Multisensory Perceptual Discrimination in Evolved Networks and Agents

Marieke Rohde[1]

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
marieke.rohde@tuebingen.mpg.de

Abstract

The fact that humans and animals have several sensory modalities and use them together to make sense of the world imbues their behaviour with an immense richness and robustness. In this study, recurrent neural networks and minimal agents with active vision are evolved for a perceptual discrimination task (unimodal and bimodal). The purpose of this study is mainly exploratory: to test which of the characteristics of human perceptual discrimination evolve easily (with a focus on statistically optimal integration), how they are realised and what active perception does in this process. Whilst some of the systems evolved to perform perceptual discrimination well, they did not conform to the predictions from statistical optimality. Analyses of the systems point towards a number of relevant issues, noticeably towards the lack of a good account of 'unimodality' in existing models of multisensory perception.

Introduction

Humans and animals use several sensory modalities to make sense of the world and to judge on and distinguish objects in the environment. For instance, the size of an object can be judged both by touching the object or by looking at it, or by doing both at the same time. In humans, it could be shown that subjects, when estimating object size, integrate visual and tactile cues in a statistically optimal fashion to decrease uncertainty (Ernst and Banks, 2002). Similar findings were reported from other multisensory tasks, e.g., audio-visual sound localization (Alais and Burr, 2004).

These kinds of results are usually obtained using a psychophysics approach, where subjects are asked to perform perceptual judgments on stimuli that are varied systematically along a physical dimension. Comparing the human behaviour to that of an 'ideal observer' using maximum likelihood estimation (MLE), the mentioned findings of optimality are derived. This approach is *prima facie* behaviour-based; the underlying mechanisms of (optimal) multisensory integration are not yet well understood. Under the dominant representationalist paradigm, we would expect a dedicated internal neural mechanism to implement MLE. Accordingly, Knill and Pouget (2004) rephrase the problem of statistically optimal multisensory integration as follows: "(i) how do neurons, or rather populations of neurons, represent uncertainty, and (ii) what is the neural basis of statistical inferences?" and review candidate neural correlates.

By contrast, Artificial Life and dynamical approaches in cognitive science have repeatedly shown that efficient, robust or plausible models exist that do not rely on local computation but on agent morphology, contingencies in agent-environment interaction or on non-linear dynamics in neural control. Examples of such models in perception research include active vision to solve a non-Markovian visual discrimination task with feed-forward control (Floreano et al., 2004; Izquierdo-Torres and Di Paolo, 2005), agency detection by emergent behavioural coordination (Di Paolo et al., 2008) or olfactory perception through chaotic neural dynamics (Freeman, 1987). These models do not just point out alternatives, they also show that, if global dynamics are taken into consideration, many phenomena that appear complex emerge effortlessly.

For the study presented, recurrent neural network controllers and minimal agents with an active vision system were evolved to solve a size discrimination task. Such an evolutionary robotics (ER) approach has been argued to minimise prior assumptions about underlying mechanisms by outsourcing the design to an automated search procedure (Harvey et al., 2005). The purpose was mainly exploratory: if no constraints of optimality are imposed, which, if any of the hallmarks of MLE optimal integration evolve? How do the systems realize perceptual discrimination? How do they integrate their senses and how do they deal with varying levels of uncertainty? Comparing a disembodied network and an embodied agent, what are the differences and commonalities? Are there advantages associated with active perception in this task?

The results presented can be seen as work in progress. They point out issues that require a rethinking of the approach taken here. While some of these difficulties are of a more technical nature, others proved to be insightful with respect to the overarching question of (optimal) multisensory integration. In particular, the question of what unimodality means in a system with several sensory channels is of

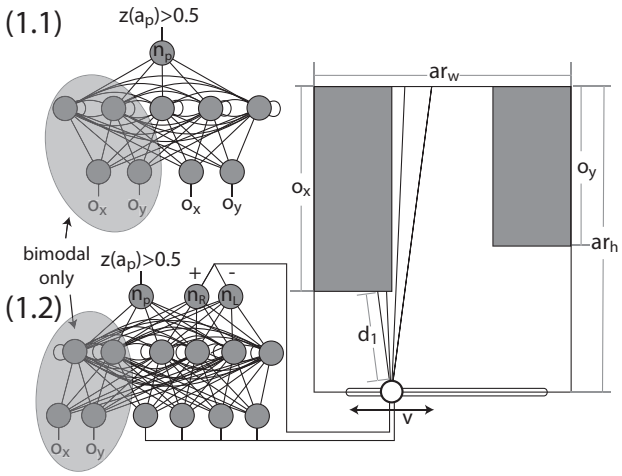

Figure 1: Evolved networks for the direct condition (1.1) and for the active vision condition (1.2).

potential importance for the study of multisensory integration in general. The results confirm that emphasizing the non-obvious is one of the key characteristics and merits of generative ER modelling.

