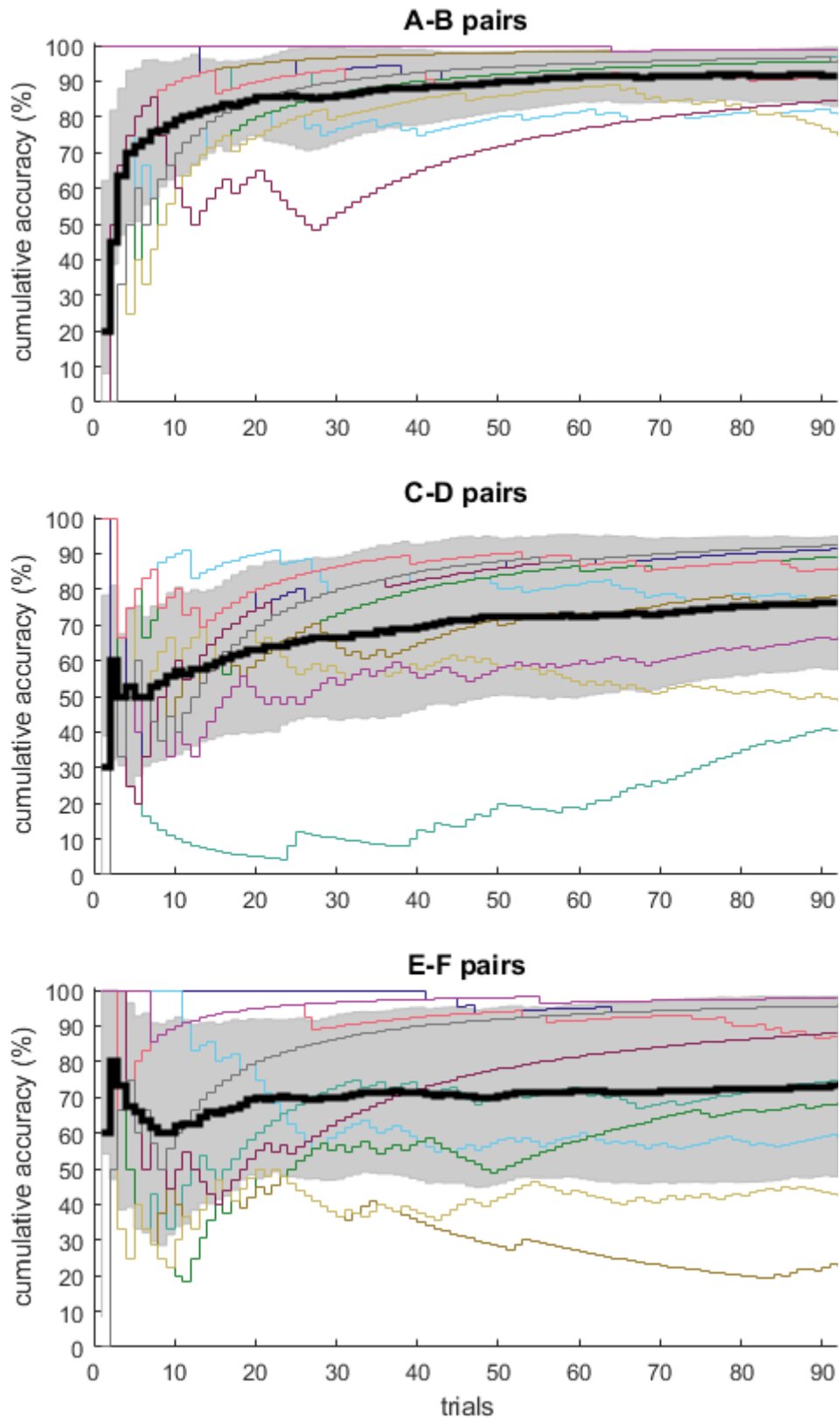
Figure 9: Cumulative accuracy in the learning phase by pair. Each individual participant's cumulative accuracy as trials progressed is shown as an individual line. Mean cumulative accuracy is shown as a bold line ± standard deviation as a grey band.

Testing Phase

Total mean accuracy for all trials was 79.3%±8.8%. Combined mean accuracy for the three original pairs from the learning phase was 88.6%±13.3% at the end of the testing phase. The accuracy for the A-B pair was 94.4%±12.7%, for C-D 88.9%±15.9% and for E-F 76.0%±32.7%. The lowest individual scores were 62.5% for both C-D and E-F pairs, thus all participants managed to perform the previously learned task at least slightly better than at chance level. The difference between the mean accuracies of the original pairs was not statistically significant (repeated measures ANOVA F(2,16)=1.61, p=0.23).

For the novel pairs, the mean accuracy was 77.5%±10.1%, with the difference to the original pairs not being statistically significant (unpaired t-test t(16)=1.62). Mean accuracy for choose A trials was 88%±20.0% and for avoid B 73%±18.0%, with the difference not being statistically significant (unpaired t-test t(16)=1.67). When comparing individual choose A and avoid B results to classify the participants using the approach suggested by Frank et al.[69,89], six of the participants were classified as approach learners, one as an avoidance learner and the remaining two had an equivalent performance in both criteria. One participant had a noticeably low accuracy at the choose A trials (37.5%) and win-win trials (25%) which also contributed to him obtaining the lowest total accuracy in the group (63%). According to his self-reporting, he incorrectly estimated the internal order of the winner cues (C>E>A).

When excluding this participant, the difference between the choose A (94.1%±7.0%) and avoid B trials (72.3%±19.1%) is statistically significant (unpaired t-test t(14)=3.05, p<0.01). Accuracy of high/low conflict pairs: win-win: 73.6%±23.5%, lose-lose: 57.9%±26.0%, low conflict: 88.4%±11.1% (repeated measures ANOVA F(2,16)=4.57, p<0.05). When excluding the participant who performed weakly: F(2,14)=4.22, p<0.05. significance between conditions does not change. (Data are given as mean±SD)

Response times: mean response time across all trials was $892 \pm 157\,ms$. For choose A trials response time was $774 \pm 119\,ms$ and for avoid B trials $964 \pm 189\,ms$ (paired t-test t(9)=4.27, p<0.005). For win-win trials, mean response time was $853 \pm 56\,ms$, for lose-lose trials $1025 \pm 60\,ms$ and for low conflict trials $861 \pm 53\,ms$ (repeated measures ANOVA F(2,16)=19.01, p<0.001). (Data are given as mean±SD)
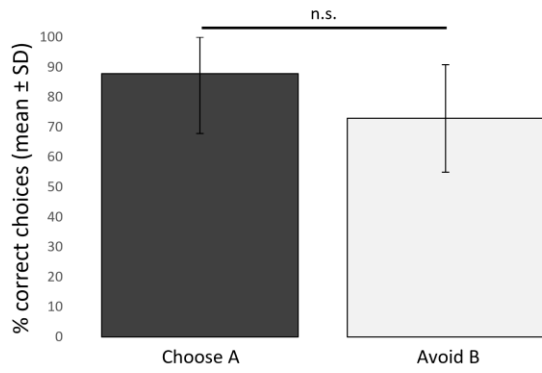
Figure 10: Mean accuracy for choose A and avoid B trials in the testing phase. n.s. = not significant
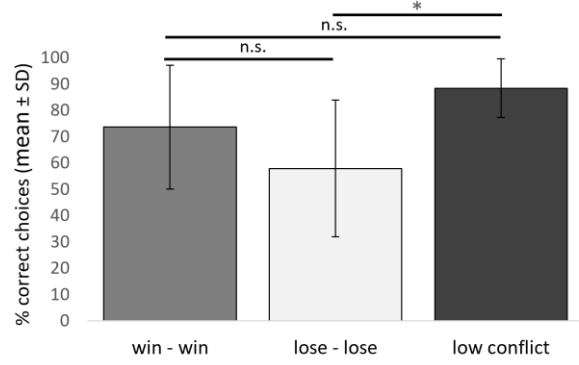
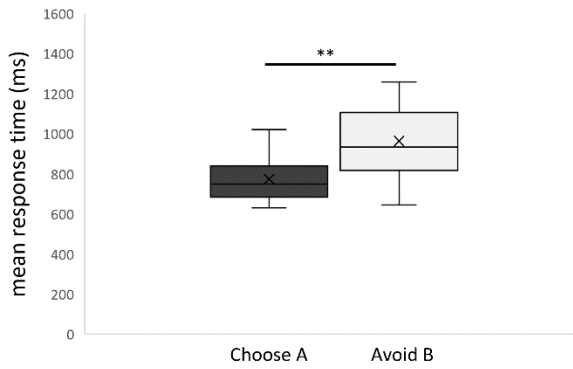*Figure 11: Mean accuracy for high- and low-conflict trials in the testing phase. *p<0.05, n.s. = not significant*

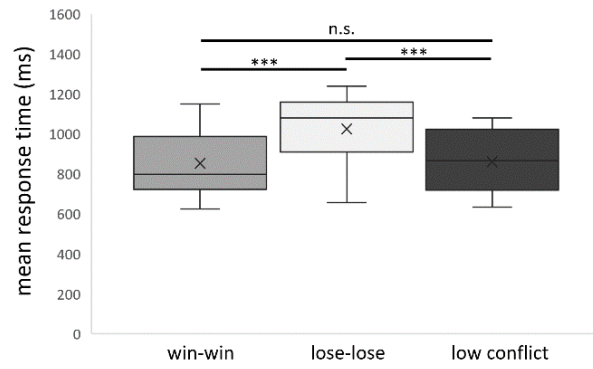*Figure 12: Mean response times for choose A and avoid B trials during the testing phase. **p<0.005*

*Figure 13: Mean response times for high- and low-conflict pairs during the testing phase. ***p<0.001, n.s. = not significant*

4.2.2. Computational modelling

Comparative analysis of the 8 proposed models for the reinforcement learning framework showed most participants could best be modelled by the "extended asymmetrical" model and to a lesser extent with the "extended unitary" model, which were both based on data from both phases and used an individually fitted temperature parameter β. The asymmetrical model used separate learning rates α_+ and α_- for positive and negative feedback and the unitary model used a single uniform learning rate α for both types of feedback.

Five participants were explained best by the extended asymmetrical model, of them three had also a partial attribution to the extended unitary model. Four participants were better modelled with the extended unitary model, yet all had partial attributions to the asymmetrical model. Noticeably, one of these participants was completely lacking data from the testing phase due to technical issues, so the results contain an artefact and should be disregarded.

One participant was best modelled solely with the "canonical" model, which used a single learning rate, an individually fitted temperature and data only from the learning phase. Interestingly, this participant was also the participant who showed the worst total accuracy (63%) and worst accuracy for novel pairs in the testing phase (57%), despite showing a very good accuracy in the learning phase (94%).

For a detailed graphical overview, see figure 24.

For the extended asymmetrical model, mean α_+ was calculated to be 0.13±0.15, mean α_- was 0.08±0.17 and mean β was 0.11±0.06.

For the extended unitary model, mean α was 0.14±0.17 and mean β was 0.13±0.08.

For the canonical model, mean α was 0.19±0.23 and mean β was 0.16±0.15.

For a detailed overview of the correlation between the participants' task performance and the parameters calculated in the computational models, refer to table 3.

Final SAV for the extended asymmetrical model were: A(80%): 80%±17%, B(20%): 40%±15%, C(70%): 71%±21%, D(30%): 46%±19%, E(60%): 65%±19%, F(40%): 50%±21%. (Data are given as mean±SD)

For a more detailed comparison refer to figure 14. For a comparison between the three models on how the individual SAV for a single cue progressed over the course of the learning phase, refer to figure 15 and to appendix figure 25 to 29 for the remaining cues.

Table 3: Correlation matrix between participants' task performance and calculated learning parameters. Results are given in Spearman's rank correlation coefficients. Positive LR, negative LR and temperature asym are derived from the extended asymmetrical model. LR = learning rate ; asym = asymmetrical learning rate.

	total accuracy	choose A	avoid B	win-win	lose-lose	low conflict	positive LR	negative LR	canonical LR
accuracy choose A	0.41								
accuracy avoid B	-0.05	0.31							
accuracy win-win	0.43	0.69	-0.36						
accuracy lose-lose	0.33	0.50	0.85	-0.14					
accuracy low conflict	0.51	-0.47	-0.58	0.00	-0.35				
positive LR	0.07	0.39	0.69	0.09	0.41	-0.43			
negative LR	0.18	0.14	-0.07	0.45	-0.23	0.13	0.53		
Canonical LR	0.07	0.14	0.37	-0.12	0.18	-0.29	0.38	0.45	
extended unitary LR	0.10	0.42	0.60	0.10	0.36	-0.36	0.88	0.67	0.60
Temperature asym	-0.37	-0.19	0.59	-0.37	0.18	-0.39	0.67	0.47	
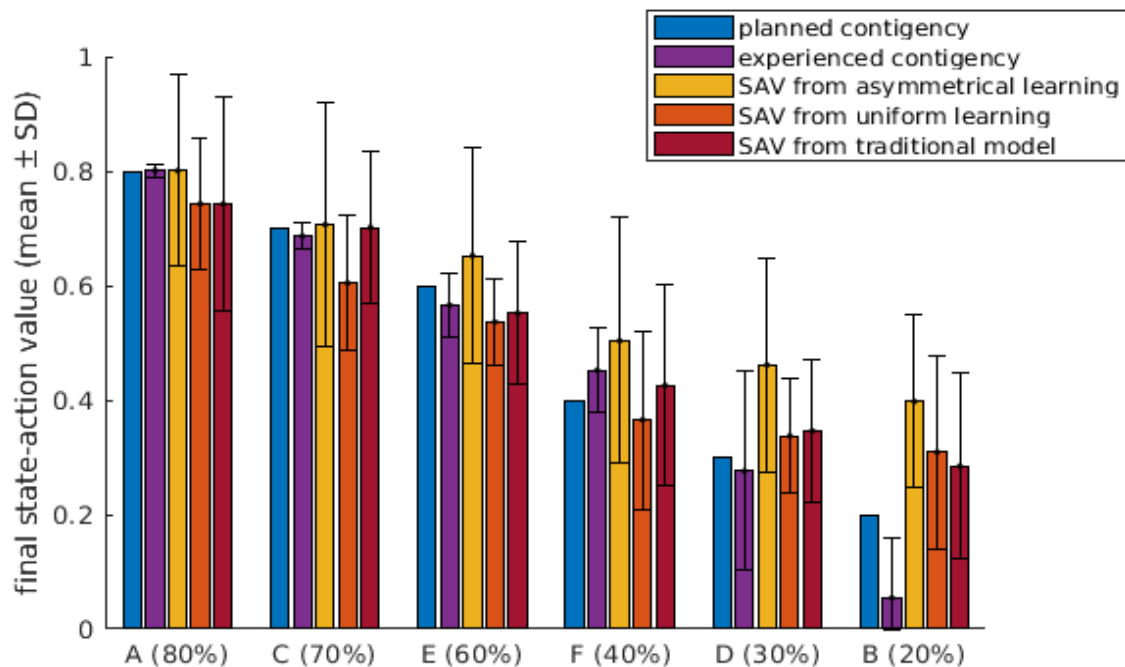
Figure 14: State-action values for each cue at the end of the task. Comparison of the theoretical values with the mean contingencies experienced by the participants and different calculations of SAV from three different models.
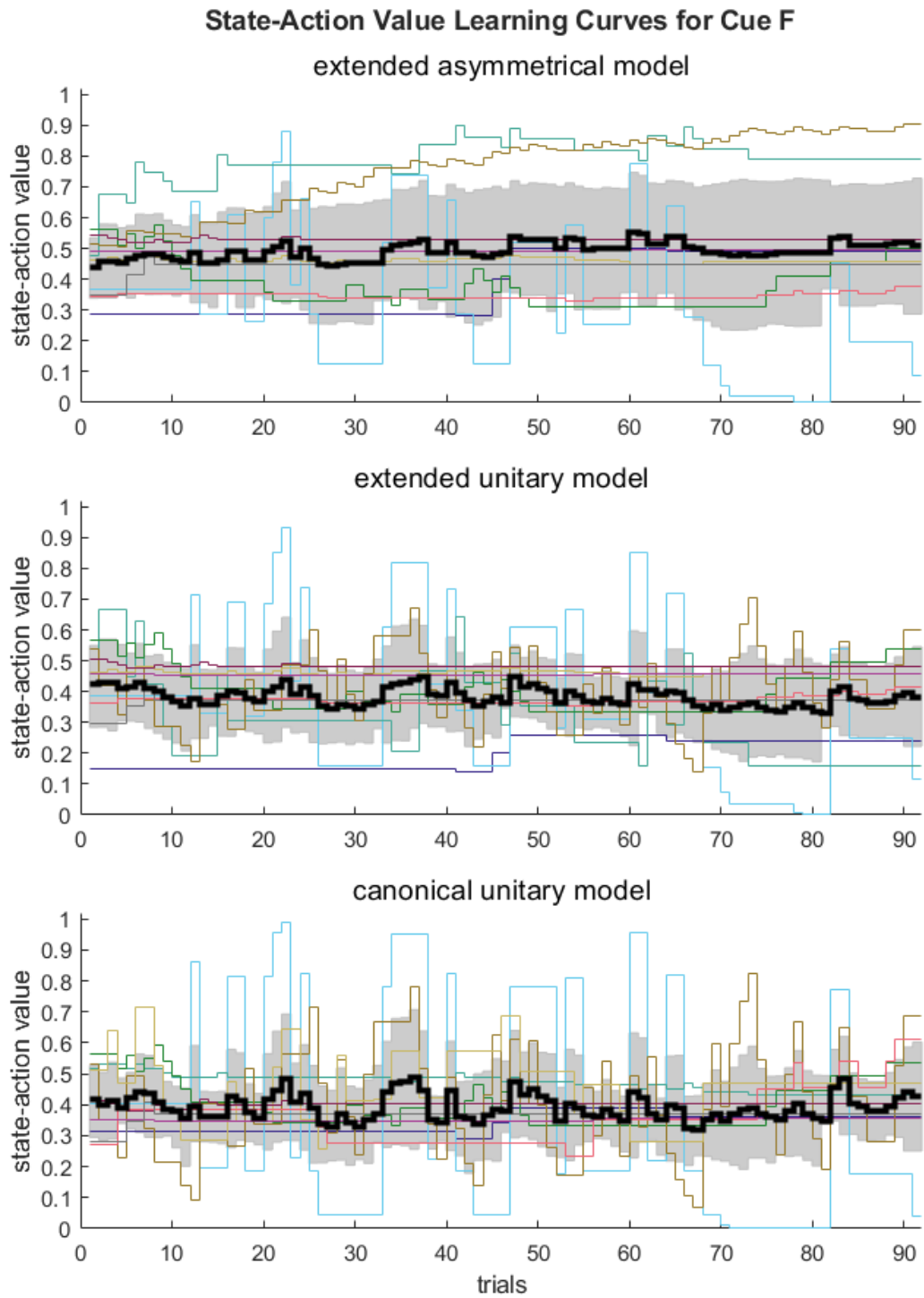
Figure 15: SAV for cue F as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.

