**Title**

**Abstract:**

**Introduction**

When people make decisions, many cognitive processes are involved, some under careful conscious control others outside the conscious realm. The cognitive process under conscious control is well known to people trying to be more active; here careful planning of their day with rewards for being active, are practices commonly done. Processes that are outside the conscious realm includes decisions that are made quickly and rely on experience, that is habits, emotions etc. The interaction between these processes and systems is best described by an example. Having decided to lose weight, you find yourself in a meeting, with snacks on the table. Here the unconscious system will resort to habits and internal desires and make you reach out for the snacks as you usually do, however with the conscious goal of losing weight you decide not to grab them as this decision would go against the goal of losing weight. This example illustrates how different processes that are both conscious and unconscious interact in deciding what decisions are made. Previous literature has investigated how good people are at consciously determining why they make certain decisions. For instance, in a study by Mazar & Wood (2022) participants decisions and explanations for these decisions was investigated. It was found that there was a great discrepancy between the observed behavior and the explanations given. In the first of the two studies it was found that people attributed their willingness to help a researcher after a finished experiment to valence, which had no effect on their actual willingness to help, all group level results. It was also found that habit, which was manipulated in the study influenced their willingness to help, but its effect was not acknowledged by the participants to the degree it influenced their behavior. In the second study the researchers found that this effect of misattribution was also present in a more ecologically valid experiment utilizing ecological momentary assessment on people’s everyday coffee drinking habits. Here it was found that participants mostly attributed their coffee drinking to fatigue, however the results indicated that habits (time of day, location, etc.) was a much stronger predictor of coffee drinking (Mazar & Wood, 2022). This again demonstrates how conscious processes “I am fatigued, therefore I drink coffee” and unconscious processes such as habits both influence behavior. The results also strikingly show that attribution to what drives decisions and behavior might not be accurate, as people tend to attribute their conscious processes more weight than what is actually the case. Many instances of how unconscious processes influence the behavior and decision making of humans have been investigated in psychology and cognitive science. One of the most well-known phenomena where unconscious processing drives behavior and decision making is known as cross modal correspondence. Which is the tendency of humans to match sensory features in one sensory modality to features in another modality. This could for instance be the association between high pitched tones and light objects (Parise, 2016).

**Cross modal correspondence**

In many introductory psychology or cognitive psychology textbooks a chapter about perception is often accompanied by the description of the binding problem (Anderson, 2000; Treisman, 1996). This problem arises from the observation that the visual input received in the retina and transferred to the occipital lobe is seemingly dismantled into distinct regions and processing streams e.g. the where and what pathway as well as regions that seem to be specialized for color, motion or orientation (Bramão et al., 2010; Kravitz et al., 2013). This leads to the binding problem, which can be stated as if the brain processes these qualities individually, and independently, how does the conscious percept of a vertical red bar and a horizontal green bar appear, when sensory information from these two objects is received simultaneously? The question is how does the brain know that the green bar is horizontal and the red bar vertical, when and if color and orientation is processed separately (Anderson, 2000).

Another layer to this binding problem is how information from multiple senses are integrated into a single percept. This integration of multisensory information has in recent years been heavily investigated (de Dieuleveult et al., 2017; Gröhn et al., 2022), and a large amount of research has shown that people have tendencies to match certainty sensory modality characteristics with others. For instance, peoples tend to associate high-pitched sounds with small bright objects that are located high in space (Parise, 2016). Another extremely known example is the kiki bouba effect, which asks people to allocate the two names (e.g. Kiki or Bouba) to two objects, one with round edges and one with sharp edges. Results from now a huge body of evidence shows that people tend to ascribe the object with the sharp edges to Kiki and the object with round edges to Bouba, also across cultures (Ćwiek et al., 2022; Lammertink et al., 2016). The phenomenon described in the two examples above is commonly referred to as Cross modal correspondence. This phenomenon is interesting in of itself, as it perhaps can help solve the binding problem and help to understand the weird nature of perception and perhaps consciousness (Spence, 2011). However, the phenomenon is perhaps also necessary to study and understand for research using perceptual tasks. This is the case so that researchers can control or even elimination such effects from experimental paradigms that utilizes stimuli from different senses, without studying the effect. This could for instance be in associative learning paradigms which is beginning to see more light as computational efficiency and models are being more assessable (de Berker et al., 2016; Delamater & Lattal, 2014). It is therefore import for researchers in these fields to be aware of the cross-modal correspondence effect, preferably when designing their experiments, and otherwise to model it, so it doesn’t systematically bias their results and inferences.