Methods

Simulation and Genetic Algorithm

Continuous-time recurrent neural networks (CTRNNs; e.g., Beer, 2003) are evolved to solve a two-alternative forced-choice (2AFC) size discrimination task. The decision, which of two objects $o_x, o_y \in [1, 2.5]$ is larger is either generated by an agent controlled by a CTRNN or by a CTRNN directly. The dynamics of units in a CTRNN is governed by

$$\tau_i \frac{da_i(t)}{dt} = -a_i(t) + \sum_{j=1}^{N} w_{ij} z(a_j(t) + \theta_j) + I_i(t) \quad (1)$$

where $z(x)$ is the standard sigmoidal function $z(x) = 1/(1 + e^{-x})$, $a_i(t)$ is the activation of unit i at time t, θ_i is a bias term, τ_i is the activity decay constant, w_{ij} is the strength of a connection from unit j to unit i. The structure of the network is partially layered, network sizes vary between conditions (see Fig. 1). Neural and environmental dynamics were simulated using the forward Euler method with a time step of $h = 1ms$.

For all controllers, input signals are fed into input units n_i by $I_i(t) = Sg_i \cdot inp + \nu\epsilon$, where Sg_i is the evolved sensory gain, inp is the input signal, ϵ is a normally distributed random variable and $\nu \in [0, 3, 6, 9, 12]$ is the level of sensory noise that modulates channel reliability across trials. In the network condition, the inputs $inp = o_x, o_y$ are fed directly into the network (see Fig. 1, 1.1). The active vision agent, inspired by (Beer, 2003), can move left and right by $v = Mg \cdot (z(n_l) - z(n_r))\ units/s$ in an arena of random width $ar_w \in [3.5, 4]$ and depth $ar_d \in [4.5, 5]$ (see Fig. 1, 1.2). The agent has a vision system comprised of four rays with angles $[-7.5°, -2.5°, 2.5°, 7.5°]$ and perceives distance by $inp_i = d_i/5$ where d_i is the distance at which a ray i is intercepted. All controllers are evolved for both a 'unimodal' and a 'bimodal' condition. In the bimodal condition, controllers are given a redundant direct input channel and two additional hidden units (see Fig. 1).

An output unit n_p generates a perceptual estimate: $z(a_p) > 0.5$ means a perceived $o_x > o_y$ at the end of a trial. This leads to the following performance criterion for pairs of objects (o_x, o_y)

$$P(o_x, o_y) = \begin{cases} 1 & \text{if } (z(a_p) > 0.5) = (o_x > o_y) \\ 0 & \text{else} \end{cases} \quad (2)$$

Fitness for individual controllers is computed according to

$$F = \frac{(1 - RB)}{16} \sum_{i=0}^{16} P(o_x, o_y) \cdot P(o_y, o_x) \quad (3)$$

where $o_x, o_y \in [1, 2.5]$ are drawn from a uniform distribution. As pairs are presented in both orders for F, evaluation involves $2 \times 16 = 32$ trials. The response bias $RB \in [0, 1]$ is proportional to the amount by which $z(a_p) > 0.5$ has a bias stronger than 75% to either side. The multiplicative term and the punishment for response bias were included after piloting because evolved systems tended to be very accurate but strongly biased towards one side. Object presentation lasts $T \in [3000, 4000ms]$ for networks ($+t_{pre} \in [100, 500ms]$ without stimulus) and $T \in [16000, 18000ms]$ for agents. Networks are initialised randomly and agents are positioned on the mid point of the line along which they can move.

CTRNNs are evolved using a generational GA with a population of 30 and are selected using truncation selection $(1/3)$. Genes are real-valued $\in [0, 1]$ with vector mutation $r \in [0.3, 0.5]$ and reflection at gene boundaries. Evolved gene values are linearly mapped onto the target range for $w_{ij} \in [-8, 8], \theta_i \in [-3, 3]$ and exponentially for $Sg \in [0.1, 20], Mg \in [0.1, 100]$ and $\tau_i \in [30, 3000ms]$ (networks) or $\tau_i \in [30, 10000ms]$ (agents) respectively. For the hidden and output layer, $\theta_i = -0.5 \sum_{j=0}^{N} w_{ij}$ (center-crossing).