4.3 Neuroimaging results

4.3.1. Design efficiency analysis

Figure 16: Jitter design efficiency
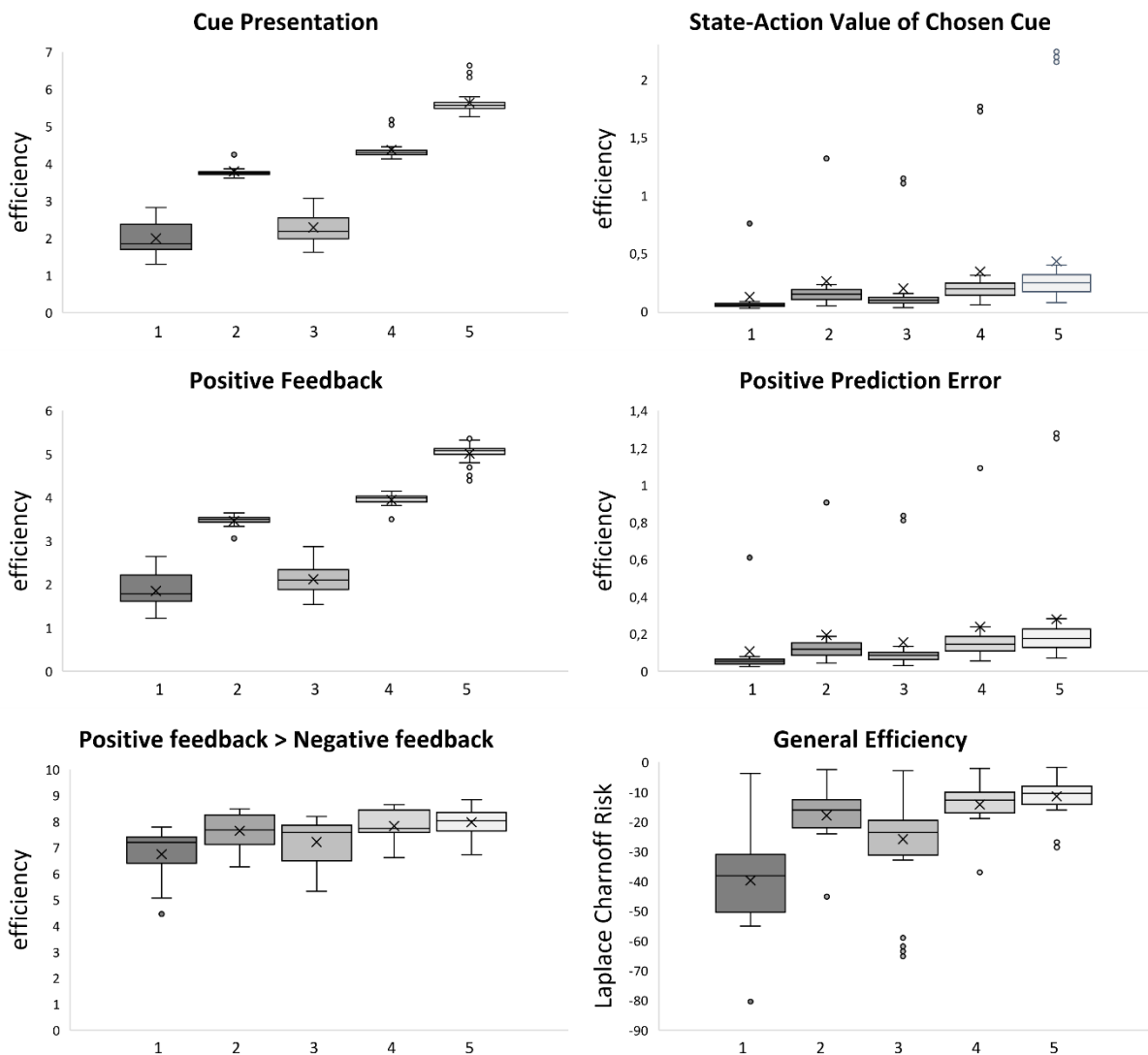
(1) Jitter design used in the study – truncated exponential distribution (m=0.6; μ=0.3) with real response time

(2) Jitter design used in the study (m=0.6; μ=0.3) with maximal response time

(3) Design split into two jitters sets (each m=0.3; μ=0.15) with simulated response time

(4) Design split into two jitters sets (each m=0.3; μ=0.15) with maximal response time

(5) Jitter design used in the study, duplicated also for the second jitter set (each m=0.6; μ=0.3) with maximal response time (adding ca. 4 minutes total scan time to design N°4)

The post hoc jitter design analysis showed that the efficiency of the fMRI analysis could have been improved both by jittering for longer lengths of time in general and especially by lengthening the interval between the onsets of cue presentation and feedback arrival in each

trial. In figure 16, the efficiency of the design we implemented in the study is compared to similar designs with increased interstimulus intervals, achieved by either complementing the response time the participants had remaining in each trial (N°2), by exchanging a part of the intertrial intervals for longer interstimulus intervals (N°3) or with both adaptations combined (N°4) as well as with a longer total scan time (N°5). For a more comprehensive analysis, also comparing designs with a different distribution of jitters and varied total jitter time refer to appendix 7.4.4.

4.3.2. Learning phase

Group-level analysis of cue presentation shows a cluster of activation in the visual cortex (cluster maximum x=22 y=-84 z=-16, T=30.59, k_E=26561 for uncorrected p<0.001, k_E=75 for FWE<0.05)
as well as some activation in the left motor cortex, likely corresponding to clicking the mouse button with the right hand (cluster maximum x=-50 y=-4 z=34, T=4.45, k_E=4261 for uncorrected p<0.001, no cluster at FWE<0.05). See figure 17.

Figure 17: Main effect of cue presentation. Learning Phase. Display threshold set to p<0.001 uncorrected.

Reward outcome, as well as outcome of negative feedback, showed bilateral activation on the face or lips part of the sensorimotor cortex - likely corresponding to the action of sensing and swallowing the fluid (cluster maximum, x=57 y=-2 z=30, T=11.83, k_E=473 uncorrected p<0.001, no cluster at FWE<0.05). The contrast of negative outcomes over positive outcomes showed activation in the anterior insula, especially of the left side (cluster maximum x=-26 y=20 z=10, T=17.62, k_E=902 for uncorrected p<0.001, k_E=3 for FWE<0.05) as well as activity in further structures such as the dorsolateral prefrontal cortex (cluster maximum x=32 y=40 z=14, T=12.20, k_E=4226 for uncorrected p<0.001, no cluster at FWE<0.05) and the anterior cingulate cortex (cluster maximum x=-2 y=-1 z=40, T=8.78, k_E=1399 for uncorrected p<0.001, no cluster at FWE<0.05). However, the reverse contrast showed no major activations, including a lack of gustatory cortex activation. See figure 18 and figure 19.

For the analysis of brain activity related to the SAV and PE parameters analysis from the extended asymmetrical model, see figure 20 and figure 21.
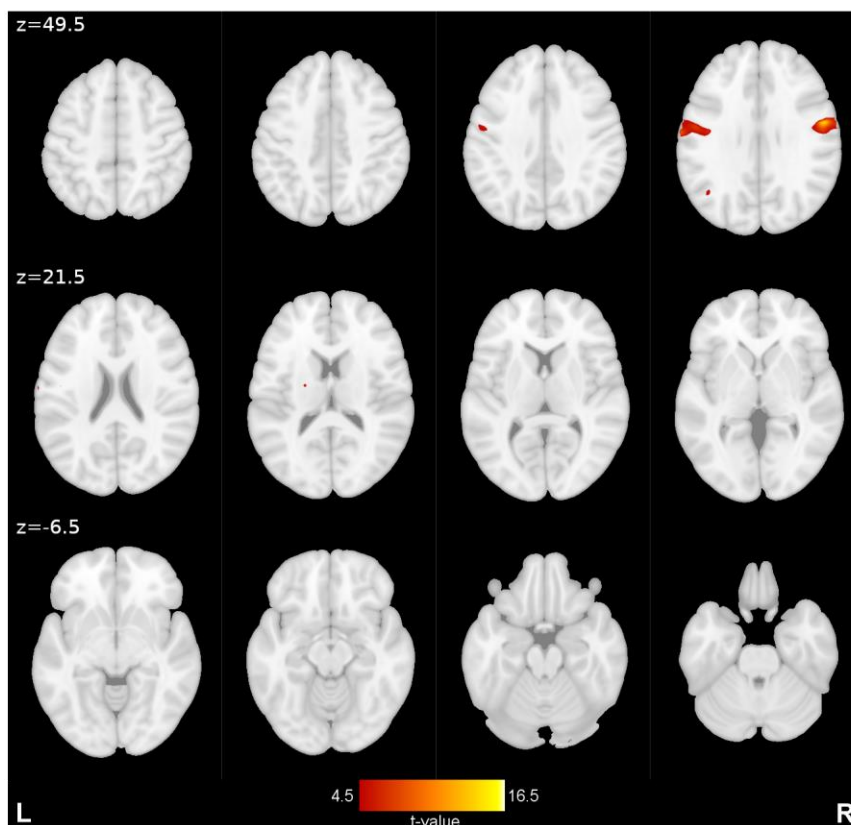
Figure 18: Main effect of reward. Learning Phase. Display threshold set to p<0.001 uncorrected.

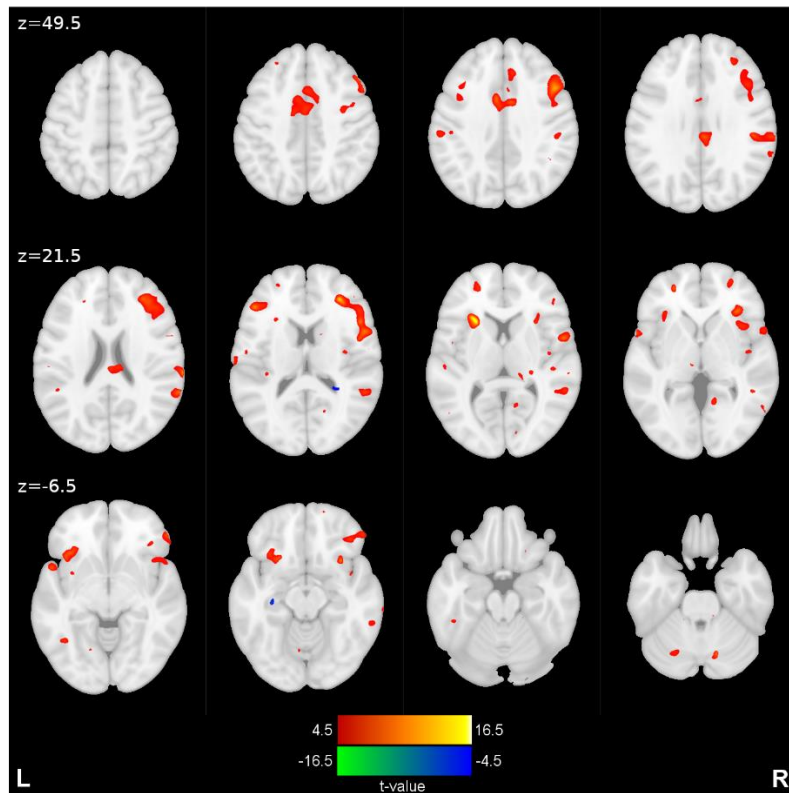# Negative Outcomes > Positive Outcomes

Figure 19: Contrast of negative outcomes (tasteless solution) over positive outcomes (milkshake). Blue voxels represent deactivation (i.e., activations in the reversed contrast). Learning phase. Display threshold set to p<0.001 uncorrected.
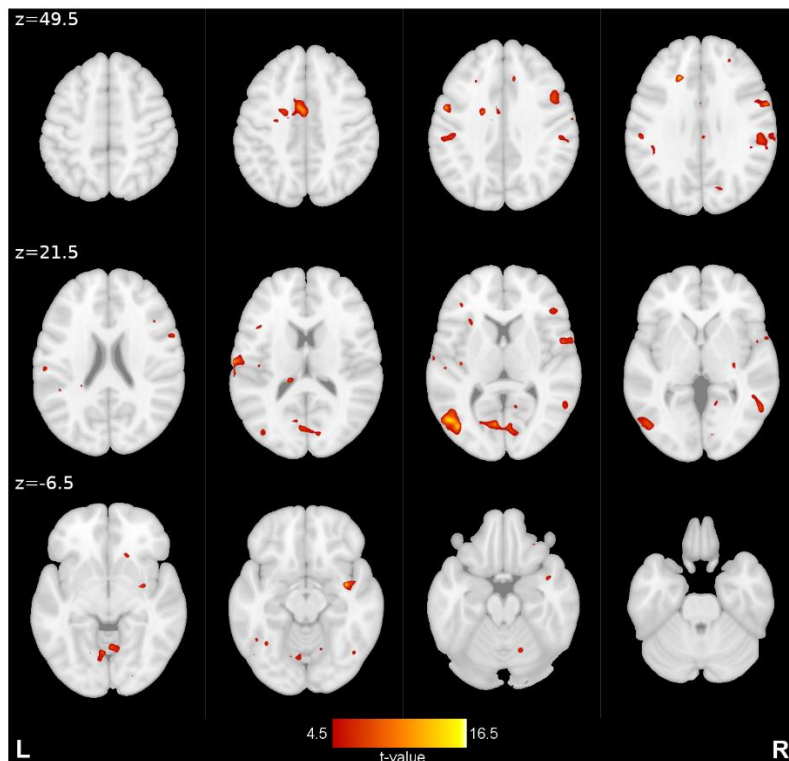
State-Action Value of Chosen Cue

Figure 20: Brain activity related to the state-action value of the chosen cue. The parameters were derived from the extended asymmetrical model. Learning phase. Display threshold set to p<0.001 uncorrected.