The current study seeks to investigate the cross-modal correspondence effect between tones of different pitches and temperatures. One study to date has investigated this kind of cross-modal correspondence and found that high pitch tones are associated with cold temperature and stimuli whereas low pitch tones are associated with warm temperature and stimuli. (Wang & Spence, 2017).

**Anchoring effect**

When participants make choices based on previous information, there is a tendency to rely to heavily on information from a reference point. This reference point in many normal psychological or cognitive science experiments could be the starting trials of the experiment. This makes individual choices biased in the direction of what happened in the start of an experiment, as this reference point anchors the participants’ belief more than what would be optimal for performance. (Furnham & Boo, 2011)

**The hierarchical gaussian filter:**

Statistical and computational modeling are usually used when trying to make inferences about human decision making. When trying to describe human decision making, when learning is involved, a time dependency is necessary in the model as learning occurs over time. The hierarchical gaussian filter (HGF) is a learning model developed by Christoph Mathys and colleges (C. Mathys et al., 2011; C. D. Mathys et al., 2014). This learning model tries to combine the computational efficiency of reinforcement learning models such as Rescorla Wagner (Rescorla & Wagner, 1972) with more theoretically driven Bayesian models that can require a huge computational load. The HGF tries to combine these approaches by using various approximation strategies of the Bayesian posterior to derive trial by trial update equations as seen with reinforcement learning models. For a full overview of the update equations of the HGF interested readers are referred to the original HGF paper (C. Mathys et al., 2011). The main idea of the HGF is that as a participant is learning an association between a cue and an outcome both usually binary. This cue outcome association is then combined in an arbitrary way into another binarization of 0 and 1, often referred to as the contingency space. This means that if the cues were high and low tones and the stimulus cold and warm, then the contingency space in this case, could be that 1 represents when the high-tone perfectly predict the cold stimulus and the low-tone *necessarily* perfectly predict the warm stimulus. This would therefore also mean that 0 would represent that the high tone would perfectly predict the warm stimulus and the low tone perfectly predict the cold stimulus, the arbitrariness to this binarization is of cause that there is no inherent meaning which association is 0 and which is 1. The contingency space can therefore be expressed as the conditional probability of the stimuli given the cues as the stimulus is preceded by the cues. Experimental paradigms using the HGF as their learning model utilizes this by changing this conditional probability over time, introducing uncertainty, which is exactly what the HGF tries to model.   
The name Hierarchical gaussian filter, comes from the fact that on a subject level the model has several layers of belief and uncertainty. In principle the model can have infinitely many layers, however in most literature published either 2 or 3 layers seem to make the best tradeoff between under and over fitting (de Berker et al., 2016; Lawson et al., 2017). Introducing more layers in the HGF has the disadvantage of being more complex and risk overfitting, however another problem is also present which is not discussed in the literature to the authors knowledge. This problem arises because of the hierarchical structure and how information flows between the different levels. If the task is not volatile enough, then there is not enough information that reaches the top level of the hierarchy and results in parameters that cannot be recovered as they basically have no explanatory power, as is seen with very large neural networks (Rehmer & Kroll, 2020). It is therefore of high importance that researchers using the HGF, performs parameter recovery for different levels of the model before data collecting to be sure that their experiment has enough complexity to recover the parameters, they are interested in.