ν is drawn randomly each trial from the available range of noise levels. Evolution starts noiseless (ν=0) and the maximum level of noise is increased every time average top performance over 50 generation exceeds $\bar{F} = 0.5$ till the full range ($\nu \in [0, 3, 6, 9, 12]$) is reached. In the bimodal condition, two quarters of the trials were unimodal trials (one quarter for each channel) to avoid specialization. This means that one modality received no signal but instead strong noise with $\nu = 15$. Otherwise, noise in the first channel was random as in the unimodal condition, whereas noise in the second channel was fixed at $\nu = 6$.

Analysis

Perceptual discrimination and integration is analysed just as in human psychophysics (e.g., Ernst, 2005). Perceptual response probability is described as a cumulative probability function ('psychometric curve') of real differences in object sizes. Evaluation is performed presenting a standard stimulus $o_s = 1.75$ to one side and a comparison stimulus $o_c \in [0.3o_s, 1.7o_s]$ to the other side. Each measurement is repeated 20 times. This procedure is repeated for both sides and for all levels of noise ν. Cumulative Gaussians are fitted to the responses using the Matlab toolbox psignifit for maximum likelihood fitting (Hill, 2005). The 50% level of a psychometric curve is called the PSE (point of subjective equality) and corresponds to the mean of the fitted Gaussian. It indicates perceptual bias. The difference between the 50% and the 84% is called the JND (just-noticeable-difference) and corresponds to $\sqrt{2}\sigma$ of the underlying Gaussian. It indicates perceptual accuracy.

Optimal integration is assessed by comparing the evolved system's perceptual discrimination with an ideal observer model using MLE and an independent channel model. In such a model, a bimodal perceptual estimate S^* is generated as a weighted sum of unimodal estimates (i.e., $S^* = w_1 S_1 + w_2 S_2$) in a way that minimizes uncertainty. MLE generates the following testable predictions (cf. Ernst, 2005; Ernst and Banks, 2002):

$$w_1 + w_2 = 1 \quad w_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + 1/\sigma_2^2} \quad \sigma^{*2} = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (4)$$

The first term indicates multisensory integration in general, whereas the second and third term are characteristic of optimal integration in particular. These criteria also clarify the significance of the noise level ν as the parameter that should modulate σ_i. According to the predictions, the weights w_i and σ^* should change with σ_i (in particular, bimodal discrimination should be more accurate than each of the unimodal discriminations).

To compute the weights, crossmodal conflicts $c \in [-.25o_s, .25o_s]$ are introduced during testing, i.e., for one modality $o_s^1 = o_s - 0.5c$ and for the other modality $o_s^2 = o_s + 0.5c$. Integration occurs if, in the presence of conflicts, PSEs are shifted along the $[o_s - 0.5c, o_s + 0.5c]$ interval according to the weights. σ_i can be computed by $JND = \sqrt{2}\sigma_i$.

Perceptual Discrimination in Recurrent Neural Networks

Evolving perceptual discrimination in recurrent neural networks is a less biased approach to the study of perceptual integration because it allows for the evolution of dynamically complex solutions and functional intertwinement: solutions evolved may not employ separate populations of neurons to perform different tasks, such as unimodal estimation, integration and measuring uncertainty. Also, given that the fitness function Eq. (3) does not require optimal integration, there is the possibility that optimality spontaneously emerges.

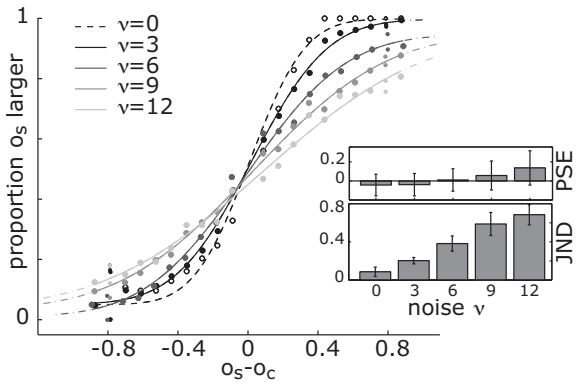

Figure 2: Unimodal networks. Psychometric curves for the different noise levels ν, data pooled from all 7 networks and both orders. Inlay: mean and s.e.m. for fitting parameters PSE (bias) and JND (accuracy) from individual fits (average of both stimulus orders; $N = 7$).