Prediction Error
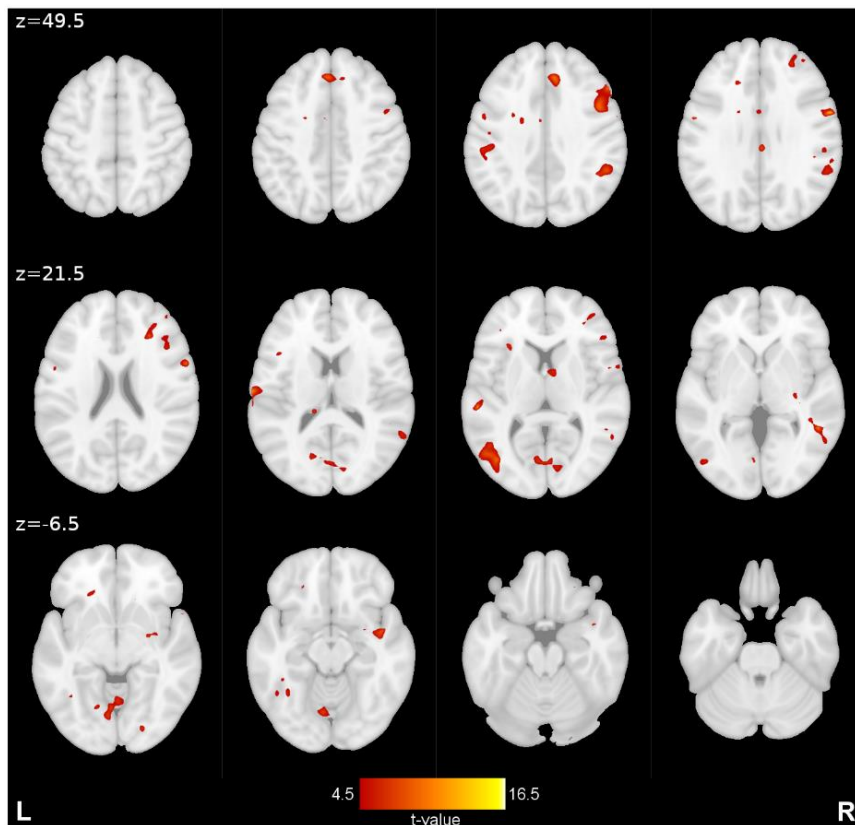
Figure 21: Prediction error-related brain activity for both outcome types. The parameters were derived from the extended asymmetrical model. Learning phase. Display threshold set to p<0.001 uncorrected.

4.3.3. Testing phase

Group-level analysis of cue presentation for all event types shows a cluster of activation in the visual cortex (cluster maximum x=20 y=-96 z=14, T=15.87, k_E=12310 for uncorrected p<0.001). Refer to figure 22.

The contrast of avoid B over choose A showed some activity in the dorsomedial prefrontal cortex (cluster maximum, x=-2 y=34 z=34, k_E=545 uncorrected p<0.001, no cluster at FWE<0.05) and the anterior cingulate cortex (cluster maximum, x=4 y=36 z=16, k_E=227 for uncorrected p<0.001). The reverse contrast showed no major activity. See figure 23.

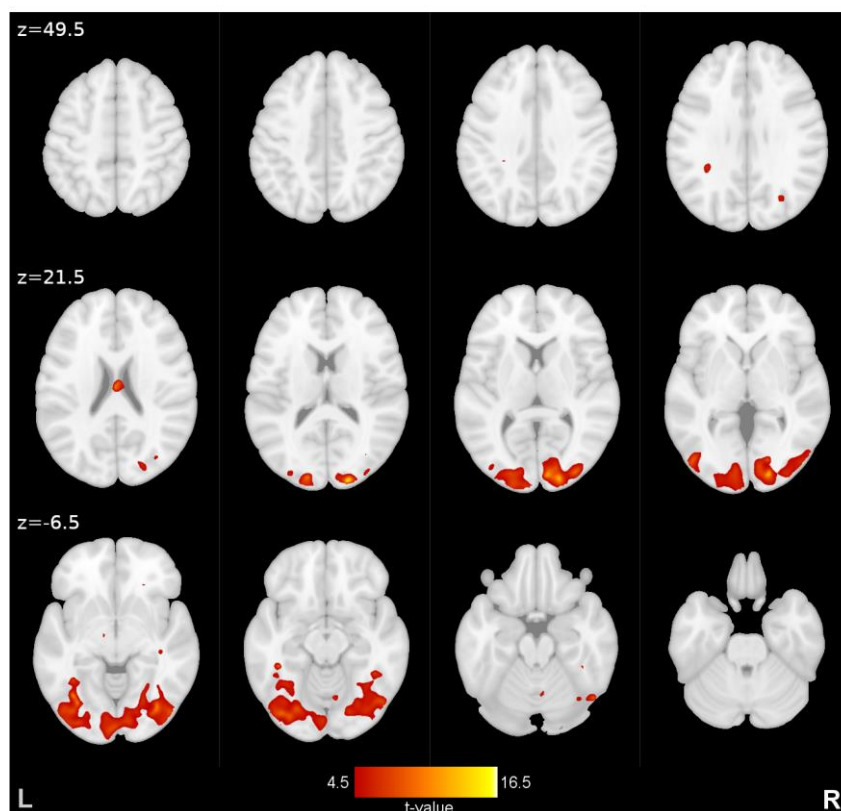# Cue Presentation. Testing Phase

Figure 22: Main effect of cue presentation. Testing Phase. Display threshold set to p<0.001 uncorrected.
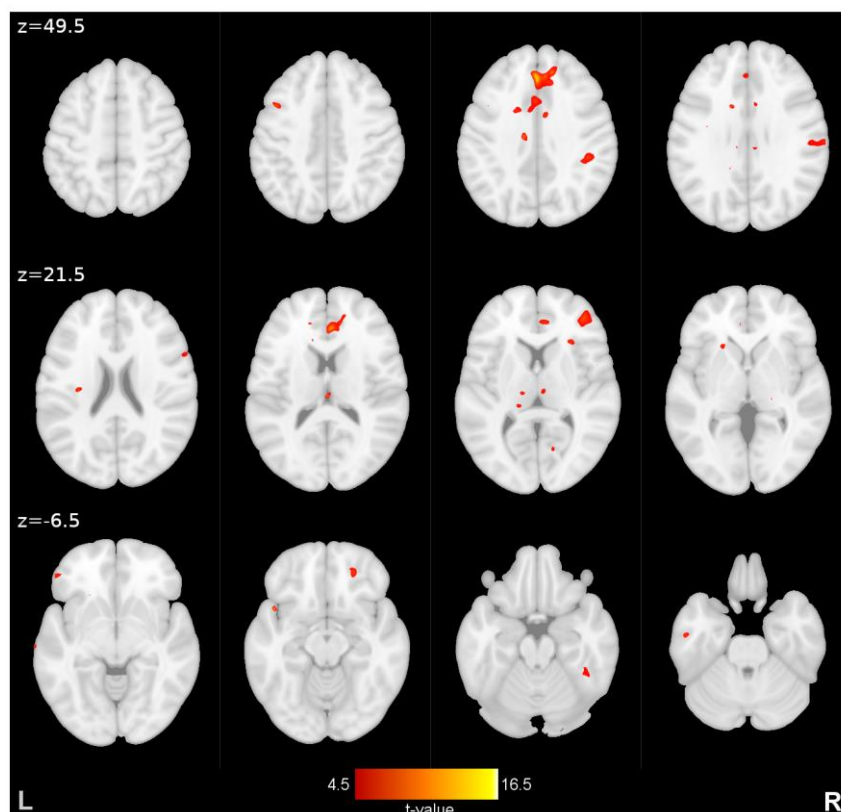
Avoid B > Choose A

Figure 23: Contrast of avoid B over choose A trials. Testing phase. Display threshold set to p<0.001 uncorrected.

5. Discussion

Obesity is a major challenge for global public health, especially in more economically developed countries.

To address this issue, we aimed to better our understanding of the dysregulated decision making and learning processes underlying feeding behaviour in humans. The main purpose of our study was to develop an experiment to evaluate such learning processes. To do so, we designed an experiment that included the PST, an established behavioural task for evaluating learning behaviour, which we modified to assess the effect of nutritional rewards as reinforcers more directly, as opposed to the secondary reinforcers used in previous studies. In addition, we acquired imaging data using a modern multiband technique to gain a better insight into the neurological mechanisms behind the measured behaviour.

5.1 Participants

As we only examined healthy volunteers, without characterising their predisposition for obesity (such as with genetic screening, or with questionnaires to evaluate external or emotional eating tendencies) we could not perform any inter-group analysis. However, the average results and variance of the group could be used as reference values for future studies

5.1.1. Taste test

One of our main questions regarding gustatory stimuli in our experiment was about the stability of their value as reward throughout the experiment. We wished to determine if the value changes during the task and if this would affect performance in the task. Unlike monetary or visual rewards, which are assumed to retain a constant value, we did indeed observe a variable decrease in the subjective value of the nutritional reward. The amount of this decrease influenced performance, when comparing between participants. The greater the reduction of either the liking or wanting ratings, the worse the performance was for the highly rewarding cues in the testing phase. This suggests that a participant whose motivation for receiving reward decreased markedly during the task, also had more difficulties learning from rewarding trials, particularly for the highest rewarding cue A. Participants whose liking of the reward decreased markedly during the task, had more difficulties to differentiate between the values of the more rewarding cues (i.e., during win-win trials). These correlations are in line with our hypotheses and indirectly confirm the validity of the task, showing a higher valuation of reward leads to better results with reward related events.

Surprisingly, participants who reported reduced liking showed a better estimation of the values of the less-rewarding cues, especially of the least rewarding cue B. This may indicate that these participants began to perceive the neutral-tasting solution as salient or even rewarding, and this heightened attention to the outcome of cue B improved their performance on the related trials. This effect is problematic, as it undermines the purpose of the task and needs to be minimized. One possible solution would be to reduce the number of trials in the learning phase, reducing the total amount of reward delivered and thus lowering the risk of developing negative reactions towards it. However, a reduced number of trials would make the task more difficult and a balance between these two factors would need to be found. A second possible solution would be to use an aversive fluid instead of the neutral tasting one. An unpleasant flavour would unlikely become perceived as a reward even in later trials of the task but may shift the neurological reaction from "lack of reward" to "punishment" or to a combined response, which we expect to be different to some extent. Finally, some participants reported developing dryness of the mouth after a certain amount of time due to their lips always being slightly opened and thus likely regarded any form of liquid as positive in later stages of the learning phase. This could possibly be countered by adding a rinse with the tasteless solution to each trial or as an occasional pause in the paradigm. However, the additional time required for such a step needs to be considered.

5.2 Behavioural task

5.2.1. Task performance

Accuracy results from the testing phase show the participants managed to satisfactorily perform on the PST, demonstrating one can use gustatory stimuli successfully as the feedback modality for the paradigm, also when the participants are lying in an MRI machine. The average accuracy for the choose A and avoid B trials was similar to or even higher than studies using the same cue set, despite the participants experiencing fewer learning trials to train on.[80,81] The results were also similar to a study with other cues that can be easily verbally encoded.[89] Furthermore, the participants in our study showed a higher average accuracy than several studies implementing the PST with the more complex hiragana symbols, reflecting the reduced difficulty with our cue set.[38,50,69,79,83,84,88,136] The participants showed the most difficulty in differentiating between less rewarding cues (B/D/F), presenting a significantly lower accuracy as well as a slower response time for lose-lose trials compared to low conflict trials. While the difference in accuracy to the win-win high conflict trials is not statistically significant, there is indeed a trend of win-win trial accuracy being higher than lose-lose accuracy, further supporting the notion that these trials were easier for the participants. Additionally, the

response speed on lose-lose trials was significantly slower, implying the participants needed more time to decide due to the increased challenge. This accuracy trend is opposite from the one reported by Jocham et al.[81], but it is possible that the cohort of that study was compromised mainly of avoidance learners, as opposed to approach learners as in our group, reversing the performance results. However, response times for the three conflict types were similar between that study and ours, which also tested healthy young adults [81]. In another study, healthy seniors generally showed slower responses, but the relationship between response time for low- and high-conflict trials was similar to our own.[83]

Overall, the similarity of results for the behavioural task to previous studies suggests our setup with gustatory stimuli is a viable implementation of the PST, and the modified trial length and number do not impede performance, allowing comparison between studies using the different configurations.

5.2.2.   Computational modelling

The functional structure of the basal ganglia and their Go/NoGo pathways is well established.[44,48] This structure makes distinct learning rates for approach and aversive based learning plausible.

Our data seems to be mostly in line with this hypothesis, with the asymmetrical model best explaining most participants from the group. However, this was not the case for all participants, possibly reflecting the contradiction between these participants performing during the testing phase as approach learners but the calculation of their parameters from both phases better-reflecting avoidance or completely balanced learners.

Previous studies which reported model selection showed mixed results, with some studies finding the asymmetrical model superior[88,89], while others found no advantage for the additional learning rate parameter[79,80]. Importantly, these studies used somewhat weaker calculations, partially using cruder model inversion techniques such as a grid-search-based approach and other model selection criteria such as the Akaike's information criterion or the Bayesian information criterion, which is likely inferior to the variational free energy which we used in our analysis.[137] This may have hindered the success of the model selection for the mentioned studies, so it is possible the asymmetrical model would have been more appropriate for their data as well, when fitted with our method. In addition, in our cohort most of the participants were not fully associated with only a single model and this may have also been the case with previous studies. It is presumably most appropriate to account for this uncertainty and perform random-effects analyses and Bayesian model averaging.

For both asymmetrical and unitary models, we found a clear advantage for using data from both phases as opposed to only using the learning data. The only exception was a participant who showed a strong performance in the learning phase but a weak performance in the testing phase. Thus, the contradiction between the choices he "should" have made according to the learning phase and choices he made in practice in the testing phase likely worsened the model evidence for all models which were fitted on the additional phase. However, this failure of learning is of interest, and it may be prudent to still use the extended model for such participants, as it may more closely reflect the real learning behaviour.

It should be noted that the calculated parameters for the extended models and the shorter models may reflect different mechanisms of learning and so possibly not directly comparable. Frank et al. proposed that the purely learning phase based calculation assesses working memory while the calculation on extended data better reflects habitual learning.[84] As we are more interested in the habitual learning mechanism, this only strengthens the advantage of using the extended models.

While it seems the optimal learning rates for reinforcement tasks are task-specific,[138] for the PST, studies have suggested that a lower α_+ improves performance on the approach-based choose A trials and a lower α_- improves performance on the avoid B trials. According to the hypothesis, lower learning rates allow the slow integration of feedback from multiple trials into the cue's value instead of relying on the most recent outcomes.[84,89] However, our results are not in line with this hypothesis. The suggested correlation between α_- and avoid B trials is poor, while α_+ and choose A show a fair correlation in the opposite direction.

