Chapter 5

VISUAL DISCRIMINATION WITH UNKNOWN STIMULUS ONSET

Sequential Reasoning with a Nonstationary Probabilistic Model

Our last project is a psychophysics study of visual discrimination with uncertain stimulus onset. Unlike the previous problems, in this problem a probabilistic model is given, but the model is not stationary over time.

5.1 Motivation

An organism's survival is critically dependent on its ability to detect change (e.g. the sound/sight of something moving in the distance), and classify its nature (e.g. a predator, prey, or meaningless clutter). In ecological conditions, change detection and object classification frequently co-occur: approaching animals need to be detected and classified as friend or foe. Despite the ecological significance of considering detection and classification jointly, the two tasks are typically studied in isolation. Consequently, it remains unknown how humans jointly perform classification and detection, and whether and how humans trade off speed and accuracy.

Psychologists have studied the phenomenology of visual discrimination as well as computational approaches [1]–[3]. We have reviewed that the optimal model for trading off speed and accuracy is the sequential probability ratio test (SPRT) [4]. When the discrimination is between two simple templates, the diffuse-to-bound process [5] is also optimal. These discrimination models require knowing when change happens, i.e. when to start accumulating evidence, which is not a realistic hypothesis in most ecological conditions.

The phenomenology of change point detection is relatively less explored. Earlier studies examine *whether* change occurred [6]–[9], and, more recently, *when* it occurred [10], [11]. The optimal model for minimizing detection error and reaction time [12] dates back to the cumulative sum control chart (CUSUM) [13], [14], which utilizes a diffuse-to-bound mechanism with only one absorbing boundary. When the change could bring the world into one of multiple states, a network of diffusions is required [15] to integrate changes attributable to different categories optimally. Despite addressing the uncertainty in change onset, these models do not consider the question of classification.

Contributions

(1) We study the the **joint** detection and classification task (the 'dual task' for brevity). Our experiment is a variant of random dot motion discrimination [2] where the motion is completely incoherent at first. After a random delay it becomes coherent in one of two directions. The subject is asked to both detect change and classify the coherent motion. We manipulated the motion directions to control the relative difficulty of detection and classification.

(2) We developed three **computational models** for the dual task. The first model 'Classifies and then Detects' (CD), which is optimal [16]. CD applies SPRT on the probabilistic model of both classification and detection. The second and third models are computationally simpler and sub-optimal, where they apply SPRT separately on the detection and the classification problem. The two models differ in the temporal order in which the SPRT modules are executed. Model two conducts 'Detection and Classification in Parallel' (DCP), while the third model conducts 'Detection and Classification in Series' (DCS).

(3) We **test human subjects** on the dual task as well as a pure detection task. Fitting the parameters of our models to data collected from both tasks reveals that the only model that is consistent with human SAT behavior is the **conceptually simple but sub-optimal** DCS model. Primates have been found to be near-optimal in detection and classification [11], [17]–[19] and our findings deviate from this pattern.

(4) To fit our models on random-dot motion patterns, we develop a simple model of early vision [20], [21] based on **quantized** sensory input, which are action potentials from motion-tuning neurons in area MT [22]. This model is parsimonious and versatile: with one free parameter it simulates sensory inputs for detection tasks and dual tasks with arbitrary coherent strengths and motion directions. This generalization ability is an improvement over other decision models of random dot motion discrimination [3], [23], [24], which typically are independently parameterized across tasks and only generalize across the level of coherence of motion stimuli.

5.2 Framework for visual discrimination with unknown onset

Chapter-specific notations

Formally in the dual task, the world exists in one of three states at any given time bin t: $C_t \in \{0, 1, 2\}$, where time bin t represents the duration $((t-1)\Delta, t\Delta]$, where 0 is the initial state (e.g. incoherent motion), 1 and 2 are two post-change states to be

distinguished (e.g. coherent motion along one of two directions). The world always starts from $C_0 = 0$ and changes to either class 1 or class 2 at a random time t_δ . The change occurs *only once*. The observer has information regarding the distribution of the change time t_δ, but not the actual value of t_δ. The goal is to infer the stimulus category $C \in \{1, 2\}$ as quickly as possible, but not earlier than t_δ, in which case the response is considered a false detection error. **Fig. 5.1a** illustrates the setup for the dual task in the context of random dot motion discrimination (see **Sec. 5.3**).

Models

Our three models (CD, DCP, and DCS) vary in optimality and simplicity. CD is optimal. The initial incoherent motion and the two coherent motions are modeled as three separate stimulus categories, and the dual task is reduced to a multi-category classification task, which may be solved optimally [16]. DCS and DCP are computationally simpler and sub-optimal. Both use a detector to identify any kind of coherent motion, and a classifier to distinguish between the two motion directions. In DCP, the detector and the classifier operate **simultaneously**, and as soon as the detector reveals a change, the classifier is consulted to reveal the nature of the change. In DCS, the coherent motion detector **triggers** the integration time for a classical diffuse-to-bounds classifier which eventually reaches a decision.

Our models of the dual task assume optimal evidence accumulation from input sensors [21]: all models have access to the following two statistics computed according to Bayesian inference. The first statistic is the log posterior ratio between any pair of classes (derived in **Eq. 5.10**):

$$S_t^{i,j} \triangleq \log \frac{P(C_t = i|\boldsymbol{X}_{1:t})}{P(C_t = j|\boldsymbol{X}_{1:t})}, \tag{5.1}$$

where $S_t^{i,j}$ is the log posterior ratio between class i and j ($i, j \in \{0, 1, 2\}$) given evidence $\boldsymbol{X}_{1:t}$ collected up to time t. We overload this notation to represent ratios between sets of classes. For example, $S_t^{1,\bar{1}}$ means the log posterior ratio of class 1 versus 'not 1', which contains class 2 and class 0.