Unimodal Networks

The purpose of the unimodal condition was primarily to verify that the task is suitable for the study of perceptual discrimination. In order to allow the evolution of optimal integration, controllers have to perform perceptual discrimination sufficiently well. Their accuracy should decrease with the level of noise (JND should increase) to make it possible to test for statistically optimal integration.

CTRNNs were evolved in 20 evolutionary runs with 1000 generations. 7 of the 20 networks evolved performed sufficiently well according to these criteria. The main exclusion criterion pointed towards a very successful but trivial local maximum for this task (up to $F \approx 0.6$): 7 networks were excluded because they considered only one stimulus and judged if it is 'big or not', which means that performance is good during testing for the standard o_s on one side, but at chance level or substandard for the other side.

Figure 2 depicts the psychometric curves for the different noise levels ν for all 7 successful networks together, as well as the JNDs and PSEs from individual fits. Increase in ν leads to a clear increase in JND (1 factor ANOVA: $F(4, 2) = 7.55, p < 0.001$), while PSEs are not influenced by noise ($F(4, 2) = 0.25, p = 0.91$). The successfully evolved networks show that, given the task and the fitness criterion, artificial systems can evolve to generate behaviour and simulated data that can be compared to human data and that can be analysed the same way.

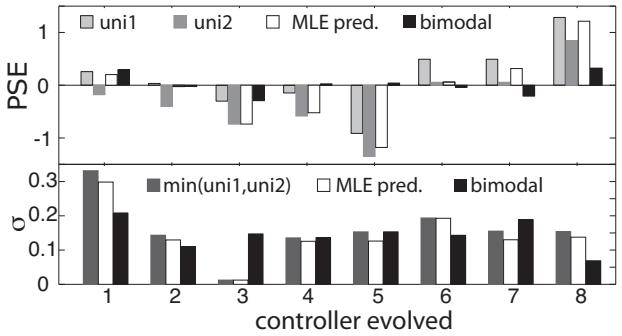

Figure 3: Unimodal, bimodal and predicted PSE (top) and σ (width of fitted Gaussians, bottom) for all networks evolved to perform (partial) bimodal discrimination.

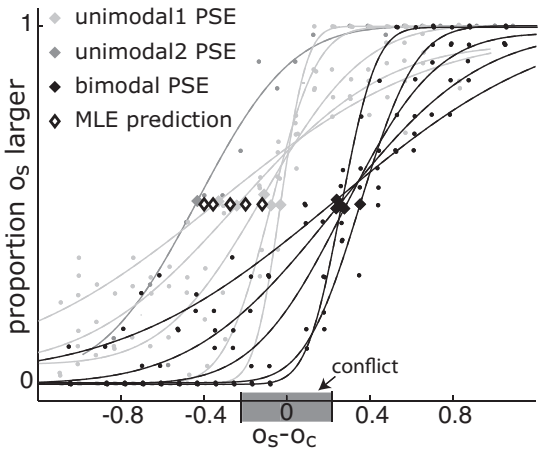

Figure 4: Example psychometric curves for the most successful network with $\nu = 0$ in the silent channel. $c = -0.25$, all noise levels ν. Data pooled for c_s left/right. Unimodal curves are shifted along the x-axis according to the conflict.

Bimodal Networks

In the bimodal condition, the emphasis is on the kind of integration behaviour that the networks exhibit and if it conforms to the predictions from MLE in Eq. (4).

Controllers for the bimodal condition were evolved in 20 evolutionary runs with 2000 generations. Only one network evolved to successfully discriminate between objects for all orders in both the unimodal and the bimodal conditions. The simulated data was fitted and analysed like in the previous simulation. When comparing the JND of the unimodal and the bimodal condition for the successfully evolved network, at first glance it appeared to exhibit the most important hallmark of MLE, i.e., that the probability distribution of bimodal estimates was more accurate than either of the unimodal estimates. However, testing the exact predictions from MLE (Eq. (4)) on this controller, the network proved to be *super-optimal*, i.e., the accuracy (in terms of σ of the fitted Gaussian) was dramatically better than expected from MLE (Fig. 3, bottom left).

7 of the other controllers evolved performed satisfactorily for both modalities if the standard o_s was presented to one side only. They were analysed and compared to the predictions of MLE as well. Even if lateral specialization is unsatisfactory concerning the main question, it involves some degree of integration. Figure 3 (bottom) depicts σ for the bimodal condition, averaged over noise levels ν, in comparison to the lower of the unimodal σ and the predicted σ using Eq. (4). All controllers were either grossly super-optimal or less accurate than the better of the uni-modal conditions, i.e., there was no evidence for optimal integration.