Also unexpected was the moderate correlation between α_+ and avoid B trials as well as the fair correlations between high conflict trials and their opposing learning rates.

It is possible performance on high-conflict trials may benefit from a different strategy than low-conflict trials. A larger opposite learning rate may be beneficial in high-conflict trials by greatly changing the value of a cue according to the rare event (reward for cues B/D/F and omission of reward for cues A/C/E) expanding the value difference between cues where this gap is otherwise subtle. As the lose-lose trials also include half of the avoid B trials, this effect would also influence their results. As the majority of participants in our group were approach learners, this alternative strategy possibly came stronger into play than the low α_- they would otherwise need for the avoidance strategy.

Nevertheless, the suggested correlation between low learning rates and performance does seem to exist for the low conflict trials, which compromise most of the trials in the testing phase and thus had the highest effect on overall task performance. In addition, this effect for the asymmetrical model is for α_+ and not α_-, and this is more pronounced for α_+ than for the single

learning rates from the unitary models, in agreement with the bias of learning type in the group and further supporting the use of the asymmetrical model.

Calculations of state-action values show that according to all models the more rewarding cues were on average estimated quite well by the participants, while less rewarding cues were overestimated. While this is plausible for cue F, as most participants experienced a contingency higher than the planned 40%, this is somewhat surprising for cues B and D, as the experienced contingency for some of the participants was even lower than planned. This overestimation may be an artefact of the computation, as some participants only rarely choose these cues. When combined with a low (negative) learning rate, the posterior value could only be slightly adjusted from the prior value of 0.5. However, this may indeed reflect the true difficulties the participants had with evaluating the less rewarding cues and consequently performing worse on lose-lose and avoid B trials.

5.3 Neuroimaging

5.3.1.   Jitter design efficiency analysis

The jitter design of a task is essential for an effective analysis of fMRI data[93] and our design turned out to be detrimental to our analysis. We made several suboptimal choices for the experiment. Firstly, the choice of contrasts for the original efficiency analysis was not extensive enough and did not allow us to properly choose a jitter design optimized for the contrasts we later focused on. Secondly, we originally decided not to use an explicit jitter between cue presentation and feedback arrival, assuming the participants will jitter naturally by taking a variable decision time from trial to trial. This decision was made to shorten scan time. However, the small variance in this form of jitter proved to be less efficient in separating activations and likely weakened results, especially for the feedback outcomes stimuli and the derived PE, as discussed below.

Finally, we completely randomized a new jitter-list for each participant, which resulted in a variation of efficiency between participants. The analysis showed the efficiency for the same design varied significantly between iterations and it would have been preferable to select the jitter-list(s) from the iteration that showed the best efficiency. In addition, even if the efficiency of each list were equal, using a different list for each participant likely introduced unwanted variability between participants. Using a randomized cue set for each participant likely introduced unwanted variability in a similar manner. On the other hand, using an identical list of jitters and cues for all participants could introduce a bias, so our recommendation for future studies would be to pseudo-randomize and balance a few select lists across participants.

The post hoc jitter design efficiency analysis showed that the specific jitter design we implemented in the experiment for the learning phase had a relatively low efficiency score compared to some of the alternatives and a relevant increase in efficiency would have been easy to implement, in part even without extending the total scan time.

A large boost in efficiency could have been achieved by lengthening the interstimulus interval between cue presentation and the feedback arrival. By splitting the jitters into two separate sets, with half of the time invested in the intertrial intervals and the other half in the interstimulus intervals between the two mentioned stimuli of each trial, efficiency would have been significantly increased even without spending additional scan time. Another, even greater increase in efficiency could have been achieved by lengthening the interstimulus intervals by presenting the cues for the maximal allotted 1.7s response time instead of continuing to feedback delivery immediately when participants chose a cue. This would have likely also reduced unwanted variability for the cue presentation contrast. While this would have required adding time to the experiment (in this group 2.9-5.0 minutes for each participant, depending on their individual speed) this time frame was already calculated for the BOLD sequence and was already taken into account as part of the maximal time participants might spend in the scanner. Implementing both changes would have had an additive effect.

Finally, adding more time to either intertrial intervals, interstimulus intervals or both would have improved efficiency with a cumulative effect proportional to the additional time. While lengthening the trials' duration may have negatively affected participants' attention span, the improved efficiency would have made it plausible to reduce the number of trials and thus keep the experiment length reasonable.

5.3.2.   Learning phase

The presence of the main effects for visual cortex and motor cortex activation at cue presentation (and button click) as expected from the literature[100,101] suggest the acquisition and analyses of the fMRI data were generally successful.

However, the results for feedback outcomes were only partially as expected from the literature. The main effect of reward outcomes showed only weak activity in the sensorimotor cortices without any activity in the gustatory cortices or the ventral striatum.[58,91] The contrast of negative outcomes over positive outcomes was somewhat closer to what we expected, showing some activity in the anterior insula and the anterior cingulate. However, expected activity in the dorsal striatum was missing.[139] The results of the parametric modulators were also lacking, showing no activity in the ventromedial prefrontal cortex for SAV and only weak activity in the dorsolateral prefrontal cortex, with no activity in the striatum or amygdala for PE.[92] These partial results suggest the acquisition of the functional data was considerably suboptimal.

Modifications we made in the behavioural task, such as a smaller set number of trials compared to some of the previous studies could have contributed to this issue by making the task less robust. However, this is unlikely the main cause, as other studies implementing the PST with an even smaller number of trials were successful in observing activity in the reward system.[79]

It is more probable that the use of a suboptimal MRI-sequence led to a drastically worse quality of the functional data. In addition, specifically observation of activity in the ventromedial prefrontal cortex, which was expected for the SAV regressor, was hampered by tilting the field of view. Visually inspecting our data showed this region appears somewhat distorted also after correction, especially compared to other studies on the same MRI-scanner which used an optimized multiband sequence with the field of view parallel to the commissural line. Finally, as previously mentioned, the inefficient jitter design we used also likely contributed to this problem, especially weakening the analysis of the feedback outcomes and the related PE.

5.3.3. Testing phase

The analysis of the testing phase was likely also severely impacted by the suboptimal MRI-sequence. The inefficient jitter design likely also impacted our results, but to a lesser extent, as the trials in this phase did not encompass a second stimulus. For this phase, we also observed activity in the visual cortex for the general stimulus of cue presentation, but the investigated contrasts did not present the activity we expected for the reward system and the prefrontal cortex. However, in this case, it is important to note that the other studies investigating this phase did not show any or showed only very little activity in the ventromedial prefrontal cortex or reward system for participants who were not pharmaceutically manipulated.[79,81] Nevertheless, the contrast of avoid B over choose A trials presented activity in the dorsomedial prefrontal cortex and anterior cingulate cortex, structures which also play a role in decision making and conflict management,[140–142] possibly reflecting the fact that these trials were more difficult for most of the participants.

5.4 Limitations

A major limitation of the analyses of our study was the sample size, which was small in comparison to similar fMRI studies and likely underpowered our statistical tests.

Interpretation of the neuroimaging data was limited by the technical quality of the scans. Visual inspection of the scans aroused suspicions that the parameters of sequence we used for acquired functional data were suboptimal. All scans had quite low spatial resolution and structures of the brain were difficult to identify.

A finger tapping task experiment comparing the sequence we used to one with modified parameters (voxel size $2x2x2\ mm^3$ and TR of $0.81\ ms$) performed on the same MRI machine, showed that optimisation of the sequence parameters resulted in significantly improved results for the same stimuli.

5.5 Conclusions and future prospects

Our study showed that it is possible, using our setup, to conduct the PST whilst using gustatory stimuli as feedback and simultaneously acquiring fMRI data, and that this can be used to investigate learning. This system could prospectively be used to investigate the difference between groups of participants, such as between obese and normal-weight participants or between participants carrying genes predisposing to obesity. It would be of interest to see if gustatory rewards play a role in the dysregulated learning mechanism which are related to obesity, in a manner that is different to other reward modalities.

Several changes and additions could be made to improve interpretation of results in future studies. Extending the characterisation of the participants, both by using focused questionnaires and by relevant genetic characterisation could help differentiating between the various types of learners. Another improvement could be accomplished by adding the neutral-tasting solutions to the on-screen questionnaire in the taste test. Results of this test could aid in ratings-normalisation for the milkshake questionnaire and enable better comparison between participants.

Our computational modelling analysis showed that it is advantageous to compare alternative models using modern techniques such as variational Bayes and that use of observed data from both phases of the PST is superior to the traditional methods used by most previous studies with the PST, which relied only on the data from the learning phase.

Finally, while we did not observe the expected neuronal activity for model-free reinforcement learning in our fMRI analysis, we did manage to observe expected effects of visual and sensorimotor activity, which suggests an improvement of the fMRI settings, jitter design and better correction of artefacts could enable future studies to acquire better results of the activity in the reward system.

The combination of all these elements would hopefully allow a future study to gain important insight into the mechanisms of dysregulated feeding behaviour in obese adults.

6. References

1   Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. *JAMA* 1999; **282:** 1523–29. https://doi.org/10.1001/jama.282.16.1523.

2   World Health Organization. Global Health Observatory. Overweight and obesity. The 2018 update. https://www.who.int/gho/ncd/risk_factors/overweight/en/ (accessed May 04, 2020).

3   Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 2018; **392:** 1923–94. https://doi.org/10.1016/S0140-6736(18)32225-6.

4   Institute for Health Metrics and Evaluation. Findings from the Global Burden of Disease Study 2017. Seattle, WA, 2018.

5   Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet* 2018; **392:** 2052–90. https://doi.org/10.1016/S0140-6736(18)31694-5.

6   OECD. The Heavy Burden of Obesity: The Economics of Prevention. Paris: OECD Publishing, 2019.

7   Owen OE, Smalley KJ, D'Alessio DA, Mozzoli MA, Dawson EK. Protein, fat, and carbohydrate requirements during starvation: anaplerosis and cataplerosis. *Am J Clin Nutr* 1998; **68:** 12–34. https://doi.org/10.1093/ajcn/68.1.12.

8   Coelho M, Oliveira T, Fernandes R. Biochemistry of adipose tissue: an endocrine organ. *Arch Med Sci* 2013; **9:** 191–200. https://doi.org/10.5114/aoms.2013.33181.

9   Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr Today* 2015; **50:** 117–28. https://doi.org/10.1097/NT.0000000000000092.

10 Ode J, Knous J, Schlaff R, Hemenway J, Peterson J, Lowry J. Accuracy of body mass index in volunteer firefighters. *Occup Med (Lond)* 2014; **64:** 193–97. https://doi.org/10.1093/occmed/kqt143.

11 Shah NR, Braverman ER. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin. *PLoS ONE* 2012; **7:** e33308. https://doi.org/10.1371/journal.pone.0033308.

12 Jones LM, Legge M, Goulding A. Healthy body mass index values often underestimate body fat in men with spinal cord injury11No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit

upon the author(s) or upon any organization with which the author(s) is/are associated. *Archives of Physical Medicine and Rehabilitation* 2003; **84:** 1068–71. https://doi.org/10.1016/S0003-9993(03)00045-5.

13 Pischon T, Boeing H, Hoffmann K, et al. General and abdominal adiposity and risk of death in Europe. *N Engl J Med* 2008; **359:** 2105–20. https://doi.org/10.1056/NEJMoa0801891.

14 Schneider HJ, Friedrich N, Klotsche J, et al. The predictive value of different measures of obesity for incident cardiovascular events and mortality. *J Clin Endocrinol Metab* 2010; **95:** 1777–85. https://doi.org/10.1210/jc.2009-1584.

15 Kyle UG, Bosaeus I, Lorenzo AD de, et al. Bioelectrical impedance analysis--part I. review of principles and methods. *Clin Nutr* 2004; **23:** 1226–43. https://doi.org/10.1016/j.clnu.2004.06.004.

16 Kyle UG, Bosaeus I, Lorenzo AD de, et al. Bioelectrical impedance analysis-part II. utilization in clinical practice. *Clin Nutr* 2004; **23:** 1430–53. https://doi.org/10.1016/j.clnu.2004.09.012.

17 Kushner RF, Gudivaka R, Schoeller DA. Clinical characteristics influencing bioelectrical impedance analysis measurements. *Am J Clin Nutr* 1996; **64:** 423S-427S.

18 Matsuzawa Y, Funahashi T, Nakamura T. Molecular mechanism of metabolic syndrome X: contribution of adipocytokines adipocyte-derived bioactive substances. *Ann N Y Acad Sci* 1999; **892:** 146–54. https://doi.org/10.1111/j.1749-6632.1999.tb07793.x.

19 Kern PA, Ranganathan S, Li C, Wood L, Ranganathan G. Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *Am J Physiol Endocrinol Metab* 2001; **280:** E745-51. https://doi.org/10.1152/ajpendo.2001.280.5.E745.

20 Banerjee RR, Rangwala SM, Shapiro JS, et al. Regulation of fasted blood glucose by resistin. *Science* 2004; **303:** 1195–98. https://doi.org/10.1126/science.1092341.

21 Liao C, Gao W, Cao W, et al. Associations of Body Composition Measurements with Serum Lipid, Glucose and Insulin Profile: A Chinese Twin Study. *PLoS ONE* 2015; **10:** e0140595. https://doi.org/10.1371/journal.pone.0140595.

22 Makris MC, Alexandrou A, Papatsoutsos EG, et al. Ghrelin and Obesity: Identifying Gaps and Dispelling Myths. A Reappraisal. *In Vivo* 2017; **31:** 1047–50. https://doi.org/10.21873/invivo.11168.

23 Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985; **28:** 412–19. https://doi.org/10.1007/BF00280883.

24 Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care* 2004; **27:** 1487–95. https://doi.org/10.2337/diacare.27.6.1487.

25 Hall JE, Crook ED, Jones DW, Wofford MR, Dubbert PM. Mechanisms of obesity-associated cardiovascular and renal disease. *Am J Med Sci* 2002; **324:** 127–37. https://doi.org/10.1097/00000441-200209000-00003.

26 Prentice AM. Overeating: the health risks. *Obes Res* 2001; **9 Suppl 4:** 234S-238S. https://doi.org/10.1038/oby.2001.124.

27 Berthoud H-R. Metabolic and hedonic drives in the neural control of appetite: who is the boss? *Curr Opin Neurobiol* 2011; **21:** 888–96. https://doi.org/10.1016/j.conb.2011.09.004.

28 Scarborough P, Burg MR, Foster C, et al. Increased energy intake entirely accounts for increase in body weight in women but not in men in the UK between 1986 and 2000. *Br J Nutr* 2011; **105:** 1399–404. https://doi.org/10.1017/S0007114510005076.

29 Swinburn B, Sacks G, Ravussin E. Increased food energy supply is more than sufficient to explain the US epidemic of obesity. *Am J Clin Nutr* 2009; **90:** 1453–56. https://doi.org/10.3945/ajcn.2009.28595.

30 Urlacher SS, Snodgrass JJ, Dugas LR, et al. Constraint and trade-offs regulate energy expenditure during childhood. *Sci Adv* 2019; **5:** eaax1065. https://doi.org/10.1126/sciadv.aax1065.

31 Saper CB, Chou TC, Elmquist JK. The Need to Feed. *Neuron* 2002; **36:** 199–211. https://doi.org/10.1016/S0896-6273(02)00969-8.

32 Lutter M, Nestler EJ. Homeostatic and hedonic signals interact in the regulation of food intake. *J Nutr* 2009; **139:** 629–32. https://doi.org/10.3945/jn.108.097618.

33 Schultz W. Getting formal with dopamine and reward. *Neuron* 2002; **36:** 241–63.

34 Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology* 2009; **9:** 65–73. https://doi.org/10.1016/j.coph.2008.12.014.

