METHODS

**GSR-recording device.**

In electrical terms, the skin can be modeled as a series of parallel resistors whose conductivity (or inverse resistance) increases with secretion of sweat. Because resistances are hard to measure using a microcontroller, the measuring device is based on a voltage divider, where R1 is a fixed resistor and R2 represents the skin. Custom-built electrodes were fabricated from 5-Rappen coins, which were selected because of their high ratio of highly conductive copper (92 %). An Arduino Uno \cite{arduino} microcontroller was connected to the voltage divider, supplying 5 Volts to R2 and the node between R1 and R2 was connected to a 10-bit analog-to-digital converter (ADC). Thus, the measured values range from 1023 (short circuit) to 0 (open circuit). To balance the tradeoff between sensitivity and dynamic range, we selected R1 for each individual subject such that the baseline input value was in the vicinity of 600. Typically, resistors between 100 kOhm and 1 MOhm were used.

The Arduino Uno was programmed to interface with MATLAB, providing samples at a rate of about 65 Hz.

\begin{figure}

\includegraphics{GSR\_device.png}

%\includegraphics{GSR\_device\_foto.png} ?

\caption{Arduino-based GSR recording device.}

\end{figure}

**GSR-processing and analysis (Jannis)**

…

**Experimental design**

The behavioral experiment was based on Iglesias et al \cite{Iglesias2013} and consists of two blocks à 150 trials each. Each trial consisted of a binary visual cue shown for 400 ms, 1200 ms response time, 1000 ms auditory stimulus presentation and a variable inter-trial-interval of 1000 ms \pm 500 ms. Cues and cue-stimulus contingencies were governed by the following probabilities:

Every 30 trials, the (?) changed in a discrete fashion in the following order: 0.9, 0.1, 0.5, 0.7, 0.3, adding up to a total of 5\*30 = 150 trials per block. Importantly, the two blocks differed solely in the nature of the auditory stimulus: a neutral 200 Hz tone in the first block and an aversive white noise tone in the second block.

**Experimental procedure**

*Initialization*. First, subjects were handed an individual sheet containing instructions (see Appendix) and a brief questionnaire, which they were instructed to start filling out. To guarantee their anonymity with respect to data analysis, subjects were first asked to pick an ID from an envelope and write it on their sheet. Next, any remaining questions as to the experimental procedure were answered, ensuring that no participant was told how or when the predictive strengths of the cues would change. Finally, to improve motivation, subjects were told that they would receive a performance-dependent compensation of 4 Rappen per correct trial, i.e., a maximum of 12 CHF.

*GSR recording*. Electrolyte gel was applied to two spots on the participants’ left hand and custom-made GSR electrodes were thoroughly mounted using medical tape. The data was recorded using a custom-written MATLAB \cite{mathworks?} interface specifically designed to sample GSR data at an effective sampling frequency of around 65 Hz.

*Stimulus amplitude calibration*. Before starting the experiment, the amplitude of the aversive stimulus was calibrated for each participant individually to maximize comparability. To this end, subjects were fitted with over-ear headphones, presented with a randomized sequence of the aversive white noise stimulus of 1000 ms duration at different amplitudes and were asked to rate each one on a painfulness scale from 0 (harmless) to 100 (pain threshold). Subsequently, the stimulus amplitude corresponding to 90 % of their individual pain threshold was extrapolated, presented to the participant for approval, and adjusted if necessary.

*Experiment*. Subjects sat in front of a screen with their left hand laying on the table (and being recorded from) and their right middle and index fingers on a keyboard to provide behavioral output. Presentations of cues and stimuli as well as recordings of behavioral and physiological data were carried out in MATLAB. (Supplementary Video?)

\*\*\*

INTRODUCTION

\section{Introduction}

\subsection{Towards an objective and quantitative strategy for diagnostics in psychiatry}

\textit{The status quo.} Traditionally, psychiatric diagnostics are performed on the basis of a classification dictated by the Diagnostic and Statistical Manual (DSM), or similar, which comprises a comprehensive list of psychiatric disorders along with the signs and symptoms that make patients eligible for diagnosis. In the case of schizophrenia, one of the most prevalent psychiatric disorders, this list includes delusions, hallucinations, anhedonia, flat affect, asociality, disorganized behavior, and many more. According to the DSM, exhibiting at least two symptoms (at least one of the positive variety) over the course of at least one month is sufficient to obtain the label \say{schizophrenic}. While this approach to diagnostics has proven helpful in effectively dealing with great numbers of returning soldiers post Wold War II, it inherently suffers from at least two major flaws. First, the pool of patients under the same diagnostic umbrella inevitably becomes highly heterogeneous in terms of pathologies, and second, there is no systematic and informed way of prescribing drugs, let alone predicting treatment courses: Drugs tend to be prescribed in a trial-and-error fashion, often based on what side effects the individual patient could tolerate best. \\

\textit{The computational approach.} Deeming this status quo untenable--- both for clinicians and patients, but also their relatives--- the emerging field of computational psychiatry has set out to develop methods that allow for the objective and quantitative classification of patients. Moving away from subjective reports of ailments, towards diagnostics that are grounded in aetiology rather than phenomenology, requires external sources of information in the form of neuroimaging or behavioral data. Naively, one could imagine that differential diagnosis may be feasible based on raw, anatomical data, such as magnetic resonance imaging (MRI) scans. While projects of this kind have been successfully undertaken (cite 7T-schizo/non-schizo ML paper), they bring us no closer to the bottom-up understanding of disease mechanisms that would enable us to prescribe targeted treatments.