Why is it so easy to be 'better than optimal'? Is it because of the noise $\nu = 15$ of the inactive channel disturbs the network in the unimodal condition? Controllers were tested again with $\nu = 0$ in the unimodal condition to test this assumption. Contrary to the expectations, taking out the noise, in most cases (5 of the 8 networks), did not improve unimodal accuracy, but led to a complete break-down of unimodal discrimination. This indicates that the noise served a functional purpose in integration.

Defining the unimodal condition as noise with $\nu = 15$ and the absence of a signal had been an arbitrary design decision. However, as it is the case in biological evolution, the GA worked with what was there and thus incorporated this noise functionally into the solution, with surprising effects on perceptual accuracy across conditions. This result raises the question of what 'uni-modality' means in a multimodal system which will be picked up in the discussion. For those networks that also worked in the absence of noise, discrimination during unimodal trials became better than during bimodal case, eliminating the super-optimality. This result supports the hypothesis that noise in the silent channel is the reason for bimodal super-optimality.

Maybe more surprising still is the fact that the controllers did not evolve to integrate the two estimates. Introducing a cross-modal conflict, networks would be expected to generate PSEs in between the PSEs that the unimodal data predicts. Figure 3 (top) shows that, in the large majority of cases, the PSE of bimodal networks is far outside this range and, therefore, also far away from the PSE predicted from MLE. Figure 4 shows this behaviour for the most successful network (with $\nu = 0$ in the inactive channel): the discrimination is successful for all noise levels for both the unimodal and the bimodal stimuli. Accuracy for the bimodal trials is comparable to the unimodal trials. However, the PSE is far outside the range that would indicate integration. Rather than to integrate uni-modal estimates, the networks had evolved to perform a different and comparably viable way of discriminating size in the presence of redundant signals. The result indicates that multi-modal integration, as it is characteristic of humans, is not a process that simply

emerges as an epiphenomenon of the existence of redundant sensory channels but probably evolved due to more specific adaptive needs. The previously mentioned tendency of networks to evolve solutions with strong perceptual biases in this task is likely to also play a role in this result.

The solutions evolved do not make use of the dynamic complexity afforded by the recurrent network structure - they rely mainly on feed-forward principles. The passive open-loop nature of the task for disembodied recurrent networks does not encourage the use of dynamic complexity.

Perceptual Discrimination in Simple Agents

Living organisms are always in dynamic interaction with the environment. The surge of sensorimotor approaches in perception research (e.g., O'Regan and Noë, 2001) reflects an increasing awareness that such closed-loop dynamics afford alternative and clever ways of solving perceptual tasks. Existing models of optimal integration assume that integration, as well as estimation of channel certainty and weight adjustment are performed internally. The objective of evolving simple vision agents for this task was to explore if and how active perceptual strategies can play a role in multisensory integration and perceptual discrimination.

To bootstrap the evolution of active perceptual strategies, the performance criterion Eq. (2) was amended such that agents receive $P = 0.1$ if their visual system perceives both objects at least once, even if the wrong decision is made. If they do not move to see both objects, they receive $P = 0$, even if the right decision was made. In 20 evolutionary runs with 1000 generations, not one controller evolved that could reliably distinguish objects of different sizes for the whole problem space: local maxima, in most cases the mentioned solution to only pay attention to one of the stimuli, could not be overcome. Variations of the task were explored to mitigate this problem, including a punishment for lateral specialization and the administration of an extra position sensor, but performance never exceeded the stable local maximum, i.e., to focus just on one side. This suggests that a more radical change of fitness criterion/task may be necessary.

Controllers were also evolved for the bimodal condition in 16 runs for 2000 generations. The possibility exists that the presence of a direct sensory channel serves as a guidance for the evolution of active visual discrimination. Instead, the agents evolved rely heavily on their second (direct) input channel (see Fig. 1) and did not evolve to use their active sense according to demand. Where partially viable behaviour evolved, it replicates the general results from disembodied networks.

While these performance deficits mean that the predictions of the ideal observer model could not be tested, it is still interesting to test whether the partial solutions evolved exhibit sensorimotor strategies for sub-parts of the problem space. If agents evolve to base their decision on one input only, they could just evolve to move over to one side (passing the other side briefly to fulfill the revised performance criterion) and, otherwise, act as if they had a direct input channel. Instead, nearly all agents exploit their capacity to act in the closed-sensorimotor loop in order to make the 'big or not' strategy more effective. The remainder of this section presents examples of such active sub-strategies.