35 Friston K, Kilner J, Harrison L. A free energy principle for the brain. *Journal of Physiology-Paris* 2006; **100:** 70–87. https://doi.org/10.1016/j.jphysparis.2006.10.001.

36 Friston K. The history of the future of the Bayesian brain. *NeuroImage* 2012; **62:** 1230–33. https://doi.org/10.1016/j.neuroimage.2011.10.004.

37 Daw ND, O'Doherty JP. Multiple Systems for Value Learning. In: Neuroeconomics. Elsevier, 2014: 393–410.

38 Coppin G, Nolan-Poupart S, Jones-Gotman M, Small DM. Working memory and reward association learning impairments in obesity. *Neuropsychologia* 2014; **65:** 146–55. https://doi.org/10.1016/j.neuropsychologia.2014.10.004.

39 Horstmann A, Dietrich A, Mathar D, Pössel M, Villringer A, Neumann J. Slave to habit? Obesity is associated with decreased behavioural sensitivity to reward devaluation. *Appetite* 2015; **87:** 175–83. https://doi.org/10.1016/j.appet.2014.12.212.

40 Kube J, Mathar D, Horstmann A, Kotz SA, Villringer A, Neumann J. Altered monetary loss processing and reinforcement-based learning in individuals with obesity. *Brain Imaging and Behavior* 2018; **12:** 1431–49. https://doi.org/10.1007/s11682-017-9786-8.

41 Mathar D, Neumann J, Villringer A, Horstmann A. Failing to learn from negative prediction errors: Obesity is associated with alterations in a fundamental neural learning mechanism. *Cortex* 2017; **95:** 222–37. https://doi.org/10.1016/j.cortex.2017.08.022.

42 Haber S. Neuroanatomy of Reward. In: Gottfried J, ed. Neurobiology of Sensation and Reward. CRC Press, 2011: 235–61.

43 Wise RA. Roles for nigrostriatal--not just mesocorticolimbic--dopamine in reward and addiction. *Trends Neurosci* 2009; **32:** 517–24. https://doi.org/10.1016/j.tins.2009.06.004.

44 Alexander GE, Crutcher MD. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci* 1990; **13:** 266–71. https://doi.org/10.1016/0166-2236(90)90107-L.

45 Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 1998; **80:** 1–27. https://doi.org/10.1152/jn.1998.80.1.1.

46 Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997; **275:** 1593–99. https://doi.org/10.1126/science.275.5306.1593.

47 Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 2003; **299:** 1898–902. https://doi.org/10.1126/science.1077349.

48 Frank MJ. Dynamic dopamine modulation in the basal ganglia. a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* 2005; **17:** 51–72. https://doi.org/10.1162/0898929052880093.

49 Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. *Nat Rev Neurosci* 2006; **7:** 464–76. https://doi.org/10.1038/nrn1919.

50 Cox SML, Frank MJ, Larcher K, et al. Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *NeuroImage* 2015; **109:** 95–101. https://doi.org/10.1016/j.neuroimage.2014.12.070.

51 Mathar D, Wilkinson L, Holl AK, et al. The role of dopamine in positive and negative prediction error utilization during incidental learning - Insights from Positron Emission Tomography, Parkinson's disease and Huntington's disease. *Cortex* 2017; **90:** 149–62. https://doi.org/10.1016/j.cortex.2016.09.004.

52 Vandecasteele M, Glowinski J, Venance L. Electrical synapses between dopaminergic neurons of the substantia nigra pars compacta. *J. Neurosci.* 2005; **25:** 291–98. https://doi.org/10.1523/JNEUROSCI.4167-04.2005.

53 Allison DW, Ohran AJ, Stobbs SH, et al. Connexin-36 gap junctions mediate electrical coupling between ventral tegmental area GABA neurons. *Synapse* 2006; **60:** 20–31. https://doi.org/10.1002/syn.20272.

54 Grace AA, Bunney BS. Intracellular and extracellular electrophysiology of nigral dopaminergic neurons—3. Evidence for electrotonic coupling. *Neuroscience* 1983; **10:** 333–48. https://doi.org/10.1016/0306-4522(83)90137-9.

55 Daw ND, Tobler PN. Value Learning through Reinforcement. In: Neuroeconomics. Elsevier, 2014: 283–98.

56 Abler B, Walter H, Erk S, Kammerer H, Spitzer M. Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage* 2006; **31:** 790–95. https://doi.org/10.1016/j.neuroimage.2006.01.001.

57 Schultz W. Dopamine signals for reward value and risk: basic and recent data. *Behav Brain Funct* 2010; **6:** 24. https://doi.org/10.1186/1744-9081-6-24.

58 Thanarajah SE, Backes H, DiFeliceantonio AG, et al. Food Intake Recruits Orosensory and Post-ingestive Dopaminergic Circuits to Affect Eating Desire in Humans. *Cell Metab* 2019; **29:** 695-706.e4. https://doi.org/10.1016/j.cmet.2018.12.006.

59 Carr KD. Feeding, drug abuse, and the sensitization of reward by metabolic need. *Neurochem Res* 1996; **21:** 1455–67. https://doi.org/10.1007/BF02532386.

60 Cabeza de Vaca S, Carr KD. Food Restriction Enhances the Central Rewarding Effect of Abused Drugs. *J. Neurosci.* 1998; **18:** 7502–10. https://doi.org/10.1523/JNEUROSCI.18-18-07502.1998.

61 Shalev U, Yap J, Shaham Y. Leptin Attenuates Acute Food Deprivation-Induced Relapse to Heroin Seeking. *J. Neurosci.* 2001; **21:** RC129-RC129. https://doi.org/10.1523/JNEUROSCI.21-04-j0001.2001.

62 Stratford TR, Kelley AE. Evidence of a Functional Relationship between the Nucleus Accumbens Shell and Lateral Hypothalamus Subserving the Control of Feeding Behavior. *J. Neurosci.* 1999; **19:** 11040–48. https://doi.org/10.1523/JNEUROSCI.19-24-11040.1999.

63 Heyes CM. Social learning in animals: categories and mechanisms. *Biological Reviews* 1994; **69:** 207–31. https://doi.org/10.1111/j.1469-185X.1994.tb01506.x.

64 Cartoni E, Balleine B, Baldassarre G. Appetitive Pavlovian-instrumental Transfer: A review. *Neurosci Biobehav Rev* 2016; **71:** 829–48. https://doi.org/10.1016/j.neubiorev.2016.09.020.

65 Dolan RJ, Dayan P. Goals and habits in the brain. *Neuron* 2013; **80:** 312–25. https://doi.org/10.1016/j.neuron.2013.09.007.

66 Volkow ND, Wang G-J, Baler RD. Reward, dopamine and the control of food intake: implications for obesity. *Trends Cogn Sci (Regul Ed)* 2011; **15:** 37–46. https://doi.org/10.1016/j.tics.2010.11.001.

67 Smith AC, Frank LM, Wirth S, et al. Dynamic analysis of learning in behavioral experiments. *J. Neurosci.* 2004; **24:** 447–61. https://doi.org/10.1523/JNEUROSCI.2908-03.2004.

68 Glautier S. Revisiting the learning curve (once again). *Front. Psychol.* 2013; **4:** 982. https://doi.org/10.3389/fpsyg.2013.00982.

69 Frank MJ, Seeberger LC, O'Reilly R C. By carrot or by stick. cognitive reinforcement learning in parkinsonism. *Science* 2004; **306:** 1940–43. https://doi.org/10.1126/science.1102941.

70 Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci U S A* 2004; **101:** 13124–31. https://doi.org/10.1073/pnas.0404965101.

71 Collins AGE, Shenhav A. Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacol.* 2021: 1–15. https://doi.org/10.1038/s41386-021-01126-y.

72 Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage* 2009; **45:** S173-86. https://doi.org/10.1016/j.neuroimage.2008.10.055.

73 Collins AGE, Cockburn J. Beyond dichotomies in reinforcement learning. *Nat Rev Neurosci* 2020. https://doi.org/10.1038/s41583-020-0355-6.

74 Daw ND. Advanced Reinforcement Learning. In: Neuroeconomics. Elsevier, 2014: 299–320.

75 Daunizeau J, Adam V, Rigoux L. VBA. A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLOS Computational Biology* 2014; **10:** e1003441. https://doi.org/10.1371/journal.pcbi.1003441.

76 Nummenmaa A, Auranen T, Hämäläinen MS, et al. Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *NeuroImage* 2007; **35:** 669–85. https://doi.org/10.1016/j.neuroimage.2006.05.001.

77 Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. *NeuroImage* 2014; **84:** 971–85. https://doi.org/10.1016/j.neuroimage.2013.08.065.

78 West R, Huet A. The Effect of Aging on the ERP Correlates of Feedback Processing in the Probabilistic Selection Task. *Brain Sci* 2020; **10. https://doi.org/**10.3390/brainsci10010040.

79 Shiner T, Seymour B, Wunderlich K, et al. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain* 2012; **135:** 1871–83. https://doi.org/10.1093/brain/aws083.

80 Sevgi M, Rigoux L, Kuhn AB, et al. An Obesity-Predisposing Variant of the FTO Gene Regulates D2R-Dependent Reward Learning. *J Neurosci* 2015; **35:** 12584–92. https://doi.org/10.1523/JNEUROSCI.1589-15.2015.

81 Jocham G, Klein TA, Ullsperger M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci* 2011; **31:** 1606–13. https://doi.org/10.1523/JNEUROSCI.3904-10.2011.

82 Solomon M, Frank MJ, Ragland JD, et al. Feedback-driven trial-by-trial learning in autism spectrum disorders. *Am J Psychiatry* 2015; **172:** 173–81. https://doi.org/10.1176/appi.ajp.2014.14010036.

83 Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science* 2007; **318:** 1309–12. https://doi.org/10.1126/science.1146157.

84 Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A* 2007; **104:** 16311–16. https://doi.org/10.1073/pnas.0706111104.

85 Batouli SAH, Sisakhti M. Some Points to Consider in a Task-Based fMRI Study: A Guideline for Beginners. *fbt* 2020. https://doi.org/10.18502/fbt.v7i1.2725.

86 Schutte I, Slagter HA, Collins AGE, Frank MJ, Kenemans JL. Stimulus discriminability may bias value-based probabilistic learning. *PLoS ONE* 2017; **12:** e0176205. https://doi.org/10.1371/journal.pone.0176205.

87 Frank MJ, Rudy JW, Levy WB, O'Reilly RC. When logic fails: implicit transitive inference in humans. *Mem Cognit* 2005; **33:** 742–50. https://doi.org/10.3758/bf03195340.

88 Grogan JP, Tsivos D, Smith L, et al. Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease. *eLife* 2017; **6. https://doi.org/**10.7554/eLife.26801.

89 Aberg KC, Doell KC, Schwartz S. Correction: Linking Individual Learning Styles to Approach-Avoidance Motivational Traits and Computational Aspects of Reinforcement Learning. *PLoS ONE* 2017; **12:** e0172379. https://doi.org/10.1371/journal.pone.0172379.

90 Sescousse G, Caldú X, Segura B, Dreher J-C. Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neurosci Biobehav Rev* 2013; **37:** 681–96. https://doi.org/10.1016/j.neubiorev.2013.02.002.

91 O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ. Neural Responses during Anticipation of a Primary Taste Reward. *Neuron* 2002; **33:** 815–26. https://doi.org/10.1016/S0896-6273(02)00603-7.

92   Chase HW, Kumar P, Eickhoff SB, Dombrovski AY. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cogn Affect Behav Neurosci* 2015; **15:** 435–59. https://doi.org/10.3758/s13415-015-0338-7.

93   Amaro E, Barker GJ. Study design in fMRI: basic principles. *Brain Cogn* 2006; **60:** 220–32. https://doi.org/10.1016/j.bandc.2005.11.009.

94   Henson RN. Efficient Experimental Design for fMRI. In: Friston KJ, ed. Statistical parametric mapping. the analysis of functional brain images. Amsterdam, Boston: Elsevier / Academic Press, 2007: 193–210.

95   Crosson B, Ford A, McGregor KM, et al. Functional imaging and related techniques: an introduction for rehabilitation researchers. *J Rehabil Res Dev* 2010; **47:** vii–xxxiv. https://doi.org/10.1682/jrrd.2010.02.0017.

96   Scarapicchia V, Brown C, Mayo C, Gawryluk JR. Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies. *Front. Hum. Neurosci.* 2017; **11:** 419. https://doi.org/10.3389/fnhum.2017.00419.

97   Lystad RP, Pollard H. Functional neuroimaging: a brief overview and feasibility for use in chiropractic research. *J Can Chiropr Assoc* 2009; **53:** 59–72.

98   McRobbie DW. MRI from picture to proton. Cambridge: Cambridge University Press, 2007.

99   Ogawa S, Lee TM, Kay AR, Tank DW. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A* 1990; **87:** 9868–72. https://doi.org/10.1073/pnas.87.24.9868.

100   Ogawa S, Tank DW, Menon R, et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A* 1992; **89:** 5951–55. https://doi.org/10.1073/pnas.89.13.5951.

101   Kwong KK, Belliveau JW, Chesler DA, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U S A* 1992; **89:** 5675–79. https://doi.org/10.1073/pnas.89.12.5675.

102   Fox PT, Raichle ME. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc Natl Acad Sci U S A* 1986; **83:** 1140–44. https://doi.org/10.1073/pnas.83.4.1140.

103   Lindquist MA. The Statistical Analysis of fMRI Data. *ss* 2008; **23:** 439–64. https://doi.org/10.1214/09-STS282.

104   Barth M, Breuer F, Koopmans PJ, Norris DG, Poser BA. Simultaneous multislice (SMS) imaging techniques. *Magn Reson Med* 2016; **75:** 63–81. https://doi.org/10.1002/mrm.25897.

105   Demetriou L, Kowalczyk OS, Tyson G, Bello T, Newbould RD, Wall MB. A comprehensive evaluation of increasing temporal resolution with multiband-accelerated

protocols and effects on statistical outcome measures in fMRI. *NeuroImage* 2018; **176:** 404–16. https://doi.org/10.1016/j.neuroimage.2018.05.011.

106   Todd N, Josephs O, Zeidman P, Flandin G, Moeller S, Weiskopf N. Functional Sensitivity of 2D Simultaneous Multi-Slice Echo-Planar Imaging: Effects of Acceleration on g-factor and Physiological Noise. *Front Neurosci* 2017; **11:** 158. https://doi.org/10.3389/fnins.2017.00158.

107   Chen JE, Glover GH. Functional Magnetic Resonance Imaging Methods. *Neuropsychol Rev* 2015; **25:** 289–313. https://doi.org/10.1007/s11065-015-9294-9.

108   Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* 1994; **2:** 189–210. https://doi.org/10.1002/hbm.460020402.

109   Wang GJ, Volkow ND, Logan J, et al. Brain dopamine and obesity. *Lancet* 2001; **357:** 354–57.

110   Burger KS, Stice E. Greater striatopallidal adaptive coding during cue-reward learning and food reward habituation predict future weight gain. *NeuroImage* 2014; **99:** 122–28. https://doi.org/10.1016/j.neuroimage.2014.05.066.

111   Stice E, Spoor S, Bohon C, Veldhuizen MG, Small DM. Relation of reward from food intake and anticipated food intake to obesity: a functional magnetic resonance imaging study. *J Abnorm Psychol* 2008; **117:** 924–35. https://doi.org/10.1037/a0013600.

112   Kroemer NB, Sun X, Veldhuizen MG, Babbs AE, Araujo IE de, Small DM. Weighing the evidence: Variance in brain responses to milkshake receipt is predictive of eating behavior. *NeuroImage* 2016; **128:** 273–83. https://doi.org/10.1016/j.neuroimage.2015.12.031.

113   Babbs RK, Sun X, Felsted J, Chouinard-Decorte F, Veldhuizen MG, Small DM. Decreased caudate response to milkshake is associated with higher body mass index and greater impulsivity. *Physiol Behav* 2013; **121:** 103–11. https://doi.org/10.1016/j.physbeh.2013.03.025.