**Walk-through of the HGF**

In this section the 2 level HGF is explained in detail, as this model is used for the computational modeling in the current paper, a brief introduction to the 3-level HGF is also given to illustrate the importance of parameter recovery in the computational modeling section.  
The HGF’s first level evolves in time in the contingency space where the participants’ belief about the association evolves in time as a Bernoulli distribution. This means that as one’s belief is strengthened (closer to either 0 or 1) the uncertainty of the belief decreases, as the variance of the Bernoulli distribution is described by its mean. The parameter for this Bernoulli distribution is governed by the second level of the HGF through a sigmoid transformation. The second level evolves in times as a gaussian distribution which has a mean which is governed by the previous time point, the standard deviation is crucially influenced by a subject specific parameter , which controls the subjects’ learning rate. The crucial update equations of the HGF can be seen in equation 1 through 4. Reader familiar with the Rescorla Wagner learning model will notice that the update of the mean of the second level closely resembles that of the Rescorla Wagner learning model, here the learning rate is however varying by trial and not only by subject as with the original Rescorla Wagner learning model (Rescorla & Wagner, 1972).

These equations form closed loop update equations, which only need priors on the initial belief on the second level and its uncertainty and priors for the subject specific parameter and the inverse decision noise , which will be described in the reparameterization of the HGF section.

To make an additional layer to the HGF two new subject specific parameters are introduced and . describes the crosstalk between the second and third level whereas describes the variance of the third level. This new level of the HGF evolves in time as the second level, that is as a gaussian random walk with mean and the variance being . This new third level only influences the uncertainty on the second level which means that the update equation for the uncertainty on the second level at time t, is now:

Where is the mean of the third level on the previous trial. The update equations for third level are not written out here, but interested readers are encouraged to referred to the original paper by Mathys and colleges (C. Mathys et al., 2011; C. D. Mathys et al., 2014).

**Objectives**

The goal of the current paper is to implement the HGF in its current form into R, utilizing just another Gibbs sampler (JAGS) for estimation. Next a hierarchical, hierarchical gaussian filter (HHGF) is going to be implemented such that effects like partial pooling is considered when fitting several participants at once. After implementing the HHGF an analysis of a large scale behavioral associative learning experiment is going to be analyzed utilizing the HHGF in R. When analyzing the behavioral data two hypotheses will be tested:

**Hypotheses**

Participants will have a stronger tendency to predict a cold stimulus, when preceded by a high pitch tone and equally have a stronger tendency to predict a warm stimulus when preceded by a low pitch tone (cross-modal-correspondence).

Participants starting in the sequence of trials where the warm stimulus is associated with the high pitch tones will display a weaker effect compared to participants starting in the sequence of trials where the cold stimulus is associated with the high tone (anchoring bias).

**Methods**

As a part of Aarhus brain project, a large-scale research program, over 300 participants went through several behavioral experiments. One of the behavioral experiments in the project was a probabilistic thermal association learning task. In this experiment participants’ (n = 247) thermal sensitivity and pain threshold were assessed to make sure that the temperatures used in the experiment was noticeable but not painful. After assessment and adjustment, a short introduction to the main experiment was given whereafter 10 practice trials were performed to make sure the temperatures used were appropriate and the participants knew the trial structure. These 10-practice trial were crucially random, and participants were told that they did not have any inherent structure. The main experiment consisted of 60-90 minutes of associative learning. Participants were asked to try and predict an upcoming stimulus either cold or warm from a preceding cue that was either a high pitch tone or a low pitch tone. The probability that one tone was associated with one stimulus was fixed in certainty probability block but varied between them, see figure 1b and 1c. In approximately every other trial the participant was to give a rating of the perceived intensity of the stimulus, on three different scales, cold, warm and burning. To account for any order effects the main experiment counter balanced participants such that half of the participants started in the block where the high pitch tone was associated with the cold stimulus and the other half where the high pitch tone was associated with the warm stimulus see figure 1 b and c.