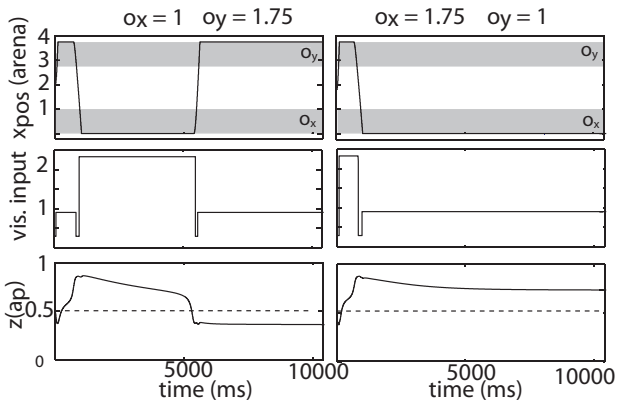

Figure 5: Selected variables across time from an agent presented with two pairs of objects with $o_x < o_y$ (left) and $o_x > o_y$ (right). Top: position. middle: sensory input from one input unit. bottom: decision output $z(n_p)$

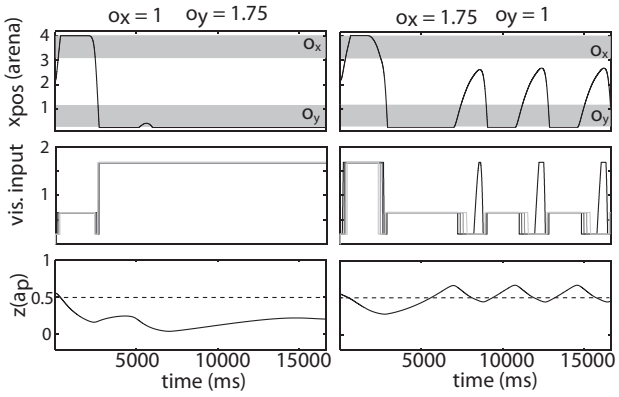

Figure 6: Selected variables across time from and agent presented with two pairs of objects with $o_x < o_y$ (left) and $o_x > o_y$ (right). Top: position middle: sensory inputs bottom: decision output $z(n_p)$

Active decision making. Figure 5 depicts the motion, inputs and decision output over time for an agent evolved. The agent evolved, under some circumstances, to steer towards the smaller of the two objects and to then make the decision contingent on the output velocity (using internal activation like an efference copy). This active decision making capacity is the most straight-forward one of the ones evolved and is an exception to the trend to pay attention to one input only.

*Active decision expression.*The agent depicted in Fig. 6

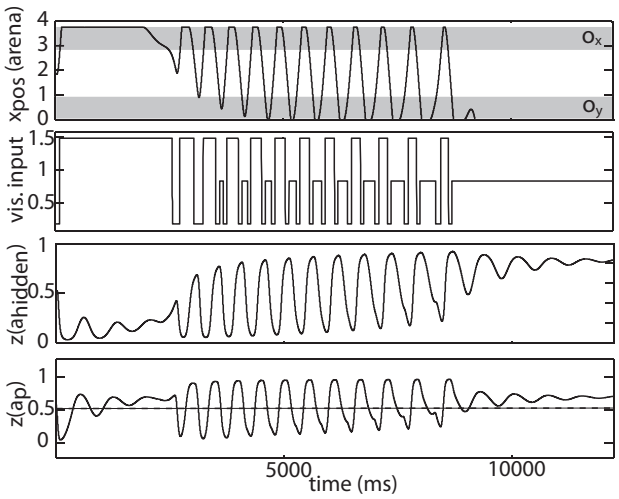

Figure 7: Selected variables across time from and agent presented with one pair of objects with $o_x > o_y$. Top: position. Second: sensory input from one unit. Third: $z(a)$ from a selected hidden unit. Bottom: decision output $z(n_p)$

evolved to only pay attention to the second input o_y. If the agent deems it large (Fig. 6 left), it comes to a halt and constantly outputs its decision ($z(a_p) < 0.5$). If, however, it deems the object small (Fig. 6 right), it initiates an oscillation towards and away from the object. Driven by this oscillation, the decision output starts oscillating around the decision boundary at $z(a_p) = 0.5$. This kind of behaviour evolved very frequently. It provides the agents with a way of expressing uncertainty: depending on when the trial ends, the same input would lead to different answers, and slight differences in object size may bias the proportion of such decisions by modulating the oscillations. Probably, such strategies evolved at least partially in response to the RB term in the fitness function Eq. (3) that punishes a strong response bias: if some of the decisions are random, it is unlikely that more than 75% of decisions would be of one kind.