The other important statistics is the log posterior ratio between the coherent motion classes (class 1 and 2) assuming that *the change has occurred* at time t (derived in **Eq. 5.13**):

$$R_{t_\delta,t} \triangleq \log \frac{P(C_t = 1|\boldsymbol{X}_{t_\delta:t})}{P(C_t = 2|\boldsymbol{X}_{t_\delta:t})}, \tag{5.2}$$

where the log posterior ratio is conditioned on observations in the time interval $[t_\delta, t]$, e.g. from the time of change t_δ and a later time t.

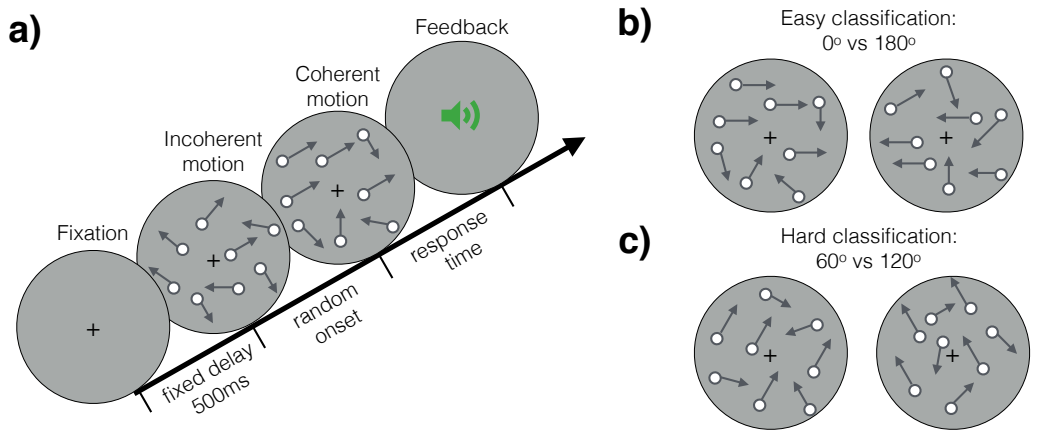

Figure 5.1: **Random dot motion discrimination with unknown stimulus onset**. (**a**) Stimulus setup. A trial begins with a central fixation cross. After $500ms$ a display of dots moving incoherently in all directions is displayed. After a random delay t_δ, a fraction z of the dots start moving coherently along one of two directions $\{\theta_1, \theta_2\}$. As quickly as possible the subject presses on a button to indicate the direction of motion. The trial ends with auditory feedback. The direction of coherent motion controls the relative difficulty between classification and detection. (**b**) Stimulus for coherent motion 0° and 180° (classification is easier than detection). (**c**) Stimulus for coherent motion 60° and 120° (detection easier than classification).

Both log posterior ratio statistics may be computed directly from the firing patterns of motion-tuning neurons in MT, to be discussed in the MT front-end section (e.g. **Eq. 5.10** and **Eq. 5.13**). Based on these statistics, we present three plausible models for the dual task.

Classify then Detect (CD)

The first system (**Fig. 5.2a,b**) is based on the posterior probability ratio of classes 1 and 2. The system employs two accumulators $S_t^{c,\bar{c}}$, one for each class $c \in \{1, 2\}$, that race to reach a threshold τ_{dis}. The class of the winner is the predicted class $\hat{C}$. Since in our tasks the two classes are completely symmetrical the same threshold τ_{dis} is set for both accumulators. Distinct thresholds may be necessary in asymmetric scenarios (e.g. one class is more frequent than the other). Let t_d denote the time of decision ($t_d - t_\delta$ is the reaction time). The CD procedure is:

$$\begin{aligned} t_c &= \text{first time } t \text{ that } S_t^{c,\bar{c}} > \tau_{dis}, \qquad c \in \{1, 2\} \\ \hat{C} &= \arg\min_{c \in \{1,2\}} t_c, \qquad t_d = t_{\hat{C}}. \end{aligned} \tag{5.3}$$