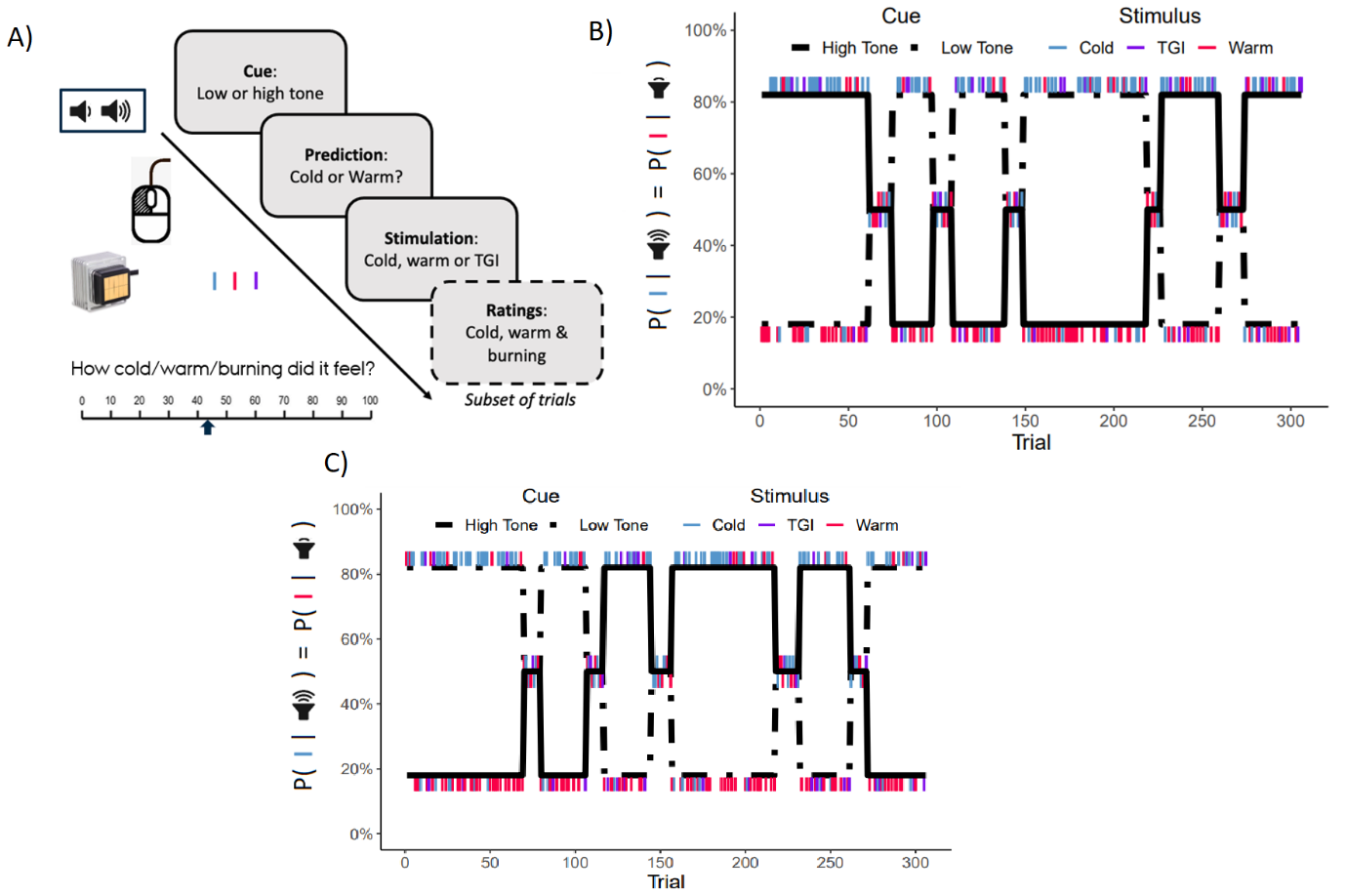


Figure 1: A) depicts the experimental setup, participants were given a cue, then asked to make a prediction on the upcoming stimulus, then the stimulus was delivered and in half of the trials participants were asked to rate the perceived intensity of the stimulus on three scales (cold, warm and burning). B) displays the cue-stimulus association structure for half the participants which had an even ID number, C) Participants with an uneven ID number received this cue-stimulation association. The crucial difference between B) and C) is that participants in B) start with high-tone being highly predictive of the cold stimulus whereas in C) participants start the experiment with the high-tone being highly predictive of the warm stimulus.