Temporal decision making. Figure 7 depicts an agent's dynamics during the presentation of a single pair of objects. The agent's strategy makes active use of the time allocated for making a decision. One hidden unit (Fig. 7, third) controls the position of the agent: it decreases activity dramatically in the beginning (steering to the right) and then slowly increases. When it reaches a certain threshold, the agent starts moving to the left. Reaching the gap between the objects, the agent starts oscillating between the two objects, which is reflected in the activity of the hidden unit, too. The output unit always decides o_x is larger ($z(a_p) > 0.5$), unless the oscillations pull it below this threshold. Therefore, oscillation stands in correlation with the decision that o_x is smaller. The oscillation can only be stopped in time before the trial ends if the second object is small enough, otherwise it will go on indefinitely or at least till the end of the trial. In that sense, this controller can be seen as a variant of the o_y only strategy. The length of the oscillatory phase is, however, not just contingent on o_y. The size of o_x appears to take influence on the time of onset of the oscillations as well as its offset in ways that are not obvious.

These are just three examples of the ways in which agents used their motion capacities in their size discrimination activity, not all of which are easy to understand. In depth analysis of only partially functional agents is an endeavour of limited value. The fact that an abundance of active strategies evolved, however, is a result worth mentioning. In systems that discriminate stimuli exploiting the agent-environment interaction dynamics, processes of multisensory integration would rely on these closed-loop dynamics. How (optimal) integration could work in the absence of explicit representation of perceptual estimates remains an intriguing open question.

Discussion

Using ER for this kind of multisensory perceptual discrimination task is a novel approach and as such the research presented has mainly exploratory character. Both technical and conceptual difficulties were encountered. Most dramatically, minimal agents could not be evolved to perform perceptual discrimination and the predictions from MLE could not be tested for the second part of the project. ER simulation modelling serves as a tool for thinking, and as such, the simulation results here presented have pointed out a number of issues that are worth reporting.

Unimodality in a Bimodal System

Possibly the most important insight gained from the simulation models is that existing models of optimal integration have a gap to fill: as humans, it is obvious for us what a unimodal and what a bimodal stimulus is. It is, however, not clear how the MLE circuits proposed (e.g. Knill and Pouget, 2004; Ernst and Banks, 2002; Alais and Burr, 2004) or a localized brain area would be able to recognise the absence of a signal in one channel and what possible noise entering through that channel can do to the decision making process. MLE assumes independent channels and independent processes of unimodal estimation and multisensory integration (cf. Method section). How the same process of generating perceptual judgments in human observers can be indicative of either of the stages is not made clear in existing models. In the model presented, the administration of random noise in the silent channel led to the evolution of apparent 'super-optimality' in bimodal trials: not because networks accurately estimate the levels of noise present, but just because additional noise sources were absent during bimodal trials. The fact that performance breaks down in most controllers when the noise is removed shows that the definition

of what 'uni-modal' means in a system is not an arbitrary one. Existing models of optimal integration would benefit from making explicit the behaviour of the inactive channel during unimodal trials and incorporating mechanisms into their models that distinguish between multimodal and bimodal trials. Testing for their existence can then confirm that the reported increase in accuracy in bimodal trials is not due to the influence of the silent channel during 'unimodal' trials.

Perception vs. Perceptual Judgments

Unlike humans, the evolved systems were surprisingly incapable to integrate their senses in a coherent way. This problem may well be due to the fact that the controllers were evolved for a laboratory task. 2AFC perceptual discrimination tasks, like the size discrimination task used here, make it possible to measure perceptual accuracy, as well as perceptual bias. The fitness criterion Eq. (3) emphasises this accuracy component. Therefore, the systems evolved tend to favour being accurate over the absence of perceptual biases (as evident from the large and variable PSEs in Fig. 3) and are rewarded for this tendency. Humans, on the other hand, develop their perceptual skills not for this kind of psychophysics task, but in real-world situations, where perception has behavioural relevance. In many real-world contexts, strong or variable perceptual biases would be extremely disadvantageous. In future research, therefore, systems will not be evolved for 2AFC tasks exclusively, but for perceptual capacities more generally (e.g., the approach taken here can be combined with a magnitude estimation task or with a sensorimotor control task that involves perceptual decision making).