114   Dreher JC, Schmidt PJ, Kohn P, Furman D, Rubinow D, Berman KF. Menstrual cycle phase modulates reward-related neural function in women. *Proc Natl Acad Sci U S A* 2007; **104:** 2465–70. https://doi.org/10.1073/pnas.0605569104.

115   Creutz LM, Kritzer MF. Mesostriatal and mesolimbic projections of midbrain neurons immunoreactive for estrogen receptor beta or androgen receptors in rats. *J Comp Neurol* 2004; **476:** 348–62. https://doi.org/10.1002/cne.20229.

116   Becker JB, Robinson TE, Lorenz KA. Sex difference and estrous cycle variations in amphetamine-elicited rotational behavior. *European Journal of Pharmacology* 1982; **80:** 65–72. https://doi.org/10.1016/0014-2999(82)90178-9.

117   Becker JB, Cha J-H. Estrous cycle-dependent variation in amphetamine-induced behaviors and striatal dopamine release assessed with microdialysis. *Behavioural Brain Research* 1989; **35:** 117–25. https://doi.org/10.1016/S0166-4328(89)80112-3.

118   DiFeliceantonio AG, Coppin G, Rigoux L, et al. Supra-Additive Effects of Combining Fat and Carbohydrate on Food Reward. *Cell Metab* 2018; **28:** 33-44.e3. https://doi.org/10.1016/j.cmet.2018.05.018.

119   Bartoshuk LM, Duffy VB, Green BG, et al. Valid across-group comparisons with labeled scales: the gLMS versus magnitude matching. *Physiol Behav* 2004; **82:** 109–14. https://doi.org/10.1016/j.physbeh.2004.02.033.

120   Green BG, Dalton P, Cowart B, Shaffer G, Rankin K, Higgins J. Evaluating the 'Labeled Magnitude Scale' for measuring sensations of taste and smell. *Chem Senses* 1996; **21:** 323–34. https://doi.org/10.1093/chemse/21.3.323.

121   Pelli DG. The VideoToolbox software for visual psychophysics. transforming numbers into movies. *Spat Vis* 1997; **10:** 437–42.

122   Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C. What's new in psychtoolbox-3. *Perception* 2007; **36:** 1–16.

123   Brainard DH. The Psychophysics Toolbox. *Spat Vis* 1997; **10:** 433–36.

124   Veldhuizen MG, Bender G, Constable RT, Small DM. Trying to detect taste in a tasteless solution: modulation of early gustatory cortex by attention to taste. *Chem Senses* 2007; **32:** 569–81. https://doi.org/10.1093/chemse/bjm025.

125   Katehakis MN, Veinott AF. The Multi-Armed Bandit Problem: Decomposition and Computation. *Mathematics of OR* 1987; **12:** 262–68. https://doi.org/10.1287/moor.12.2.262.

126   Sutton RS, Barto AG. Reinforcement learning. An introduction. Cambridge, Mass: MIT Press, 1998.

127   Daunizeau J, Preuschoff K, Friston K, Stephan K. Optimizing experimental design for comparing models of brain function. *PLOS Computational Biology* 2011; **7:** e1002280. https://doi.org/10.1371/journal.pcbi.1002280.

128   Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 2004; **23 Suppl 1:** S208-19. https://doi.org/10.1016/j.neuroimage.2004.07.051.

129   Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage* 2012; **62:** 782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.

130   Friston KJ. Statistical Parametric Mapping. The Analysis of Functional Brain Images. Burlington: Elsevier Science, 2011.

131   Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* 2002; **17:** 825–41. https://doi.org/10.1006/nimg.2002.1132.

132   Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 2003; **20:** 870–88. https://doi.org/10.1016/S1053-8119(03)00336-7.

133   Caballero-Gaudes C, Reynolds RC. Methods for cleaning the BOLD fMRI signal. *NeuroImage* 2017; **154:** 128–49. https://doi.org/10.1016/j.neuroimage.2016.12.018.

134   Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 2011; **59:** 2142–54. https://doi.org/10.1016/j.neuroimage.2011.10.018.

135   Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 2014; **84:** 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

136   Doll BB, Jacobs WJ, Sanfey AG, Frank MJ. Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Res* 2009; **1299:** 74–94. https://doi.org/10.1016/j.brainres.2009.07.007.

137   Penny WD. Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage* 2012; **59:** 319–30. https://doi.org/10.1016/j.neuroimage.2011.07.039.

138   Nussenbaum K, Hartley CA. Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience* 2019; **40:** 100733. https://doi.org/10.1016/j.dcn.2019.100733.

139   Palminteri S, Justo D, Jauffret C, et al. Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron* 2012; **76:** 998–1009. https://doi.org/10.1016/j.neuron.2012.10.017.

140   Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth MFS. Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 2006; **9:** 940–47. https://doi.org/10.1038/nn1724.

141   Umemoto A, HajiHosseini A, Yates ME, Holroyd CB. Reward-based contextual learning supported by anterior cingulate cortex. *Cogn Affect Behav Neurosci* 2017; **17:** 642–51. https://doi.org/10.3758/s13415-017-0502-3.

142   Venkatraman V, Rosati AG, Taren AA, Huettel SA. Resolving response, decision, and strategic control: evidence for a functional topography in dorsomedial prefrontal cortex. *J. Neurosci.* 2009; **29:** 13158–64. https://doi.org/10.1523/JNEUROSCI.2708-09.2009.

7. Appendix

7.1 List of Figures

7.2 List of Tables

7.3 List of Equations

7.4 Supplementary Material

7.4.1.   Digital resources

The scripts used in this study and the raw data acquired from the behavioural task are available on my GitHub repository: https://github.com/oreiner/FPST

The following software programs and packages were used in the implementation and analysis of the scientific experiment described in this dissertation:

(1)   MATLAB. Version R2014b. Natick, Massachusetts: The MathWorks Inc.
       https://de.mathworks.com/products/matlab.html

(2)   Psychtoolbox function library for MATLAB. Version 3.0.13.[121–123]
       http://psychtoolbox.org

(3)   VBA-Toolbox package for MATLAB[75] https://mbb-team.github.io/VBA-toolbox/

(4)   SPM Package for MATLAB. Version 12, Wellcome Trust Centre for Neuroimaging[108,130]
       http://www.fil.ion.ucl.ac.uk/spm

(5)   FMRIB Software Library (FSL). Version 5.0.11.[72,128,129]
       https://fsl.fmrib.ox.ac.uk/fsl/fslwiki

(6)   MCFLIRT tool for FSL.[131]

(7)   Topup tool for FSL.[132]

7.4.2.   Recipes

Table 4: Milkshake recipes from which the participants could choose the flavour.

| Commercial milkshake powder (Flavours: Chocolate, Vanilla, Banana, Strawberry)	12g
Full cream milk	170g
Whipping cream	30g

Table 5: Recipes for "tasteless" solutions from which the participants could choose.

	NaHCO3	KCl
Full concentration	2.5 mM	25 mM
75%	1.875 mM	18.75 mM
50%	1.25 mM	12.5 mM
25%	0.625 mM	6.25 mM

7.4.3. Computational modelling - model comparison

Bayesian model selection

Figure 24: Comparison of competing models for the Q-Learning algorithm
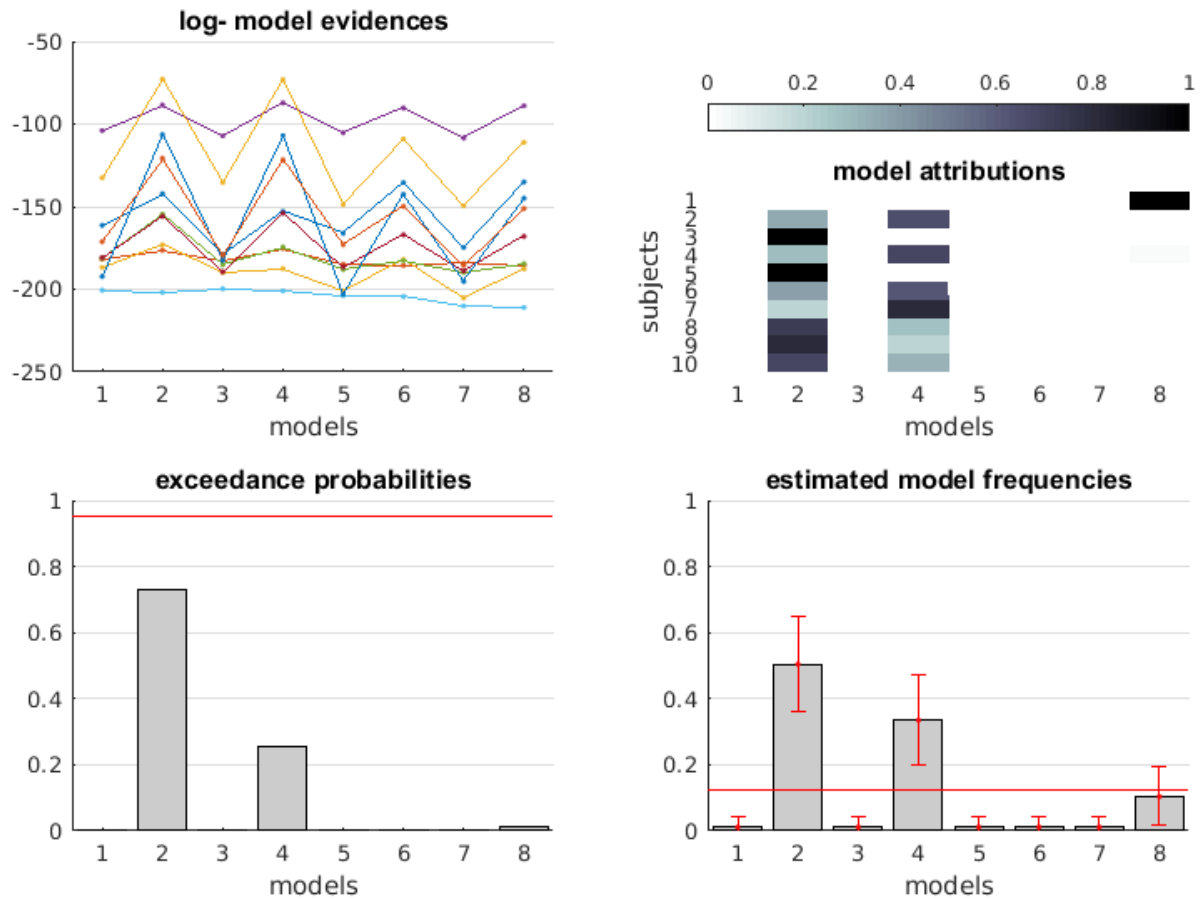
Legend: design number. data input y, learning rate α, temperature β

1. both phases, asymmetrical, fixed
2. both phases, asymmetrical, individually fitted
3. both phases, uniform, fixed
4. both phases, uniform, individually fitted
5. only learn phase, asymmetrical, fixed
6. only learn phase, asymmetrical, individually fitted
7. only learn phase, uniform, fixed
8. only learn phase, uniform, individually