**Stimulus**

Participants received both thermal and auditory stimulus. The thermal stimulation was delivered through a thermal cutaneous stimulator, which has a temperature precision of 0.1 degrees Celsius and a ramping speed of around 20 degrees Celsius. The stimulator had five different zones that could individually be cooled or warmed, and the total area of simulating surface was 10 see figure 1a. When receiving the warm stimulus uneven numbered bars were set to the desired temperature of the individual participant, whereas when the cold stimulus was delivered, the even bars were set to the desired temperature of the individual participant all other bars were set to a neutral temperature of 32 degrees Celsius. Participants were to predict this stimulus from one of two auditory tones delivered through headphones. The tones were either high-pitched 1600Hz or low pitched 400Hz which were delivered through psych toolbox (Kleiner et al., 2007).

**The Thermal Grill illusion**

The main experiment was conducted to assess how learning under uncertainty would influence veridical thermal sensation, however another aspect was also introduced which was how uncertainty would influence illusory thermal percepts. To investigate this second aspect approximately every seventh trial had the bars of the stimulator both cold and warm, inducing what is known as the Thermal Grill Illusion (TGI), which is a burning / painful sensation elicited by innocuous interleaved cold and warm stimuli (Craig & Bushnell, 1994; Shin & Chang, 2021). This addition to the experimental paradigm, does influence the results of the current study. However, all trials where the participant received the TGI stimulus was coded in the computational modeling based on the perceived thermal stimulus, as all trials with the TGI stimulus had subjective visual analog ratings (VAS). This meant that when a participant received the TGI stimulus and thereafter rated it on both a cold and warm rating VAS scale the input for that participant at that given trial was put into the contingency space of the model by the proportion the participant felt it as cold. Meaning that if the participant rated the TGI stimulus as completely cold the input to the model would be coded as a cold stimulus, and vice versa for warm.

**Reparameterization of the HGF**

To test the hypotheses of the current study a tweak to the original 2-level HGF was implemented. This tweak was directly related to how the belief of the participant in the contingency space was transformed to a probability of answering 1. The transformation of belief into a categorical choice is often done by a sigmoid transformation that transforms the belief into probabilities such that the sum of all choices equals 1 (e.g a decision is made on every trial), whereafter the choice is generated from a categorical distribution (when inverting the model), which is just a generalized Bernoulli distribution. In the case of the current experiment and others utilizing the HGF in associative learning experiments (de Berker et al., 2016; Lawson et al., 2017) there are only two choices for the participant, 0 or 1 from the models perspective in the contingency space. This together with the fact that the belief in the HGF is a continuous value between 0 and 1, normally referred to as the belief about the cue-outcome association simplifies the response model. A logistic sigmoid transformation (eq 1) is therefore often implemented that restrains the input to range from 0 to 1 while keeping the outcome a probability, as mentioned above. As with the normal sigmoid transformation in modeling of behavior a subject specific parameter is often included that depicts the participant’s tendency to answer what they believe or in other words model exploit / explore behavior. This parameter is often called the inverse decision temperature and has the Greek letter see figure 2A. For the current experiment an additional parameter here called, see eq 2 was added. This additional free parameter made it possible for the logistic sigmoid transformation to display a bias towards either 0 or 1 while still modeling the inverse decision temperature . For a visualization of how the additional parameter changes the shape of the logistic sigmoid transformation see 2B.

Eq 1

Eq 2

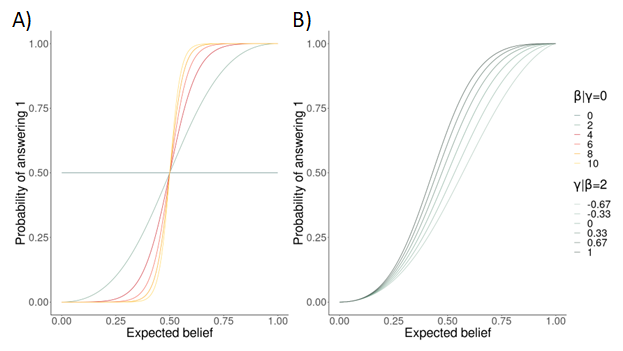


Figure 2; A) depicts the logistic sigmoid transformation with different values of the beta parameter [0;10]. B) shows the logistic sigmoid transformation with a constant beta parameter of 2 with varying gamma parameter [-2/3 ; 1].