Ideal Observing vs. Active Sensing

Ideal Observer Models of perceptual integration strongly draw on the assumptions of the dominant representationalist paradigm in cognitive science: MLE is a dedicated process that combines unimodal estimates and noise estimates. Even though behavioural approaches (e.g. Ernst and Banks, 2002; Alais and Burr, 2004) are *prima facie* agnostic about the underlying mechanisms, it is easy to jump to conclusions and assume that internal dedicated neural process perform MLE, represent the noise, represent the unimodal estimates, etc. (e.g. Knill and Pouget, 2004). Evolving embodied agents to integrate their senses optimally (on a behavioural level) can potentially challenge such underlying assumptions (on the level of the underlying mechanism). The active vision agents presented here did not arrive at a level of behaviour that would allow drawing strong conclusions about multisensory integration. However, even superficial analysis of their behaviour revealed an abundance of active sensing in the accomplishment of aspects of perceptual discrimination, including but not limited to active decision making and the expression of uncertainty through motion patterns.Thinking of the human hand and the human eye as agents, it is not unlikely that active sensing principles are exploited in a task like visuo-haptic size estimation. It is by no means clear that the introduction of noise or the variation of physical parameters, like in psychophysics, would have the same impact on such embodied processes as they have on decoupled systems that are passively cruncing representations. Even though limited in their own significance, the present results provide a good incentive to proceed with a revised version of the research on perceptual discrimination in simulated agents.

Noise and Uncertainty

The question of noise estimation, independent noise sources and reduction of uncertainty is one of the cornerstones of optimal multisensory integration research. Given that no system evolved to confirm the predictions from MLE, this question could not be direclty addressed. The first simulation confirmed that the introduction of different levels of Gaussian noise led to the expected deterioration of perceptual accuracy (cf. Fig. 2). It is arguable if adding Gaussian noise at any time step to a signal that is then fed into a rate code neural network is the most suitable approach for the evolution of systems whose behaviour is contingent on levels of noise. As a lot of the noise is filtered directly by the neurons, that have a minimal time constant of $\tau = 30ms$, such systems may have a hard time to develop sensitivity to levels of noise. In future models, noise may instead be added to a physical stimulus, which, at least in theory, would allow agents to use active strategies not just to perform perceptual discrimination, but also to perform noise estimation. Generally, it was a long shot to expect that optimal integration would evolve in evolved systems by merely adding the requirement to be accurate in perceptual discrimination. Even if the outlined technical and conceptual problems can be solved in future research, it may be necessary as a next step to explicitly require agents to integrate optimally in order to tackle this question.

Conclusion

The ambitious goal to evolve optimal multisensory integration in networks and agents has not been met in the current research. However, the difficulties encountered were informative about hidden prior assumptions on several levels: about ideal observer models (what is 'unimodal' in a bimodal system? Can noise in the silent channel explain an increase in bimodal perceptual accuracy?), about using a psychophysics task for evolution (does success in a 2AFC task equal perceptual capacity?) and about the role of action in perceptual discrimination (if active sensing is beneficial for perceptual discrimination, how does it figure in multisensory integration?). Rather than answering one question, the study generated more digestible sub-questions, which is characteristic of generative ER models. The outlined avenues for future research will be pursued to further elucidate

the relevant question of (optimal) multisensory integration from an embodied and Artificial Life point of view.

Acknowledgements

This work was supported by the HFSP Research Grant (2006) on Mechanisms of associative learning in human perception.

References

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14:257–262.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11:209–243.

Di Paolo, E., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26:278–294.

Ernst, M. O. (2005). A Bayesian view on multimodal cue integration. In Knoblich, G., Grosjean, M., Thornton, I., and Shiffrar, M., editors, *Human body perception from the inside out*, pages 105–131. Oxford University Press, New York.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433.

Floreano, D., Kato, T., Marocco, D., and Sauser, E. (2004). Coevolution of active vision and feature selection. *Biological Cybernetics*, 90:218–228.

Freeman, W. J. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological Cybernetics*, 56.

Harvey, I., Di Paolo, E., Wood, R., Quinn, M., and Tuci, E. A. (2005). Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, 11(1-2):79–98.

Hill, J. (2005). psignifit toolbox for Matlab 5 and up. http://www.bootstrap-software.org/psignifit/ retrieved 05.02.2010. Version 2.5.6 for Mac OSX.

Izquierdo-Torres, E. and Di Paolo, E. (2005). Is an embodied system ever purely reactive? In *Proceedings of the 6th European Conference of Artificial Life ECAL 2005*, pages 252–261.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712 – 719.

O'Regan, K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24:939–1011.