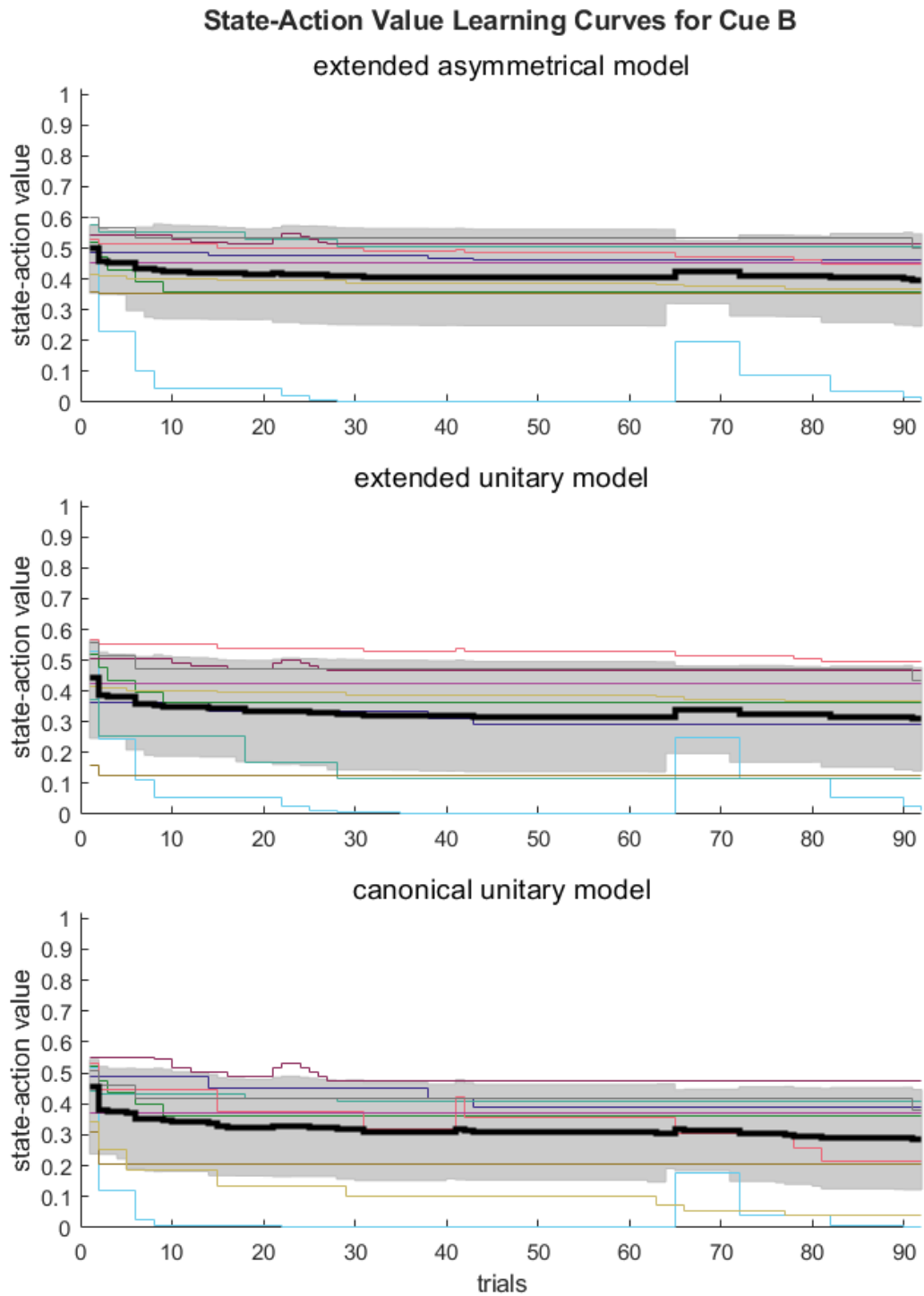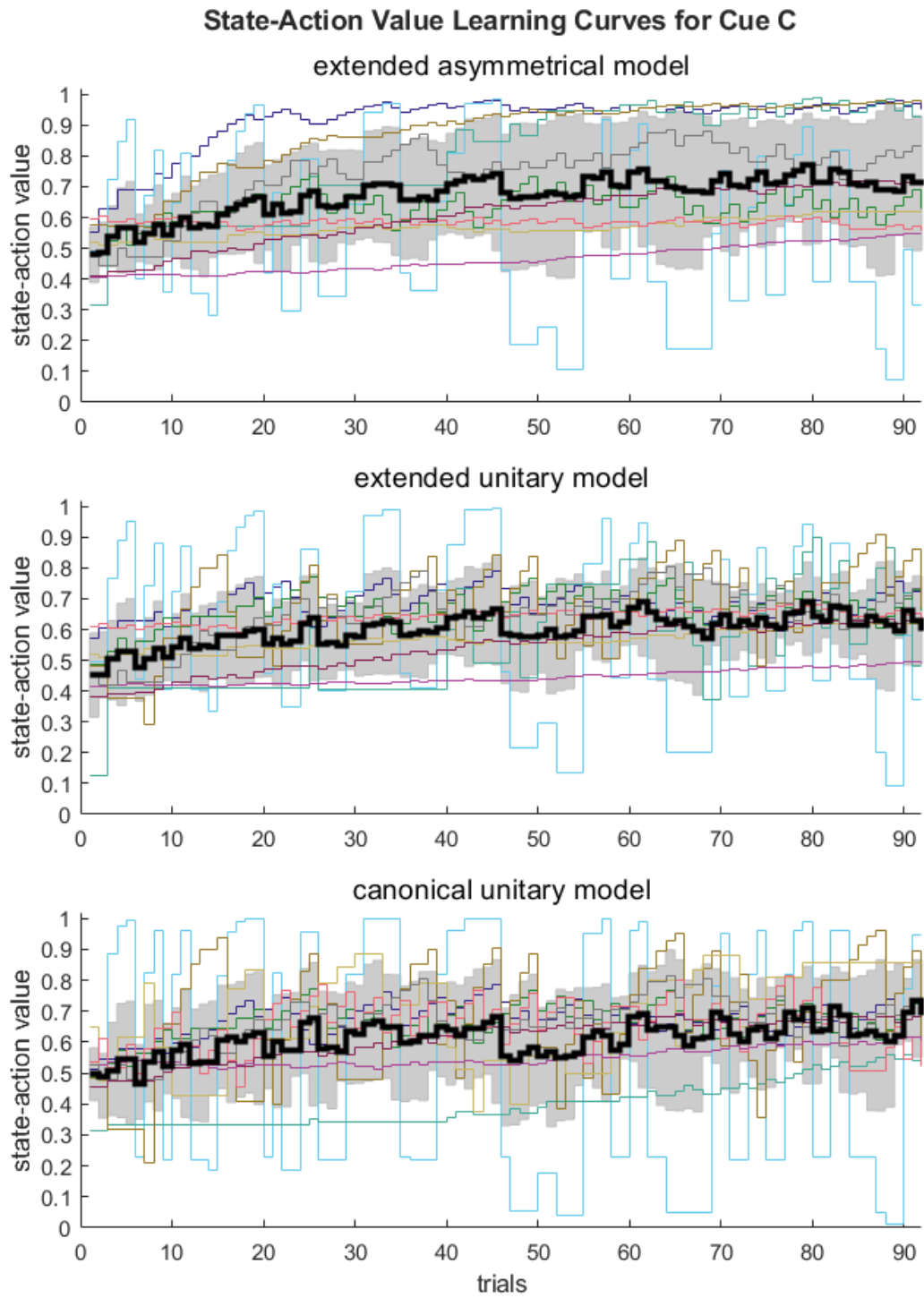**State-action values learning curves**
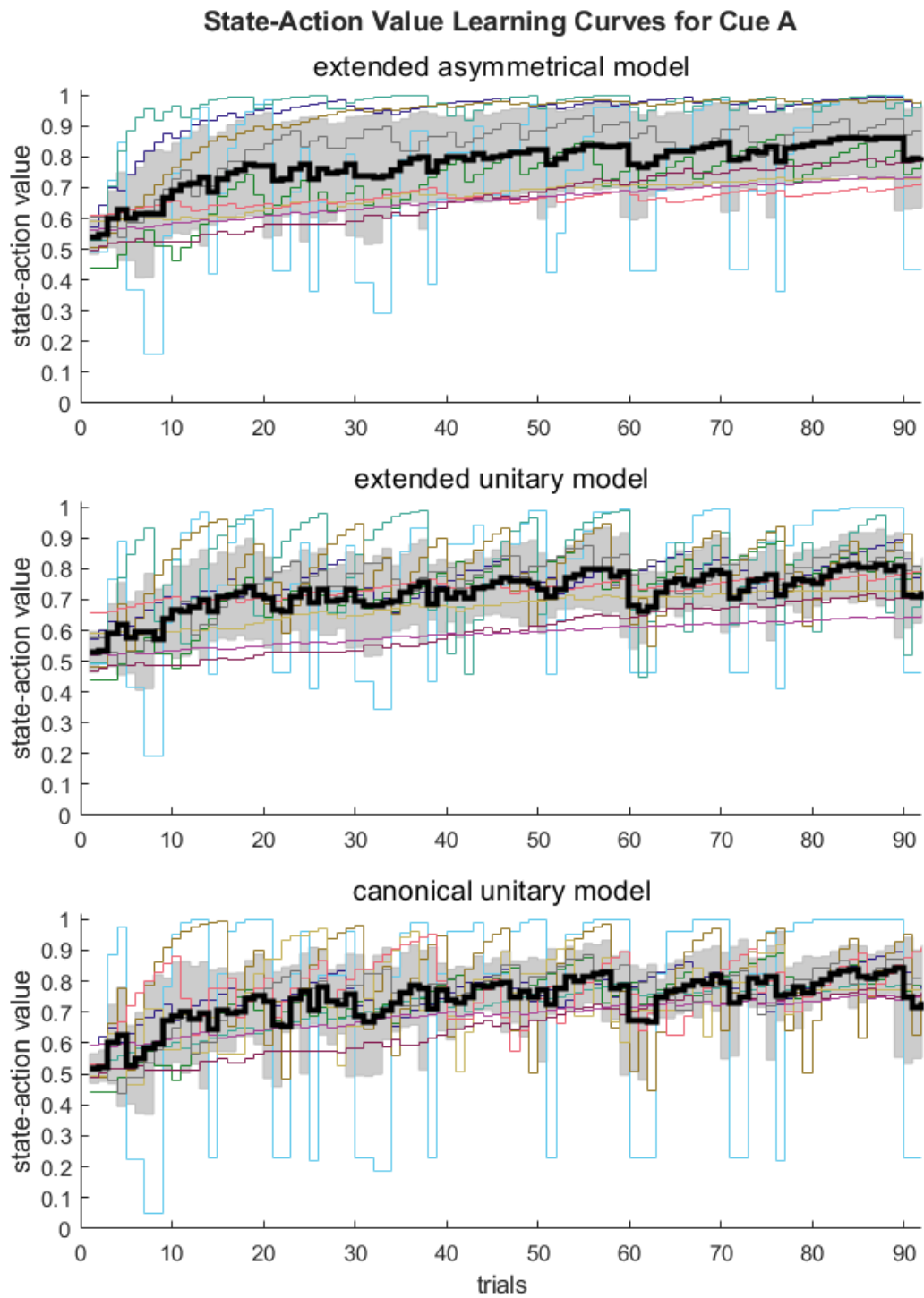
Figure 25: SAV for cue A as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.

Figure 26: SAV for cue B as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.
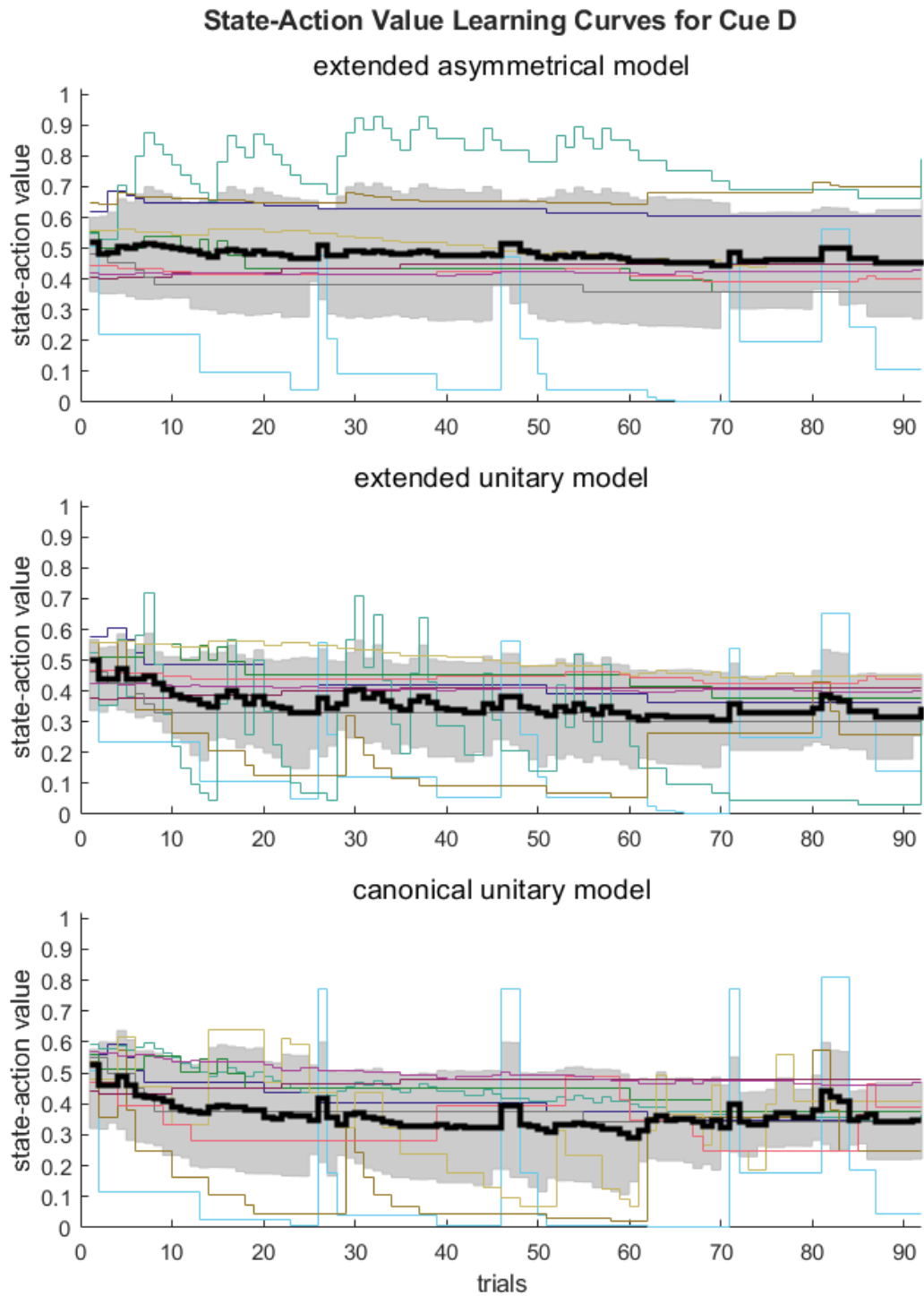
Figure 27: SAV for cue C as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.
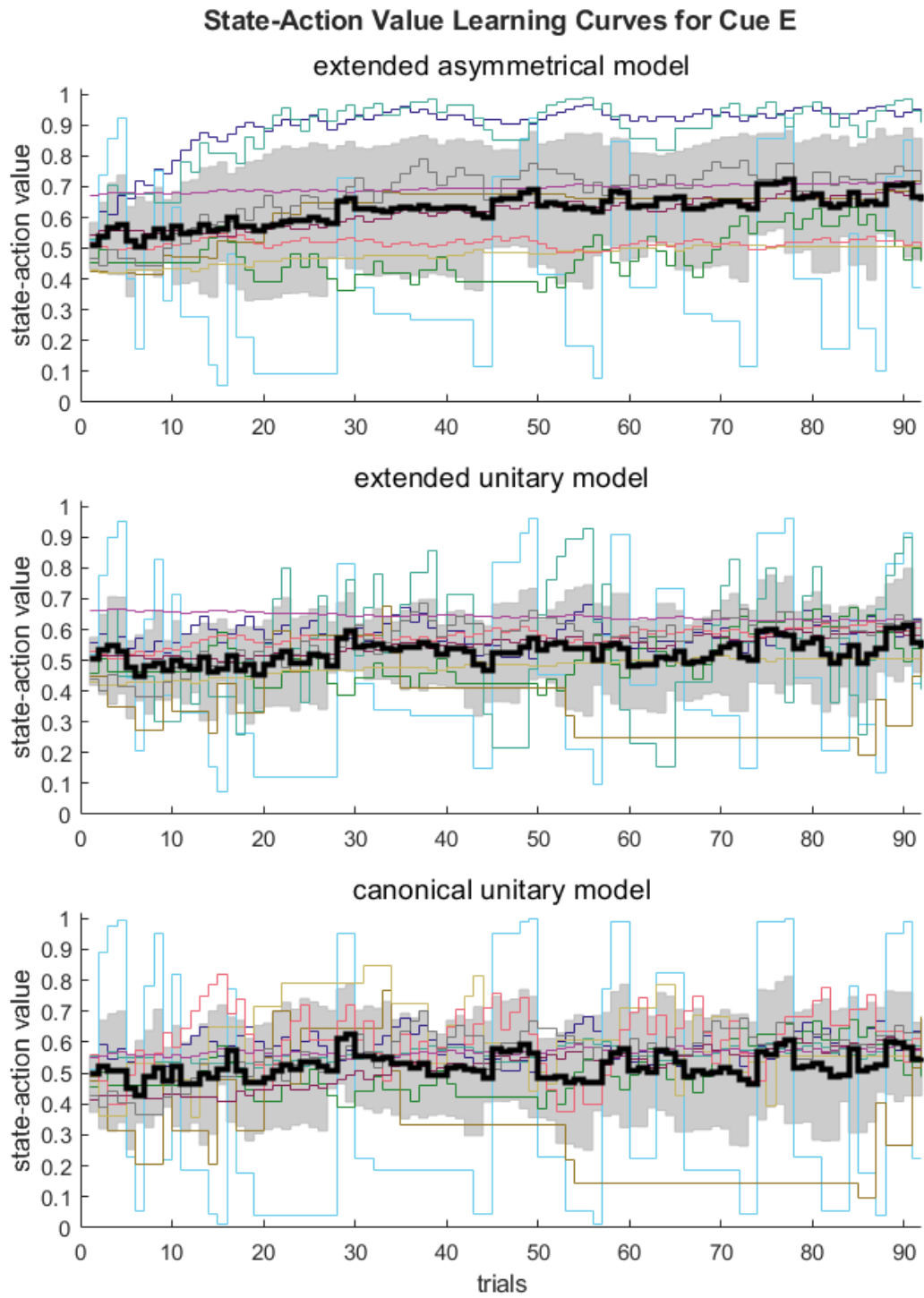
Figure 28: SAV for cue D as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.

Figure 29: SAV for cue E as calculated in three different models. Each individual participant's SAV as trials progressed in the learning phase is shown as an individual line. Mean SAV is shown as a bold line ± standard deviation as a grey band.

7.4.4. Jitter design efficiency analysis

The design efficiency analyses compared the efficiency of different jitter distributions, either with a uniform distribution defined between a minimal and maximal value, or with a truncated exponential distribution defined with a minimal value and a rate parameter.

Original analysis

In the analysis performed in preparation for the study, the data set for choices was acquired from a pilot participant, who performed a learning phase composed of 240 trials and received feedback in the form of on-screen cues: a drawing of a smiley-face as a positive outcome and a frowney-face as a negative outcome. Response times were simulated from a uniform distribution between $0.5\ s$ and $1.5\ s$.

Only the contrast positive PE > negative PE was analysed. The other contrasts (cue presentation, SAV of chosen cue, positive feedback, positive PE) and several additional designs (N° 8,9,10,12) that were tested in the post hoc test, were added here for context.

For the designs tested, refer to figure 30.

post hoc design efficiency analysis

For the post hoc analysis the 10 choice sets acquired from the participants in the study were used. The efficiency of the jitter distribution with the real response times was compared to other designs with simulated response times sampled from a normal distribution with a mean of $1\ s$ and a standard deviation of $0.2\ s$.

For the designs tested, refer to figure 31 & figure 32.