**The Hierarchical HGF**

After reparametrizing the original HGF to test the current hypothesis the hierarchical HGF (HHGF) was implemented. To test the hypothesis that participants would display a tendency towards an outcome of 1, which in this experiment meant that the high tone was associated with the cold stimulus and the low tone was associated with the warm stimulus; and that there would be a difference between participants starting in the two different probability blocks, see figure 1B and 1C. This was done by introducing another new free parameter that modelled the difference between the two group’s parameter. The full list of priors for the computational modelling and the plate notation for this new modified HHGF can be seen in figure 3.

A picture containing graphical user interface

Description automatically generated

Figure 3; Shows the plate-notation for the created hierarchical hierarchical gaussian filter, which also includes two new parameters alpha and gamma.

The priors for the model were decided to be weekly informative to help model convergence, however without restricting inference. The priors for the initial belief and uncertainty on that belief was adopted from the original paper (ref), which means that the participants are naïve to the association on the first trial with a quite high uncertainty. The model was inverted using sampling in Just Another Gibbs Sampler (JAGS), with 6000 iterations, where the first 2000 was used as burn in, with 3 chains.  
Results were analyzed by looking at the posterior credibility intervals of the two parameters and , as well as calculating a Bayes factor using the Savage Dickey density ratio test, this will be performed to see how the prior belief of these two parameters changes in light of the data.

**Parameter recovery.**

To be sure that the HGF implemented in JAGS was the same as the one originally developed in MATLAB, I took all the participants from the current study and ran them through the model both in MATLAB and R to see that the parameter estimates were similar. As can be seen in figure 4.

Chart, line chart, scatter chart

Description automatically generated

Figure 4; Shows the correlation between parameter estimates of the hierarchical gaussian filter fitted in the original MATLAB toolbox and in Just Another Gibbs Sampler in R. The black line represents the equation y=x which would indicate a perfect correlation of 1.

Next the new response model was implemented in JAGS such that the hypothesis could be tested. Parameter recovery was run on this model to make sure that the new parameter could be recovered under idealized conditions before implementing the hierarchical part see figure 5.

Diagram

Description automatically generated

Figure 5; shows the parameter recovery of the Hierarchical gaussian filter with the new response model parameter implemented in Just Another Gibbs Sampler in R. The black line depicts the equation y=x, meaning perfect correlation.

Next the HHGF was implemented with the new response model, together with the additional parameter see figure 3. Parameter recovery was then conducted on the HHGF where agents were simulated with different sets of parameter values. A total of 250 subjects were simulated with different parameter values to see if in an idealized world the parameters could be recovered. This simulation was run 50 times and the parameter recovery for the group level parameters can be seen in figure 6.

Chart, scatter chart

Description automatically generated

Figure 6; shows the parameter recovery for the hierarchical hierarchical gaussian filter model for the mean group-level parameters. Black line depicts the equation y=x which would indicate perfect correlation. The red line depicts the equation y=4x.

Before diving into the computational modeling, a mixed effects logistic regression was fit to the data to demonstrate that the behavior of the participants in the ABP did display cross modal correspondence that corresponded to the study by Wang & Spence (2017). This mixed effects logistic regression was fit in a frequentist perspective for the ease of time. The model included a varying intercept per participant with the cue the participants received as fixed effects and with trial included as a covariate, as the dependent variable the prediction of the participant was entered, that is cold or warm.