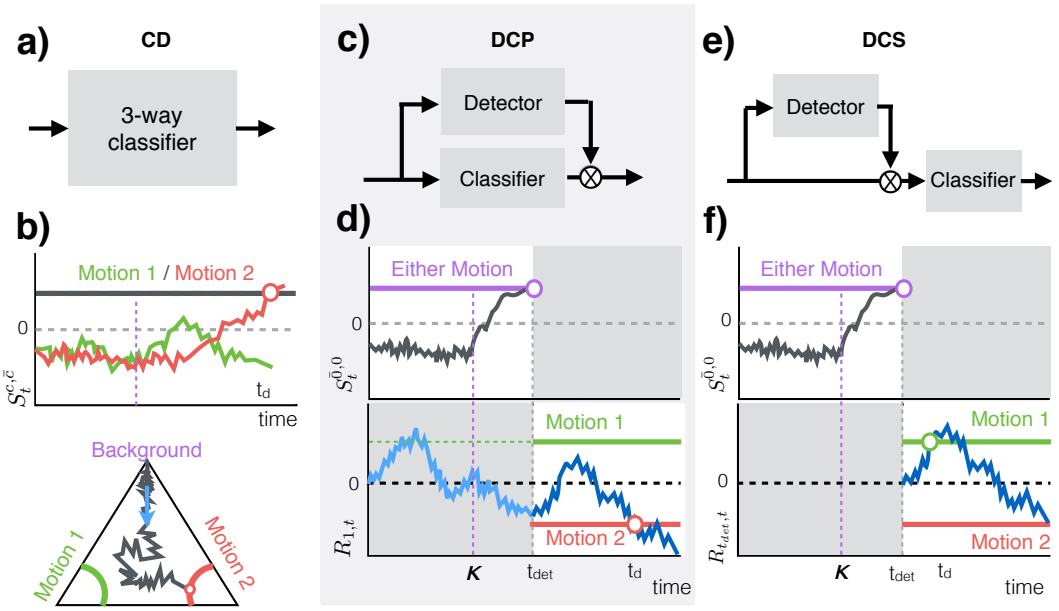

Figure 5.2: **Three models of joint detection and discrimination**. (**a**) **CD** – The world is assumed to be in one of three states: initial state (incoherent motion) and two post-change states (motion 1 and motion 2). The posterior probability of each state is computed. (**b**) (Top) In CD, log posterior ratios $S_t^{1,\bar{1}}$ and $S_t^{2,\bar{2}}$ (**Eq. 5.1**) race to a common discrimination threshold. (Bottom) An equivalent depiction showing the trajectory of the posterior over time (time direction indicated by the blue arrow) visualized in the probability simplex. When the posterior reaches one of the two lower corners the system declares motion 1 (left corner) or 2 (right corner). (**c**) **DCP** – A detector for coherent motion and a classifier of motion direction are computed in parallel. (**d**) The detector computes the log posterior ratio $S_t^{\bar{0},0}$ of any coherent motion vs. incoherent motion. The classifier computes the log posterior ratio $R_{1,t}$ of motion 1 vs. motion 2, which races towards a pair of thresholds (upper for motion 1, lower for motion 2). Until the detector fires at t_{det}, the classifier cannot fire (despite crossing dashed green threshold). After t_{det}, the classifier carries over signals prior to t_{det}. (**e**) DCS employs a detector and a classifier in series. (**f**) The classifier starts only after the detector fires and does not retain information prior to t_{det}. This lossy integration causes DCS to make a different (wrong in this example) decision than DCP.

This procedure is a Bayesian version of the multi-class CUSUM procedure [13] and proven optimal by [16]. Here optimality means that given a requirement on the false detection rate and the misclassification rate, the procedure above achieves the shortest response time on average.

Detection and Classification in Parallel (DCP)

The second model (**Fig. 5.2c,d**) separately and simultaneously performs detection

and discrimination. A detector performs a one-sided test on $S_t^{\bar{0},0}$, the log posterior ratio of 'coherent motion' (state 1 and 2) against state 0 of incoherent motion, to detect whether any coherent motion is present. Meanwhile, running in the background is a classifier that is concerned only with distinguishing between the two coherent motion classes $R_{1,t}$. The classifier is suppressed from firing until the detector fires at time t_{det}.

The decision process is parameterized by the threshold τ_{det} for detection, and the threshold τ_{dis} for classification. Again the discrimination threshold τ_{dis} is shared between classes for simplicity.

$$\begin{aligned} t_{det} &= \text{first time } t \text{ that } S_t^{\bar{0},0} > \tau_{det} \\ t_c &= \text{first time } t \geq t_{det} \text{ that } R_{1,t} > \tau_{dis}, c \in \{1,2\} \\ \hat{C} &= \arg\min_{c\in\{1,2\}} t_c, \qquad t_d = t_{\hat{C}} \end{aligned} \tag{5.4}$$

Here the detector and classifier run in parallel, and the detector functions as a gate that guards the classifier against fluctuations. Both the detector and the classifier are *lossless* in information integration, but the classifier is used sub-optimally since the information accumulated prior to stimulus onset is invalid.

> The DCP model may seem redundant as it is not optimal. It is included because of reverse compatibility and model complexity. First, DCP contains specialized and optimal components for detection and classification, respectively. By selecting the corresponding component DCP can solve pure detection or pure discrimination tasks. Second, DCP contains two independent decision thresholds, making it as complex as DCS (next subsection). Therefore any performance discrepancy between the two is directly attributable to model biases, not complexity.

Detection and Classification in Series (DCS)

In the third model, the last we consider, model detection and classification proceed in succession (**Fig. 5.2e,f**). After the detector identifies a coherent motion at time t_{det}, the classifier comes online assuming that the change has already happened ($t_\delta \leq t_{det}$). This assumption reduces the problem to pure classification starting at time t_{det}, which may be solved by the classical sequential probability test (SPRT [4]).

$$
\begin{aligned}
t_{det} &= \text{first time } t \text{ that } S_t^{\bar{0},0} > \tau_{det} \\
t_c &= \text{first time } t \text{ that } R_{t_{det},t} > \tau_{dis}, c \in \{1,2\} \\
\hat{C} &= \arg\min_{c\in\{1,2\}} t_c, \qquad t_d = t_{\hat{C}}
\end{aligned} \tag{5.5}
$$

DCS essentially concatenates the optimal detector (CUSUM) and the optimal classifier (SPRT) in time. It is *lossy* and potentially slower because the classifier does not consider any evidence before the detector fires. However it also provides *modularity*, as it completely separates the detection problem from discrimination. Mathematically, the subtle difference between DCS (Equation set 5.5) and DCP (Equation set 5.4) is the lossy evidence accumulation by the classifier. DCS discards all observations prior to detector firing, while DCP maintains them. Therefore, comparing DCP and DCP allows us to understand whether the detector functions as a gate or a trigger.

Quantized sensory input

To apply the aforementioned models on the random dot motion detection and discrimination task, we need to compute the log posterior ratios between pairs of classes (**Eq. 5.1** and **Eq. 5.2**). We chose a probabilistic strategy based on a front-end of direction tuning neurons. The front-end converts a visual stimulus (a length-Δ video segment of dots moving in space) into a set of action potentials, which are interpreted probabilistically to produce the log posterior ratios, as we see below.

$$
\lambda_\theta^k \triangleq \lambda_{min} + (\lambda_{max} - \lambda_{min}) \exp\left(-\frac{1}{2}\frac{\|\theta_k - \theta\|^2}{\sigma_Y^2}\right), \tag{5.6}
$$

where λ_{max} and λ_{min} are the maximum and minimum firing rates of a neuron (in Hz), and σ_Y is the tuning width, and the notation $\|\theta_k - \theta\|$ indicates the minimal angular distance between θ_k and θ.

> Here we have chosen a Gaussian tuning curve. This choice is not critical, e.g. a von Mises function [25] works equally well.