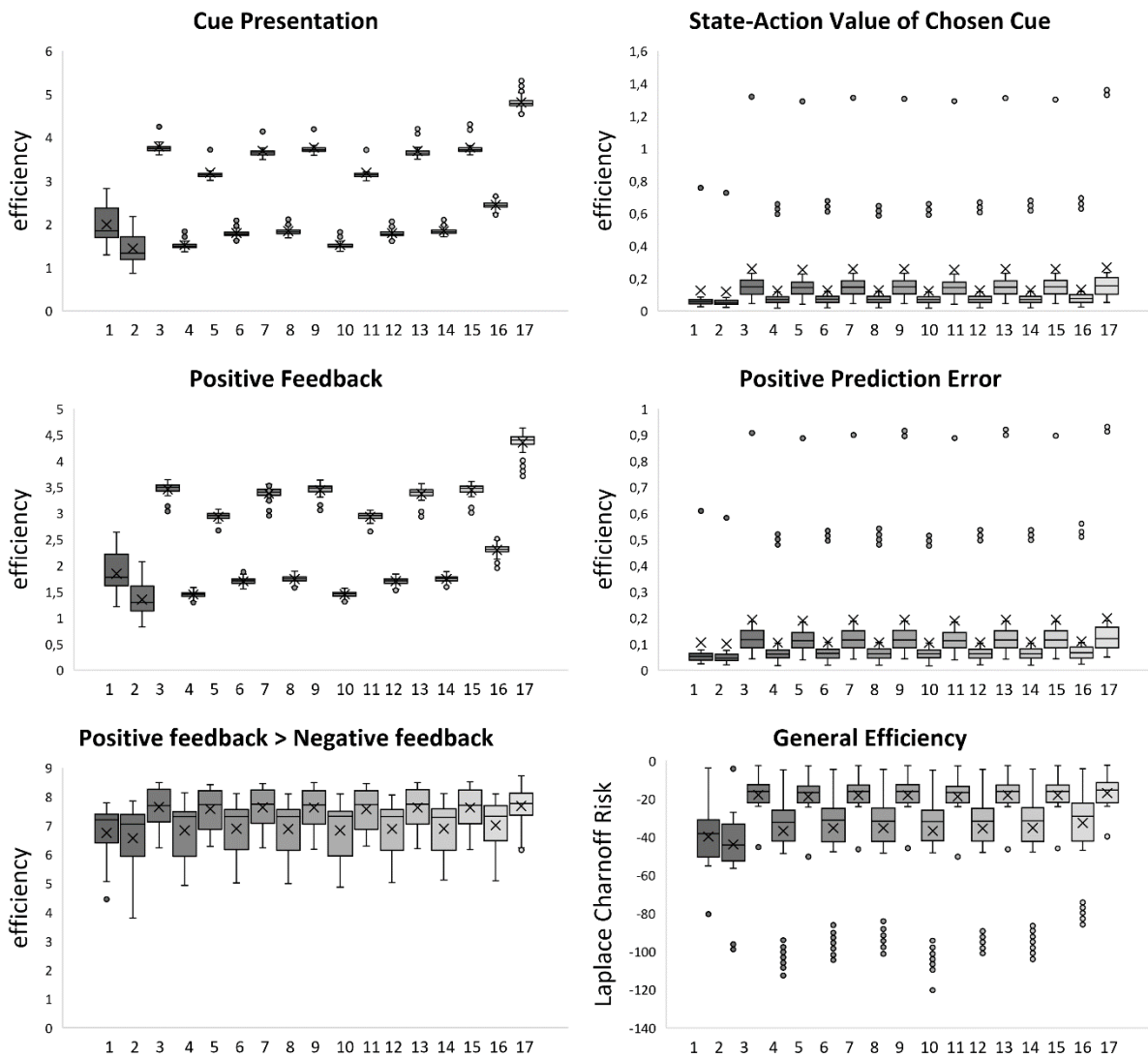*Figure 30: Jitter design efficiency - original analysis*
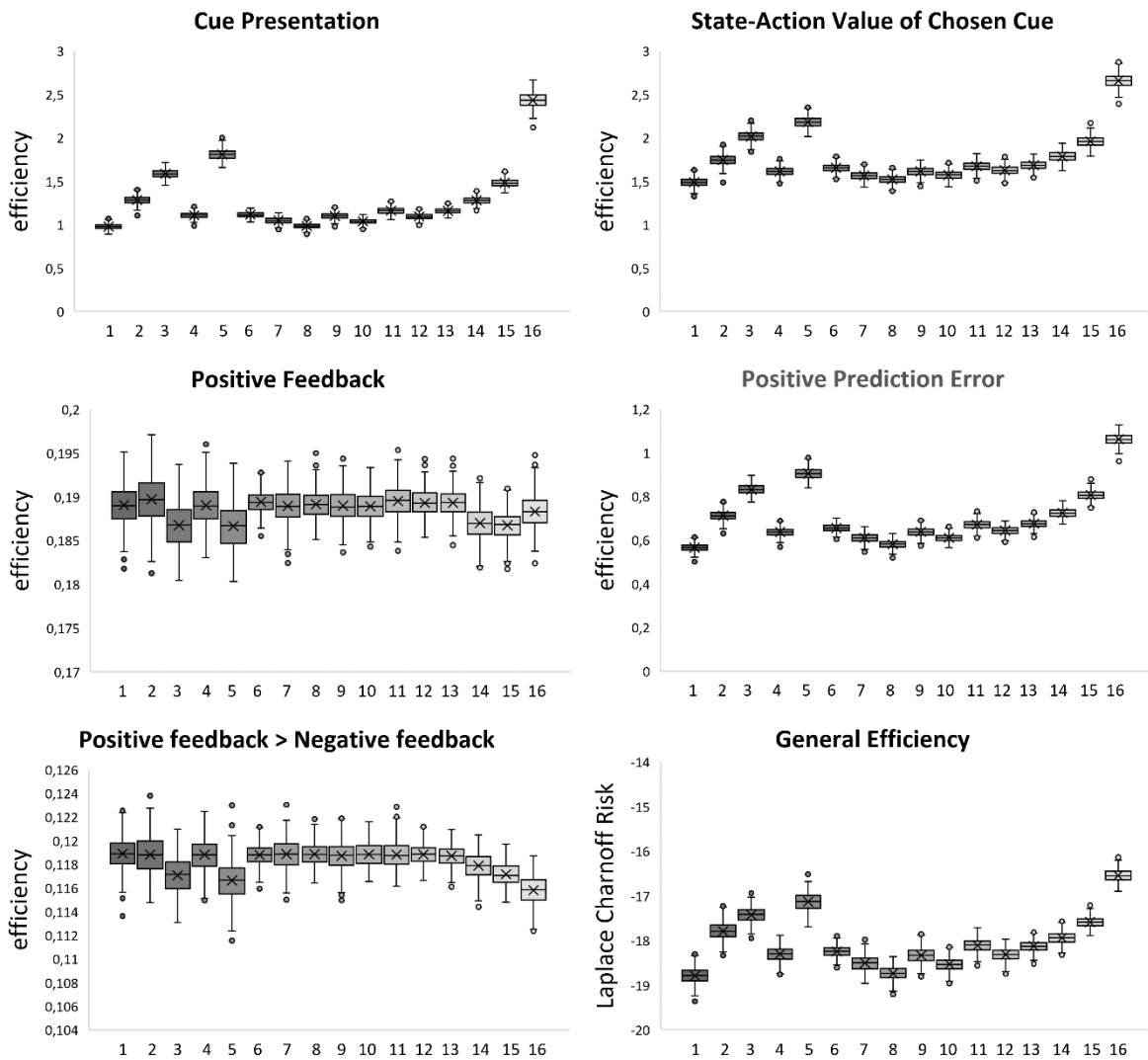
1. uniformly distributed 0-1.2s with simulated response time
2. uniformly distributed 0-2.0s with simulated response time
3. uniformly distributed 0.5-2.5s with simulated response time
4. uniformly distributed 0.2-1.5s with simulated response time
5. uniformly distributed 0.5-3.0s with simulated response time
6. fixed at 1.0s with simulated response time
7. truncated exponential distribution (m=0.5; μ=0.3) with simulated response time
8. truncated exponential distribution (m=0.6; μ=0.15) with simulated response time
9. truncated exponential distribution (m=0.6; μ=0.3) with simulated response time
10. truncated exponential distribution (m=0.7; μ=0.15) with simulated response time
11. truncated exponential distribution (m=0.7; μ=0.3) with simulated response time
12. truncated exponential distribution (m=0.8; μ=0.15) with simulated response time
13. truncated exponential distribution (m=0.9; μ=0.15) with simulated response time
14. truncated exponential distribution (m=0.9; μ=0.3) with simulated response time
15. truncated exponential distribution (m=1.2; μ=0.3) with simulated response time
16. truncated exponential distribution (m=1.7; μ=0.9) with simulated response time

Figure 31: Jitter design efficiency - comparison of exponent distributions
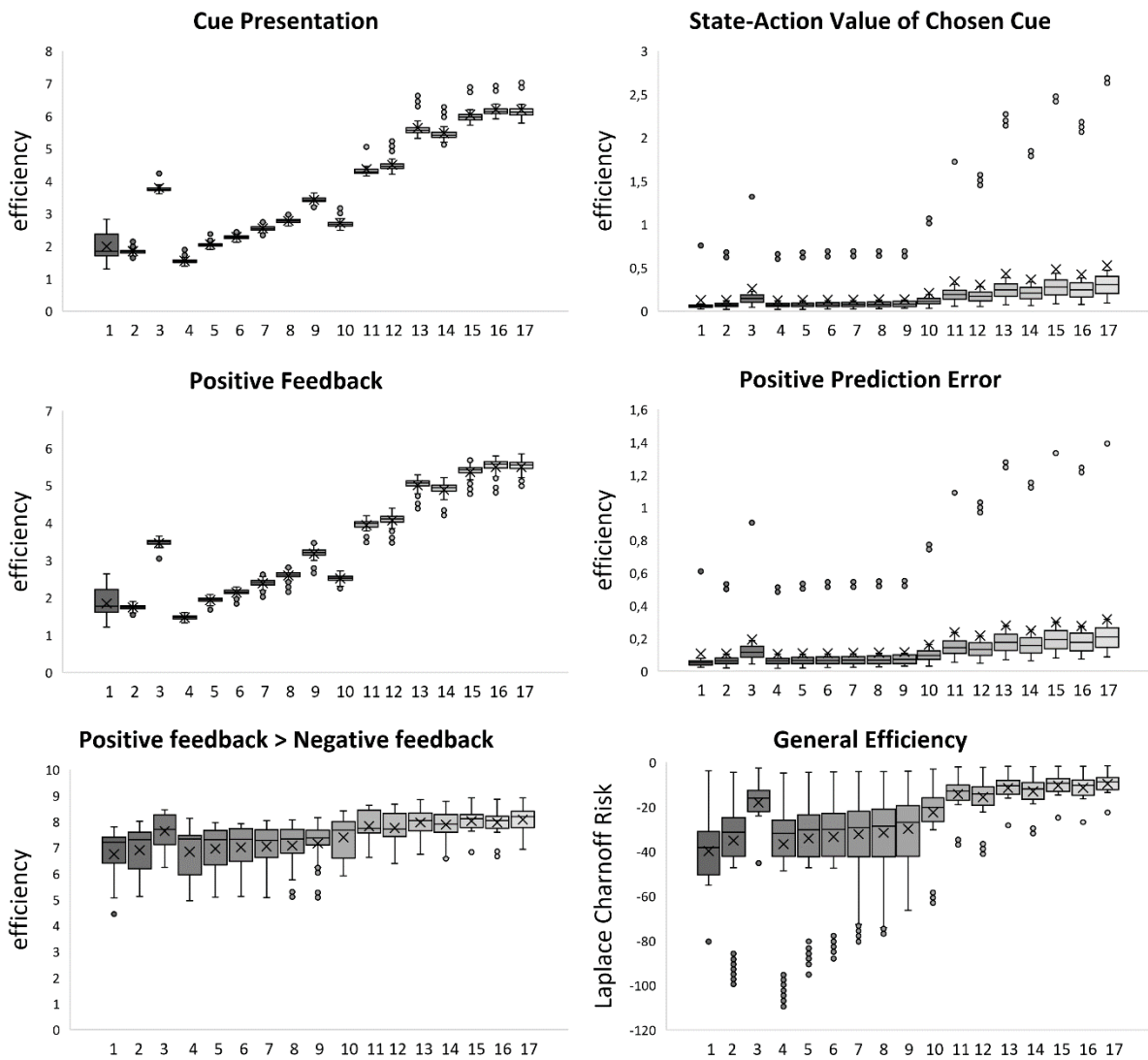
1. jitter design used in the study - truncated exponential distribution (m=0.6; μ=0.3) with real response time
2. truncated exponential distribution (m=0.6; μ=0.3) with simulated response time
3. jitter design used in the study with 1.7s response time
4. truncated exponential distribution (m=0.3; μ=0.15) with simulated response time
5. truncated exponential distribution (m=0.9; μ=0.3) with simulated response time
6. truncated exponential distribution (m=1.2; μ=0.3) with simulated response time
7. truncated exponential distribution (m=1.2; μ=0.6) with simulated response time
8. truncated exponential distribution (m=1.5; μ=0.6) with simulated response time
9. truncated exponential distribution (m=2.0; μ=0.9) with simulated response time
10. two jitter sets, each (m=0.3; μ=0.15) with simulated response time
11. two jitter sets, each (m=0.3; μ=0.15) with 1.7s response time
12. two jitter sets, each (m=0.6; μ=0.3) with simulated response time
13. two jitter sets, each (m=0.6; μ=0.3) with 1.7s response time
14. two jitter sets, each (m=0.9; μ=0.3) with simulated response time
15. two jitter sets, each (m=0.9; μ=0.3) with 1.7s response time
16. two jitter sets, each (m=1.2; μ=0.3) with simulated response time
17. two jitter sets, each (m=1.2; μ=0.3) with 1.7s response time

Figure 32: Jitter design efficiency - comparison of uniform distributions

1. jitter design used in the study - truncated exponential distribution (m=0.6; μ=0.3) with real response time
2. truncated exponential distribution (m=0.6; μ=0.3) with simulated response time
3. jitter design used in the study with 1.7s response time
4. uniformly distributed 0-0.75s with simulated response time
5. uniformly distributed 0-0.75s with 1.7s response time
6. uniformly distributed 0-1.5s with simulated response time
7. uniformly distributed 0-1.5s with 1.7s response time
8. random jitter between 0,0.5s,1s,1.5s with simulated response time
9. random jitter between 0,0.5s,1s,1.5s with 1.7s response time
10. two jitters, each uniformly distributed 0-0.75s with simulated response time
11. two jitter sets, each uniformly distributed 0-0.75s with max 1.7s response time
12. two jitter sets, each uniformly distributed 0-1.5 with simulated response time
13. two jitter sets, each uniformly distributed 0-1.5 with max 1.7s response time
14. two jitter sets, each random between 0,0.5s,1s,1.5s with simulated response time
15. two jitter sets, each random between 0,0.5s,1s,1.5s with max 1.7s response time
16. two jitter sets, each uniformly distributed 0-3s with simulated response time
17. two jitter sets, each uniformly distributed 0-3s with max 1.7s response time

7.4.5.　　Gustometer configuration

Table 6: Pump settings

Parameter	Setting
Speed	$28\dfrac{ml}{h}$
Phase 1	$0.2\ ml$
Phase 2 (milkshake | tasteless)	$0.8\ ml \mid 0.6\ ml$
Phase 3 (milkshake | tasteless)	$0.6\ ml \mid 0.4\ ml$

Testing the amount of fluid delivered showed that the pumps didn't deliver the exact amount they were programmed to deliver, and the delivered amount also differed between the fluids, presumably due to their different viscosity. The parameters were fine-tuned to output a de facto value of $0.49 \pm 0.1\ ml$ for the tasteless solution and $0.43 \pm 0.05\ ml$ for the milkshakes. (N=100 pump runs per fluid type. data given in $mean \pm SEM$)