**Results:**

The mixed effects logistic regression found that participants were significantly more likely to predict the cold stimulus given a high tone compared to a low tone, , se = 0.014, z = 16, p < .001. This was even though that the distribution of stimulus given, was evenly distributed on the cues given (XX).   
Implementing the HHGF with the two new parameters showed good model converge on all group level parameters see appendix A for trace plot. Summary statistics with credibility intervals of the posterior distributions for the parameter estimates of group-level can be seen in table 1. Prior-posterior updates for the parameters and can be seen in figure 7.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters  (Group-level) | Mean | 2.5% CI | 97.5% CI | Rhat | Effective samples |
|  | -0.90 | -1.10 | -0.70 | 1.00 | 3000 |
|  | 0.27 | 0.19 | 0.35 | 1.00 | 3000 |
|  | 0.20 | 0.05 | 0.37 | 1.00 | 3000 |
|  | 2.40 | 1.95 | 2.86 | 1.00 | 780 |

Table 1; displays the group-level estimates of the parameters of the hierarchical hierarchical gaussian filter with credibility intervals.

Bayes factor using the Savage Dickey density ratio test showed that the Bayes factor for no effect was 0.27 on , whereas the Bayes factor for no effect for was .

Diagram

Description automatically generated

Figure 7; displays the prior posterior updates for two of the parameters of the hierarchical hierarchical gaussian filter. Left side displays the results on a spread-out scale and is zoom in on the right side. The black line displays where 0 is that is the line x=0.

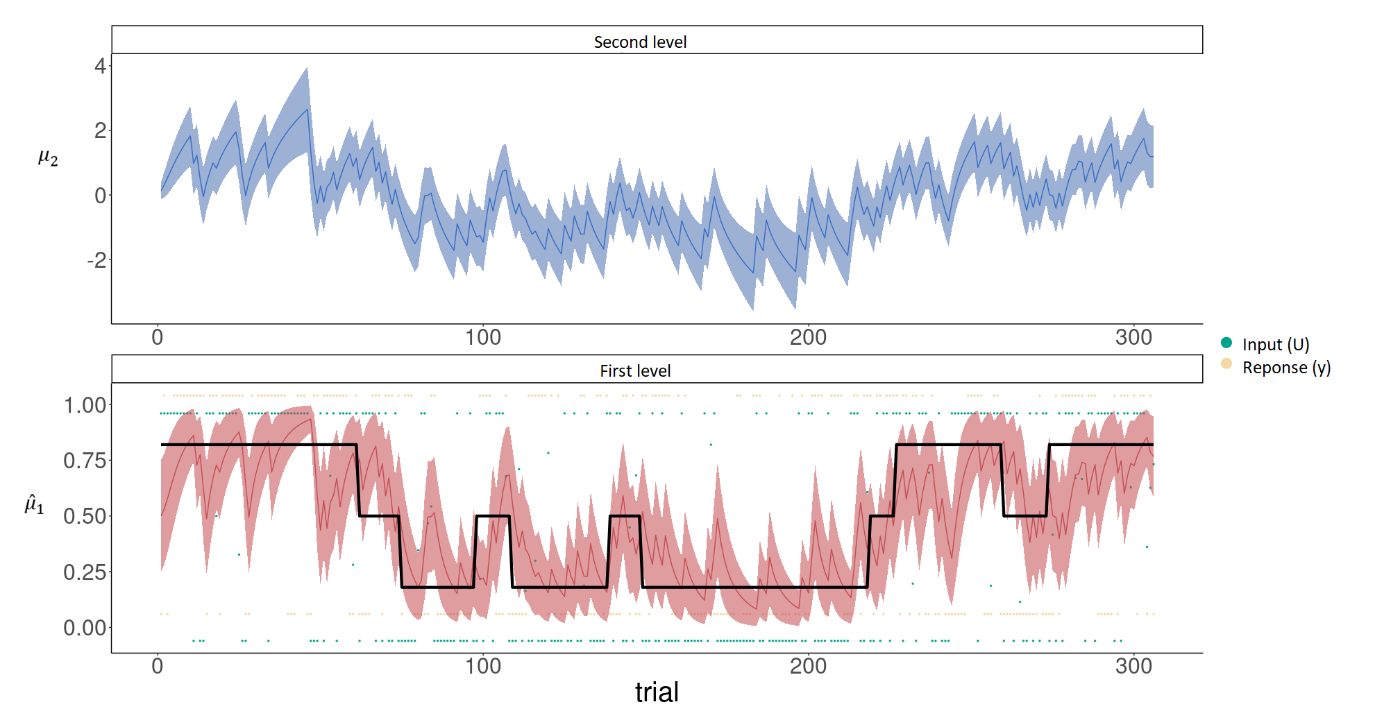
To ensure that the computational model captured the learning behavior and the objective probability blocks participants’ learning trajectories were visualized and inspected, one example trajectory can be seen in figure 8. ****