\subsection{Generative models of neuroimaging data}

A more promising approach rests in the concept of \textsf{generative modeling}, in the context of which physiologically interpretable, computational models of the brain are constructed based on neuroimaging data. More technically, those models mathematically describe the probabilistic links between hidden states of the brain and the noisy measurements we acquire using MRI or electroencephalography (EEG). The result of this approach is a set of subject-specific parameters that can be readily mapped onto physical entities (such as the average weight of excitatory long-range connections between two specific brain regions) and therefore exploited for a white-box classification into disease cohorts. The work of (cite synaesthesia study) on synaesthetes serves as illustrious proof of concept of this approach: Based on an anatomically informed set of prior models, two groups of different phenomenology, namely the \say{projectors} and the \say{associators} could be reliably distinguished, which has led to an understanding at the mechanistic level in terms of effective brain connectivity. Several similar studies within the clinical realm suggest that this approach can in principle be applied to assay specific neurophysiological parameters (\textit{e.g.}, potassium channel \cite{Gilbert2016} or NMDA receptor activity \cite{Symmonds2018}, in an effort to facilitate objective, quantitative psychiatric diagnostics. \\

\subsection{Generative models of behavior}

Embedded in a set of ideas termed the \textsf{Free Energy Principle}, Karl Friston argues that given the objective of minimizing energy expenditure, the brain's optimal policy is to hold and constitutively update a generative model of its inner and outer milieu, in an effort to avoid costly \textsf{actions} on the word or on itself (by changing its anatomy). The ensuing generative modeling approach of behavior rests on the Bayesian Brain hypothesis (cite something), which postulates a \textit{hierarchical} generative model within the brain that is aimed at minimizing statistical \textit{surprise}: the \textsf{Hierarchical Gaussian Filter} (HGF) \cite{Mathys2014}. For the simple case of predicting a series of binary outcomes, the model would be structured as follows: At the lowest level, drawing from a Bernoulli distribution yields a binary outcome. To account for dynamic changes in the probability of the latter distribution, the second level represents it as a Gaussian random walk, the tendency, whose variance in turn depends on a random variable at the third level, termed the volatility. The parameters that govern the dependencies between different layers in the hierarchy can be fitted to individual subjects, given behavioral data from a suitable task. In addition to those parameters, individual model fits also yield interesting computational quantities for each trial, such as \textsf{precision-weighted prediction errors} (pwPEs), which can then be correlated to brain activity using conventional general linear models for fMRI. Following such an approach within the context of an audio-visual learning paradigm, Iglesias et al \cite{iglesias2013hierarchical} found neurophysiolgoical correlates of prediction errors at different levels of the generative hierarchy in different regions of the brain, e.g., low level PEs in the midbrain. \\

\subsection{Peripheral correlates of computational quantities?}

It is intuitively sensible that surprise, in the colloquial sense, draws attention by creating arousal. For instance, when absent-mindedly attempting to cross a street, and noticing a car frantically breaking, you become aware of the danger \textit{as a result of} an innate emotional response, triggered by the fact that your brain had not predicted this specific, highly salient input. Physiological responses such as increased heart rate, tense muscles and active sweat glands quickly ensue as a result of this near-fatal prediction error. Thus, it stands to reason that prediction errors--- even those of the milder variety ---or other abstract computational quantities, could in principle be assayed by reading out those peripheral states. \\

In the present study, we look for correlates of prediction errors in skin conductance. To this end, we acquired behavioral data from 16 subjects in an audio-visual learning paradigm based on \cite{iglesias2013hierarchical}, where the subjects are presented with a binary visual cue and subsequently asked to predict whether there will be an auditory stimulus or not (see \textsf{Methods}). The difficulty of the task lies in the fact that the probabilities that govern the cue-stimulus contingencies change over time. In addition to behavioral data, we measured the participants' skin conductance (see \textsf{Methods}) over the course of the experiment. Combining individual HGF fits and the respective prediction error trajectories with skin conductance data yields a promising perspective on physiological correlates (see \textsf{Results}). \\

\subsection{Do aversive vs. neutral stimuli bias behavior?}

Our experiment was designed to address a second question. In the first block of 150 trials, the auditory stimulus whose occurrence the participants were tasked to predict was a neutral, 200 Hz pure tone with a windowed onset. In the second block however, the tone was an aversive, sudden-onset white noise stimulus, at an amplitude calibrated to 90 \% of the individual acoustic pain threshold. Comparing HGF parameter fits between neutral and aversive blocks would hopefully shed some light onto how the nature of the stimulus confounds behavior. One prior hypothesis is that participants whose $\omega\_{2}$, an additive term affecting the variance of the tendency (see \textsf{Methods}), increases

1. *Peripheral* correlates of computational quantities: SCR
2. The Experiment: Effect of stimulus nature on behavioral model
3. Complementary interpretation of HGF parameters – twoI, alpha model