Within a unit interval $((t-1)\Delta, t\Delta]$, the number of spikes $X_{i,t}^k$ emitted from neuron k at location i in response to motion θ is (**Fig. 5.3b**):

$$
P(X_{i,t}^k = n) = Poisson(n|\lambda_\theta^k \Delta). \tag{5.7}
$$

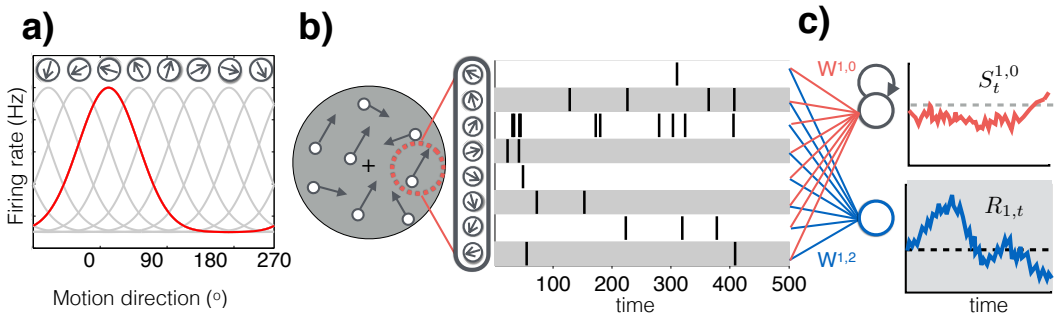

Figure 5.3: **From front-end MT neurons to log posterior ratios**. (**a**) Turing curves of a hypercolumn of 8 MT motion-tuning neurons (**Eq. 5.6**). (**b**) A hypercolumn of neurons activate in response to random dot motion at their receptive field (red dashed circle). Raster plot shows simulated homogeneous Poisson spike trains (**Eq. 5.7**) with max rate $\lambda_{max} = 20Hz$. (**c**) Two downstream neurons compute log posterior ratios from the spikes. $S_t^{1,0}$, the log posterior ratio between motion 1 and incoherent motion, may be computed by adding $W^{1,0}$-weighted spikes to a recurrent unit (**Eq. 5.10**). $R_{1,t}$, the log posterior ratio motion 1 and 2 after change onset, is a linear combination of the spike trains weighted by $W^{1,2}$ (**Eq. 5.13**). $W^{1,0}$ and $W^{1,2}$ depend on the motion coherence and motion directions, and are given in closed-form (**Eq. 5.9** and **Eq. 5.12**).

Log posterior ratios for detecting coherent motion from spikes

Consider a visual display with M moving dots. The dots are spaced sufficiently far apart such that each dot is monitored by a unique hypercolumn. At any point in time, a random fraction z ('coherence', $0 \leq z \leq 1$) of the M dots are moving along the same direction, and the remaining along random directions. Let $\bar{\lambda} \triangleq \mathbb{E}_\theta[\lambda_\theta^k]$ be a neuron's average firing rate over all stimulus directions. $\bar{\lambda}$ should be roughly identical for all neurons thanks to symmetry.

The log likelihood ratio $r_t^{c,0}$ between z fraction of coherent motion along direction θ_c of class $c \in \{1, 2\}$ and incoherent motion (class 0) is given by (derived in Methods **Eq. A.37**):

$$r_t^{c,0} \triangleq \log \frac{P(X_t|C_t = c)}{P(X_t|C_t = 0)} = \sum_i \sum_k W_k^{c,0} X_{i,t}^k, \tag{5.8}$$

$$\text{where} \quad W_k^{c,0} \triangleq \log \frac{(1-z)\bar{\lambda} + z\lambda_{\theta_c}^k}{\bar{\lambda}}. \tag{5.9}$$

The log posterior ratio $S_t^{c,0}$ between coherent motion in θ_c and incoherent motion may be computed recursively as (see Methods **Eq. A.46** for detailed derivations)

$$S_t^{c,0} = \mathcal{S}\text{rec}\left(S_{t-1} - \log \alpha_t\right) + \log \frac{\alpha_t}{1-\alpha_t} + r_t^{c,0}, \tag{5.10}$$

where $\mathcal{S}\mathrm{rec}\,(x) \triangleq \log(1+\exp(x)) \approx \max(0,x)$ is the 'soft-rectifier' function and α_t is the probability of a change happening now knowing that it has definitely not happened prior to t: $\alpha_t \triangleq P(t_\delta = t | t_\delta \geq t)$. The initial condition is $S_0 \triangleq \log \frac{P(t_\delta \leq 0)}{P(t_\delta > 0)} = \log 0 = -\infty$. See figure **Fig. 5.4** for an example of $S_t^{c,0}$.

The hardmax approximation gives an intuition for $S_t^{c,0}$. If past evidence util $t-1$ suggests that the likelihood for coherent motion is so low that $S_{t-1}^{c,0} - \log \alpha_t < 0$, then the system should "forget" about past evidence and reset to the log prior ratio $\log \frac{\alpha_t}{1-\alpha_t}$ instead. $\log \alpha_t$ thus is a threshold for triggering the forgetting mechanism. The forgetting mechanism allows $S_t^{c,0}$ to discard noisy observations in the distant past while taking in new evidence into consideration. For example, in **Fig. 5.4** we simulate the log posterior ratio $S_t^{c,0}$ for exponentially distributed change time. $S_t^{c,0}$ behaves almost like a memoryless system before the change occurs, which is crucial for an organism to detect changes whose arrival time spans a long duration.

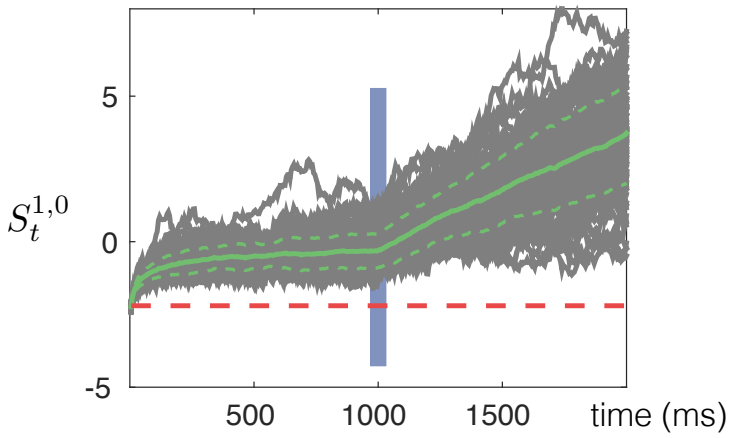

Figure 5.4: **The log posterior ratio $S_t^{c,0}$ for detecting coherent motion**. Log posterior ratio $S_t^{1,0}$ over 100 trials of Monte-Carlo simulations with mean and standard deviation overlaid. Unbeknown to the model, stimulus onset in all trials is $t_\delta = 1$ sec. The model instead uses an exponential prior for t_δ. The dash line shows the log prior ratio $\log \frac{\alpha_t}{1-\alpha_t}$. As we see from **Eq. 5.10**, the log prior ratio gives a lower bound for $S_t^{1,0}$ and cues the observer when to 'pay attention' and when to let go the past. $r_t^{1,0}$ is modeled by a Gaussian random walk $r_t^{1,0} \sim \mathcal{N}(\mu_C \Delta, \sigma^2 \Delta)$ with $\mu_C = \pm 14$ and $\sigma = 3.5$.

Log posterior ratios for classifying coherent motion from spikes

Using the same analysis we can compute the log likelihood ratio $r_t^{1,2}$ of the observation at time t between two different directions of coherent motion, θ_1 and θ_2, at

coherence level z.

$$r_t^{1,2} \triangleq \log \frac{P(X_t|C_t=1)}{P(X_t|C_t=2)} = \sum_i \sum_k W_k^{1,2} X_{i,t}^k, \tag{5.11}$$

$$\text{where} \quad W_k^{1,2} \triangleq \log \frac{(1-z)\bar{\lambda} + z\lambda_{\theta_1}^k}{(1-z)\bar{\lambda} + z\lambda_{\theta_2}^k}. \tag{5.12}$$

The log posterior ratio between coherent motions conditioning on post-change evidence is (derived in Methods **Eq. A.48**):

$$R_{t,t'} = \sum_{i=t}^{t'} r_i^{1,2}. \tag{5.13}$$

We assume even class priors $P(C=1) = P(C=2)$. Uneven prior may be incorporated by a simple shift of $R_{t,t'}$. See figure **Fig. 5.3c** for an example of $R_{t,t'}$.

> The expressions of the log posterior ratios in **Eq. 5.10** and **Eq. 5.13** suggest straightforward spiking implementations. The mechanisms are similar to those discussed in **Sec. 3.6**.

5.3 Psychophysics

Design

To test which of the proposed models (CD, DCP, DCS) best matches human detection and discrimination behavior, we recruited human subjects to participate in two experiments. Both experiments employed a dynamic random-dot display [2], where white dots were randomly distributed on a black background. All dots moved randomly except for a random fraction z that moved along a consistent direction. The direction could be one of two directions θ_1 and θ_2. The average of the two directions was always 90°, so we chose to represent them using the direction discrepancy $\Delta\theta = |\theta_2 - \theta_1|$. See **Fig. 5.1a**-c.

> Details of the display: the random dots with a density of $16.7 dots/deg^2/s$ were displayed with a 5° diameter circular aperture about the fixation center. Each dot was a white square of 5×5 pixels (0.14°). For the stimulus, on each video frame the coherently moving dots were shifted 0.125° from their positions $25ms$ earlier (three video frames, refresh rate = $120Hz$), corresponding to a speed of $5°/s$, while others were randomly repositioned.

The first experiment was dual detection and classification (**Fig. 5.1a**). The stimulus motion started incoherent ($z = 0$) and changed to one of two coherent directions

($z > 0$) after a stochastic delay t_δ. t_δ followed an exponential distribution with a mean of $800ms$. Subjects indicated the direction of coherent motion by button-press. Responses earlier than t_δ were considered false detections. Subjects were instructed to minimize both misclassification errors and response time while maintaining the false detection rate below 20%.

The second experiment was pure detection. With the identical setup as the dual experiment, here subjects were instructed to press a button as soon as they perceived coherent motion regardless of motion direction. The goal was to minimize response time while keeping the false detection rate below 20%.

We systematically varied the coherence level z and the direction discrepancy $\Delta\theta$ for each experiment. z is chosen randomly from $\{1.6\%, 3.2\%, 6.4\%, 12.8\%, 25.6\%\}$ and $\Delta\theta$ from $\{180°, 60°\}$. Both z and $\Delta\theta$ were fixed within a block of consecutive trials and varied between blocks (i.e. the subjects know the coherence level z and $\Delta\theta$).

Model fitting

We fit each model to the data that was collected in both experiments. The models were parameterized by (1) signal-to-noise ratio of the front-end and (2) decision thresholds. The front-end had four parameters: the minimum and maximum firing rates λ_{min} and λ_{max}, the tuning width w, and the number of neurons N. We fixed $\lambda_{min} = 1Hz$ and $w = 25°$ according to their physiological values in the macaque monkey [22]. Since N and λ_{max} have similar effects on the signal-to-noise ratio we did not fit both; rather, we fixed $N = 16$ neurons, and only fit the maximum firing rate λ_{max}. (2) CD only has one decision threshold τ_{dis}, whereas both DCP and DCS have two thresholds, τ_{det} and τ_{dis}, for their detector and classifier components, respectively.

We selected λ_{max} and the threshold(s) of each model and each subject to maximize the model prediction's agreement with the data in terms of the median response time, the misclassification rate and the false detection rate. The same λ_{max} parameter was used across different "conditions", parameterized by the coherence level z, the motion discrepancy $\Delta\theta$ and the experiment type (dual versus detection only). This parameter-sharing was made possible by the generalizability of our front-end model. By contrast, we did not share the thresholds, yielding one threshold in CD and two in DCP and DCS for each condition.

Human decision-making involves a perceptual component (evidence accumulation

to decision threshold) and a non-perceptional delay (axonal propagation, motor delays, etc). Our model only accounts for the perceptual component. The non-perceptional component was modeled phenomenologically with a log-normal distribution with two additional parameters (mean and variance) per subject.

Fitting results

The fits for the response time, misclassification and detection errors of a randomly sampled subject are shown in **Fig. 5.5** and **Fig. 5.6**. All three models qualitatively explain subjects' performance, although DCS is the most faithful to the error data. Despite the parsimonuous parameterization, the models predict the key performance metrics (**Fig. 5.5**) as well as the full response time histograms (**Fig. 5.6**) of the subjects. The overall scores of fitting the dual task and for fitting both tasks are in shown **Fig. 5.8a** and b. Both plots show significantly higher fitting errors of the optimal CD model compared to the sub-optimal models, within which the DCS perform better. This trend is also consistent across all 10 subjects except one (subject JD).

To further separate the two sub-optimal models, we also visualized the posterior estimates of the parameters in **Fig. 5.7.** The posterior weights the parameter values according to their agreement with the data. The signal-to-noise parameter λ_{max} estimate is correlated with the mean non-perceptual delay. This is not surprising as higher signal-to-noise ratio means shorter perceptual times, which leaves a shorter time to be explained by the non-perceptual delay. We see qualitatively that DCP produces less consistent estimates between the two experiments than does DCS. A quantitative comparison in **Fig. 5.8c** confirms that DCS is significantly more consistent across all subjects.

5.4 Discussion and summary

Plausible model for human behavior

We proposed three candidate mechanisms for joint detection and discrimination, and we explored whether any of them can account for human performance. Our first observation is that the optimal model CD underperforms the sub-optimal models in explaining human behaviors. This trend is significant and consistently observed in the dual tasks and joint fitting of both the dual and the detection tasks. However, the discrepancy between CD and the other models may be a consequence of the degree of freedom (DoF), as CD only has one threshold while DCP and DCS each independently manipulate two thresholds. To remove DoF as a confound, we re-

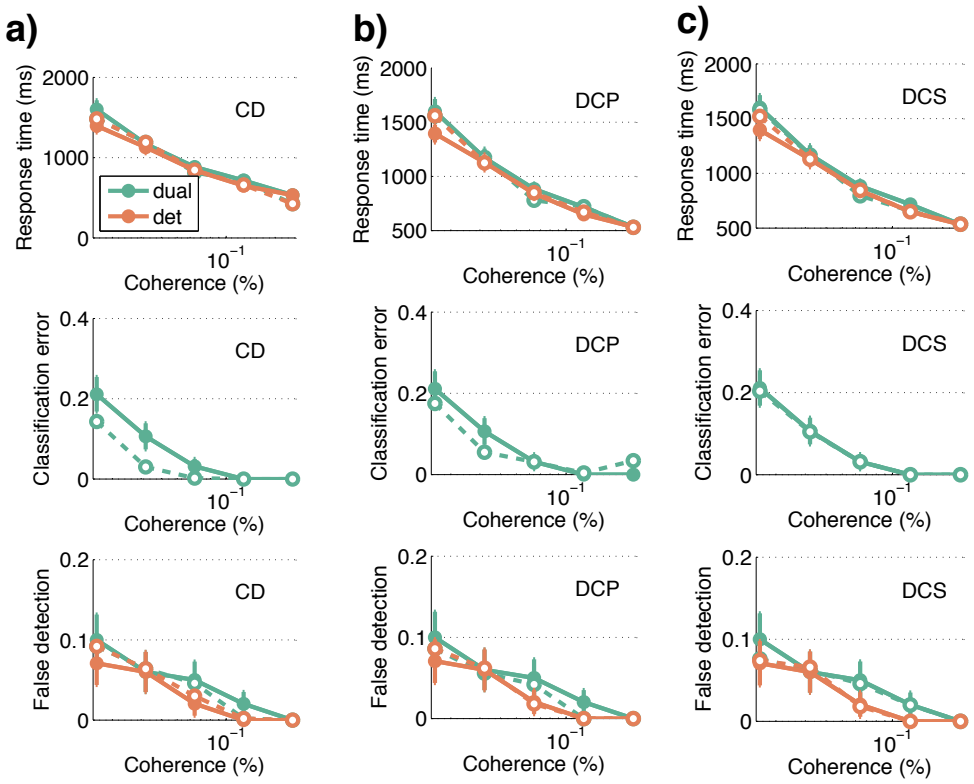

Figure 5.5: **Fitting results for a randomly selected subject**. The median response times, misclassification errors and false detection rates of a random subject (BW) and the fitted model predictions in the dual task and the detection task. The columns represent the three methods (CD, DCP and DCS). Solid lines show subject's data with 1 ste and dashed lines show the predictions. The direction discrepancy is $\Delta\theta = 180°$.

fitted the experiments while manipulating the parameter sharing across conditions. Even with more free parameters per condition, CD is less consistent with the data than the sub-optimal models are. Therefore, CD may not be the strategy of choice for humans.

DCS and DCP have the same DoF, hence may be compared fairly. The data suggests that DCS may be closer to the strategy for humans. A first clue is that DCS outperforms DCP in both fitting experiments across all subjects (**Fig. 5.8a** and b). A second cue comes from comparing the posteriors (**Fig. 5.7**). In the pure detection experiment, DCS and DCP reduce to the same algorithm, hence their parameter estimates should also be *the same*. In the dual task, however, DCP and DCS should produce *different* estimates, as explained below. DCS discards information prior to detector activation and thus requires longer evidence accumulation to achieve the

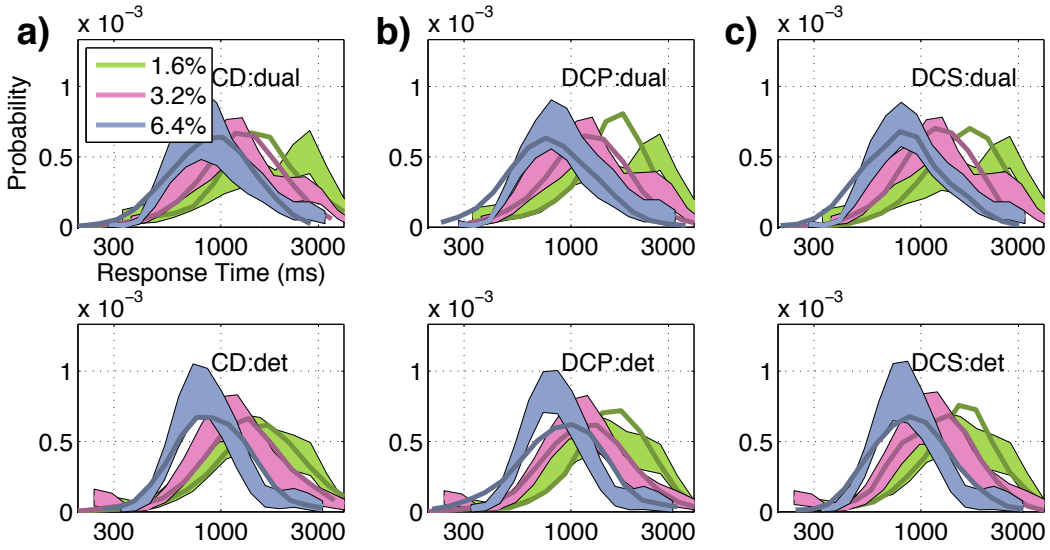

Figure 5.6: **Fitting results for a random subject (cont'd)**. The response time histograms and fits for the dual task (first row) and the detection task (second row) of a randomly selected subject (BW). Only a subset of the data, with coherence levels $\{1.6\%, 3.2\%, 6.4\%\}$ and direction discrepancy $\Delta\theta = 180°$, are shown. Each color denotes a different coherence level. The solid lines are fits from the model and the filled regions show the mean ± 1 bootstrapped standard error of the subject's data. Each column shows the fit of one model (CD, DCP and DCS).

same level of accuracy as does DCP. Therefore, to explain the same data DCS must compensate with a higher λ_{max} estimate, a shorter motor delay estimate, or both. As a result, comparing the posteriors between the two tasks will expose the incorrect model. In the case of **Fig. 5.7a**nd **Fig. 5.8c**, we see that DCS produces consistent parameter estimates, while DCP does not and should therefore be eliminated.

Sub-optimal information processing

Our analysis suggests that humans are sub-optimal in the dual detection-decision task; however, this conclusion is not inconsistent with previous findings [11], [17]–[19] that the human visual system is near-optimal in evidence accumulation. In the DCS model sub-optimality resides in the decision strategy rather than in evidence accumulation.

We speculate that the human visual system may use the DCS strategy for two reasons. 1) Modularity. DCS may be adapted to tackle a pure detection tasks by simply setting the classification threshold τ_{dis} to zero. Similarly, setting $\tau_{det} = -\infty$ would tune DCS for a pure classification task. The flexibility to switch between tasks is a desirable property. For instance, sound cues may sometimes permit detection,

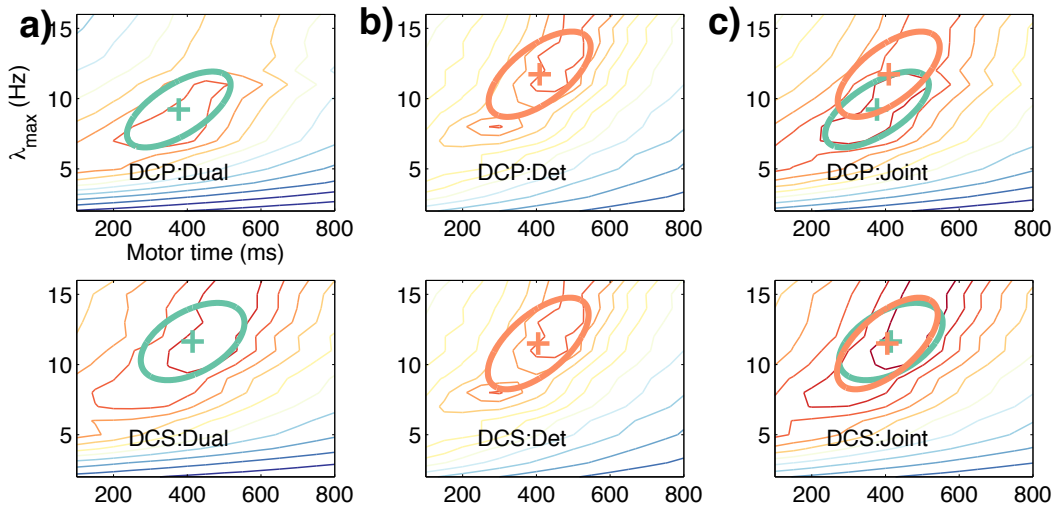

Figure 5.7: **Posterior distribution of parameters for a random subject**. The posterior of the signal-to-noise parameter (λ_{max}) and the non-perceptual delay parameter (motor time) for DCP (first row) and DCS (second row) for a random subject (BW). The three columns represent posteriors obtained from (**a**) the dual task, (**b**) the detection task, and (**c**) both tasks. For each panel in (a) and (b) the ellipse and the cross represent a Gaussian approximation to the posterior and its mean. In (c) the two ellipses from (a) and (b) are superimposed.

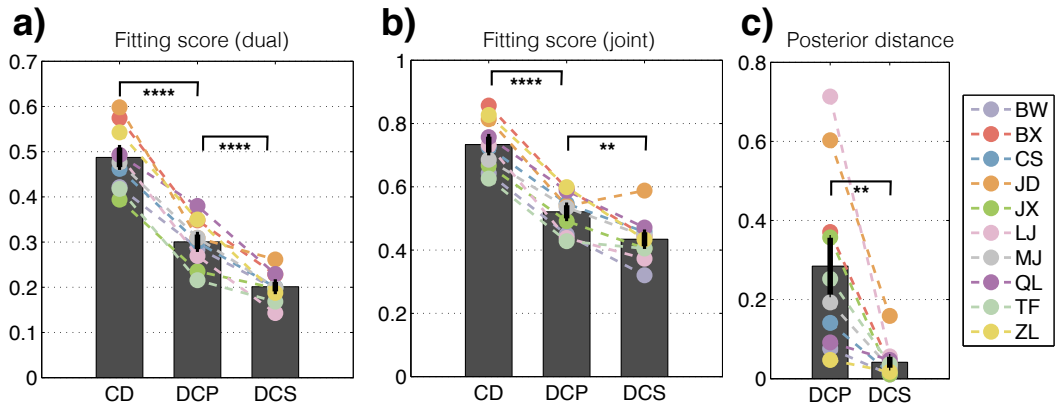

Figure 5.8: **Fitting performance**. (**a**) Fitting scores (lower means better) of the three models CD, DCP and DCS on the dual task. (**b**) Same as (a) except that the score is computed jointly over the dual and the detection-only task. (**c**) Distances (lower means better) between the posterior obtained from the dual task and that from both tasks. Colored dashed lines show performance for different subjects (see legend). Bars show average performance over 10 subjects with 1 ste. '**' represents $p \leq 0.01$ and '****' represents $p < 10^{-4}$.

which renders visual detection unnecessary. As a result, the visual system would need to switch from the dual task mode to pure classification mode. 2) Power

efficiency. While the classifiers in CD are bombarded with sensory inputs all the time, the classifier in DCS only activates for brief moments when a change in the environment has been confirmed. In other words, the classifier in DCS may be dormant most of time to conserve energy. This advantage may be more pronounced for discrimination tasks with a large number of categories, as the relative energy reduction from CD to DCS is proportional to the number of categories. On the other hand, in situations when changes happen frequently and only a small number of classes are involved, CD may be a viable strategy for humans.

References

[1] W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations.," *Psychological Review*, vol. 96, no. 2, pp. 267–314, 1989.

[2] W. T. Newsome, K. H. Britten, and J. A. Movshon, "Neuronal correlates of a perceptual decision.," *Nature*, 1989.

[3] J. Palmer, A. C. Huk, and M. N. Shadlen, "The effect of stimulus strength on the speed and accuracy of a perceptual decision," *Journal of Vision*, vol. 5, no. 5, pp. 376–404, 2005.

[4] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[5] R. Ratcliff, "Theoretical interpretations of the speed and accuracy of positive and negative responses.," *Psychological Review*, vol. 92, no. 2, p. 212, 1985.

[6] D. J. Lasley and T. Cohn, "Detection of a luminance increment: Effect of temporal uncertainty," *Journal of the Optical Society of America*, vol. 71, no. 7, pp. 845–850, 1981.

[7] H. Pashler, "Familiarity and visual change detection," *Perception & Psychophysics*, vol. 44, no. 4, pp. 369–378, 1988.

[8] J. Hohnsbein and S. Mateeff, "The time it takes to detect changes in speed and direction of visual motion," *Vision Research*, vol. 38, no. 17, pp. 2569–2573, 1998.

[9] C. W. Clifford and M. Ibbotson, "Fundamental mechanisms of visual motion detection: Models, cells and functions," *Progress in Neurobiology*, vol. 68, no. 6, pp. 409–437, 2002.

[10] E. P. Cook and J. H. Maunsell, "Dynamics of neuronal responses in macaque mt and vip during motion detection," *Nature Neuroscience*, vol. 5, no. 10, pp. 985–994, 2002.

[11] C. M. Glaze, J. W. Kable, and J. I. Gold, "Normative evidence accumulation in unpredictable environments," *Elife*, vol. 4, e08825, 2015.

[12] W. Shewhart, "The application of statistics as an aid in maintaining quality of a manufactured product," *Journal of the American Statistical Association*, vol. 20, no. 152, pp. 546–548, 1925.

[13] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[14] Y. Ritov, "Decision theoretic optimality of the cusum procedure," *The Annals of Statistics*, pp. 1464–1469, 1990.

[15] M. Pollak and D. Siegmund, "Approximations to the expected sample size of certain sequential tests," *The Annals of Statistics*, pp. 1267–1282, 1975.

[16] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, pp. 1897–1908, 1971.

[17] W. J. Ma, V. Navalpakkam, J. M. Beck, R. Van Den Berg, and A. Pouget, "Behavior and neural basis of near-optimal visual search," *Nature Neuroscience*, vol. 14, no. 6, pp. 783–790, 2011.

[18] B. Chen and P. Perona, "Speed versus accuracy in visual search: Optimal performance and neural architecture," *Journal of Vision*, vol. 15, no. 16, pp. 9–9, 2015.

[19] J. Drugowitsch, G. C. DeAngelis, D. E. Angelaki, and A. Pouget, "Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making," *ELife*, vol. 4, e06678, 2015.

[20] E. P. Simoncelli, L. Paninski, J. Pillow, and O. Schwartz, "Characterization of neural responses with stochastic stimuli," *The Cognitive Neurosciences*, vol. 3, pp. 327–338, 2004.

[21] M. Jazayeri and J. A. Movshon, "Optimal representation of sensory information by neural populations," *Nature Neuroscience*, vol. 9, no. 5, pp. 690–696, 2006.

[22] J. H. Maunsell and D. C. Van Essen, "Functional properties of neurons in middle temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed, and orientation," *Journal of neurophysiology*, vol. 49, no. 5, pp. 1127–1147, 1983.

[23] M. Shadlen and W. Newsome, "Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey," *Journal of Neurophysiology*, vol. 86, no. 4, pp. 1916–1936, 2001.

[24] J. Drugowitsch, R. Moreno-Bote, A. Churchland, M. Shadlen, and A. Pouget, "The cost of accumulating evidence in perceptual decision making," *The Journal of Neuroscience*, vol. 32, no. 11, pp. 3612–3628, 2012.

[25] B. Amirikian and A. P. Georgopulos, "Directional tuning profiles of motor cortical cells," *Neuroscience Research*, vol. 36, no. 1, pp. 73–79, 2000.