Figure 8; displays the belief trajectory of one participant going through the probabilistic thermal learning task. The solid black line depicts the underlying probability in the contingency space, whereas the green dots represent the individual trials. The yellow dots display the participants’ responses of the given trial. The lower panel displays the first level of the HGF where the red line is the belief of the participant with its uncertainty in the shaded area. The top panel displays the second level of the HGF and its uncertainty in the shaded area.

**Discussion**

Computational modeling has shown promise in modeling human behavior and decision-making both in healthy and diagnosed participants (Lawson et al., 2017). The main objective of the current study was to implement a known learning model, the Hierarchical Gaussian Filter (HGF), in a different programming language, R, with a different optimization strategy, Just Another Gibbs Sampler, with several additional features to investigate whether participants in a large-scale study would display cross modal correspondence and an anchoring effect. Parameters of the models introduced in the current paper seem to be well recover. It is however to be noted, that there is no perfect correlation between the estimates between the model from the original HGF in MATLAB and the one introduced here figure 4, this is probably to be explained by the different optimization strategies, because a Variational Bayesian optimization strategy is used in the original HGF, making it computationally efficient (C. D. Mathys et al., 2014). Figure 6 depicts the parameter recovery for the group level estimates from the Hierarchical HGF, which was also implemented in the current paper and used to answer the hypotheses at hand. The figure shows that the only group level parameter that seems to be not well recovered at least in its original scale is the beta parameter. Even though the actual value of this parameter could not be exactly recovered the model does recover the pattern, meaning that if the simulated beta estimate increases so does the recovered value however instead of matching the increase the model seems to estimate the parameter to the value times 4 see figure 6 top right panel.   
After having done parameter recovery to ensure that the model parameters could be recovered in idealized conditions the new model (see figure 3 for the plate notation and priors) was fit to the data from a large-scale association learning task with 247 subjects. The model showed that participants did display a tendency to answer 1 in the contingency space, which was coded as the association of the high-tone with the cold stimulus and the low-tone being associated with the warm stimulus. This tendency was captured in the gamma estimate which had a 97.5% credibility interval of [0.19, 0.35], indicating that this parameter value is bigger than 0, which confirms the first hypothesis. Using Savage Dickey density ratio test it was found that in light of the evidence the belief that this parameter is 0, that there is no effect, should be decreased by a factor of 3100000 from the prior belief which was set to a normal distribution with a mean of 0 and a standard deviation of 4.   
The other hypothesis tested was whether there was a difference in the gamma parameter between participants that had different probability structures see figure 1B and 1C for the two different probability structures. This difference in the gamma parameter between the two-probability groups was captured in the alpha estimate which had a 97.5% credibility interval of [0.05, 0.37], indicating that this parameter value was also bigger than 0, confirming the second hypothesis. Using Savage Dickey density ratio test it was found that in light of the evidence the belief that this parameter is 0, that there is no effect, should be decreased by a factor of 4 from the prior belief which was set to a normal distribution with a mean of 0 and a standard deviation of 4. This result should therefore be interpreted with caution as there is not very strong of decisive evidence that this parameter is not 0.

Need a real discussion on the impact and why?

**Limitations:**

One of the main limitations with the current work on the inference on the cross-modal correspondence effect and the anchoring effect is that the current study did not only have cold and warm stimulation, but also the combination of the two. This addition of the combination of the two stimuli to induce the thermal grill illusion could perhaps bias the results in the current study, but only if the participants did not update their beliefs / learned the cue-stimulus association based on the VAS ratings they provided, which ideally should reflect their perception. It is only a bias in that scenario because these trials were entered into the model based on their provided VAS ratings.

**Conclusion:**

Additional material

Github

Appendix

Model convergence:

Chart

Description automatically generated Chart, histogram

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

References: