Control of Gaze in the Context of Behavior

Mary Hayhoe
Center for Perceptual Systems
University of Texas at Austin

To understand human gaze, experimenter typically manipulates gaze via instructions, or manipulates stimuli to attract gaze.

In natural behavior, selection and timing under subject's control.

Thus we must look at natural behavior in order to understand the factors controlling gaze.

Such understanding is necessary in order to make inferences from gaze behavior.

Thus applications of pervasive eye tracking will depend very heavily on driving forces behind gaze control.

FACTORS THAT CONTROL GAZE.

TASK

REWARDS

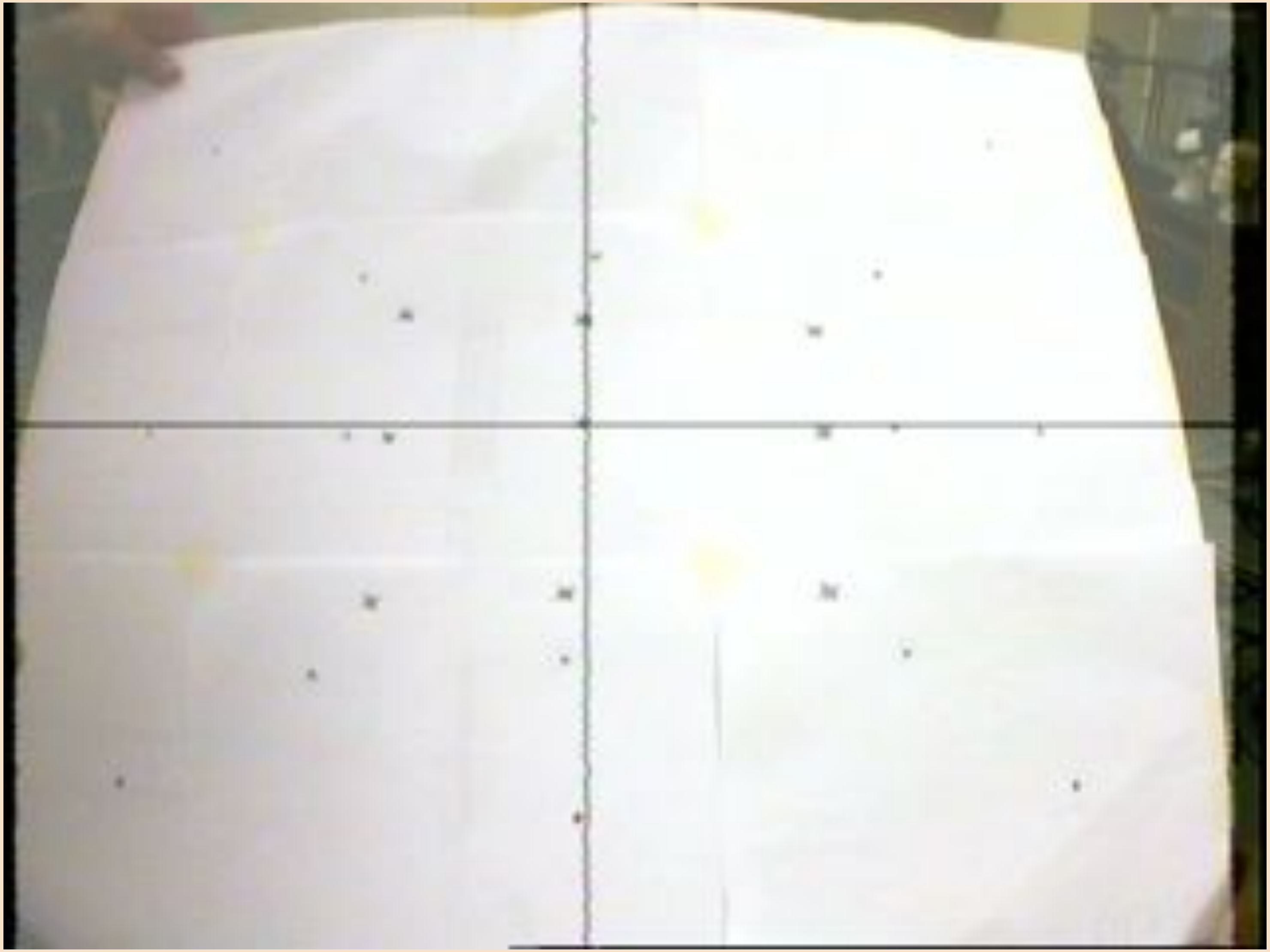
UNCERTAINTY

PRIOR BELIEFS/ Memory

IMAGE

Acquisition of visual information is goal driven

Fixations tightly linked to actions: Land (2004); Hayhoe & Ballard (2005) etc



In natural behavior, both “when” and “where” matter.

What underlies the momentary decisions of where/when to look?

Immediate behavioral goals govern gaze target selection.

- where to attend, when to attend, and what information to get. (Plate for knife placement, edge of slice for grasp etc.)

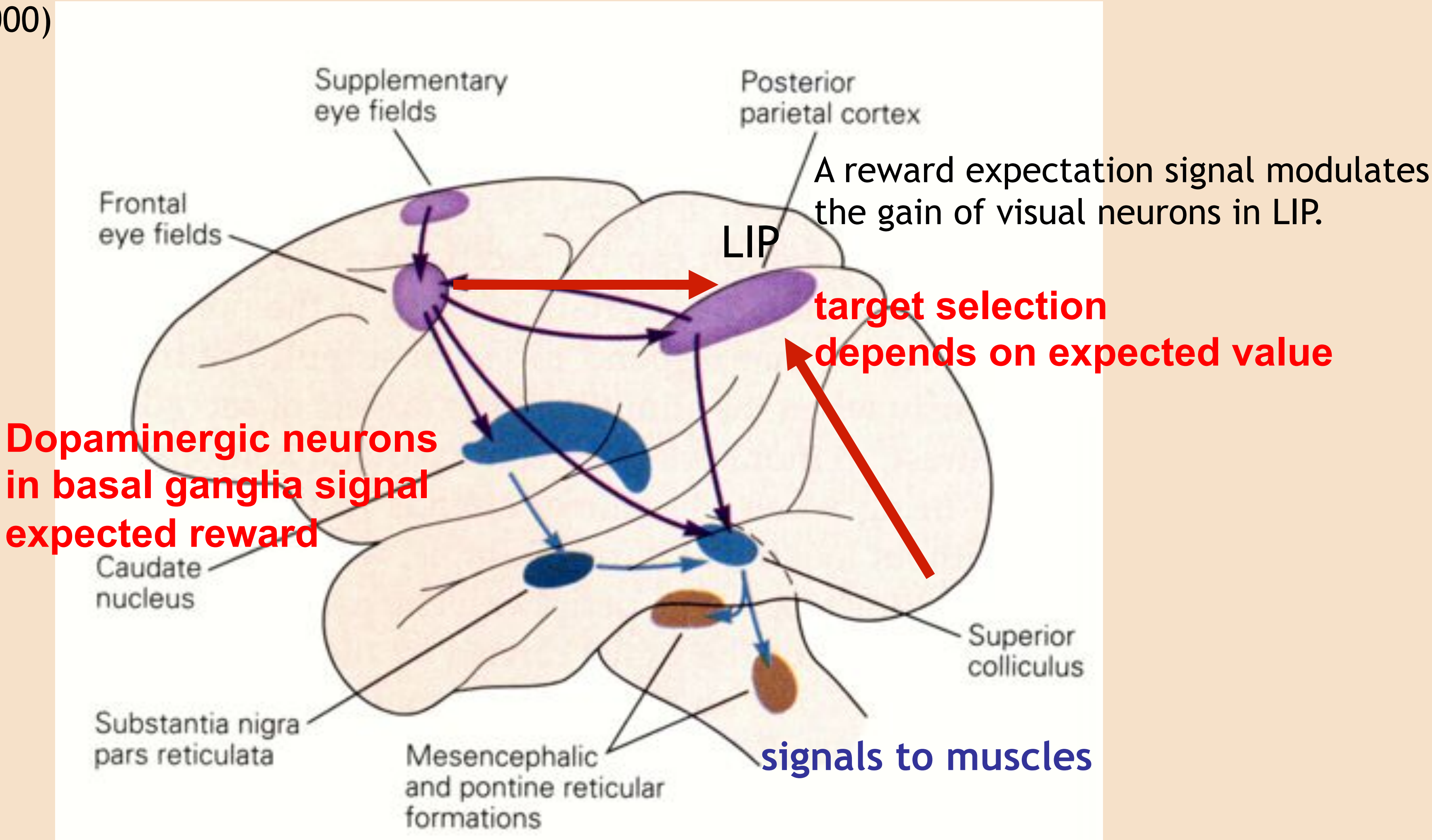
How does this come about?

Hypothesis: task control results from reward-based learning.

Reward sensitivity of Saccadic Circuitry

Neurons at all levels of saccadic eye movement circuitry are sensitive to reward (eg juice).

Neural basis for reinforcement learning models of gaze behavior. (Schultz, 2000)



What about human behavior? Any evidence for role of reward?

Reward effects in neurons have been observed with very simple choice response paradigms eg “look to left target for a drop of juice”.

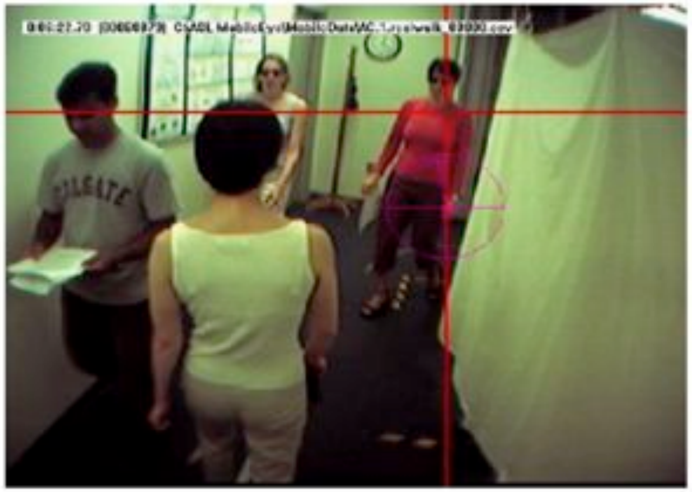
But eye movements are for getting information and are not directly rewarded in natural vision.

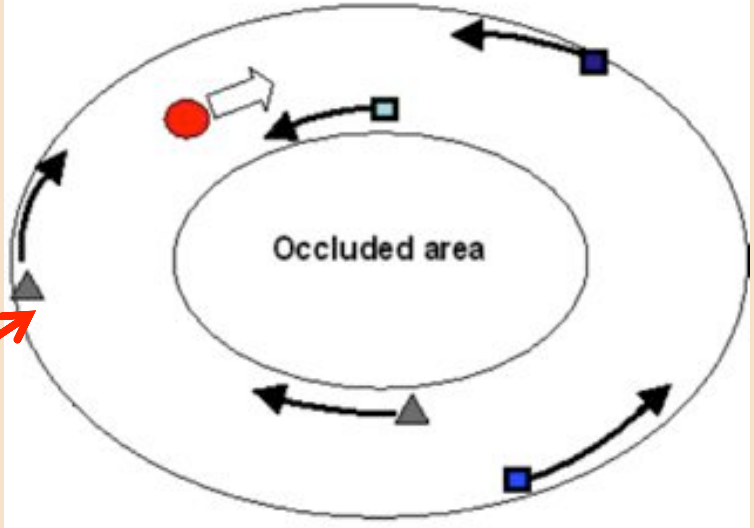
Need evidence for task (reward-based) control of gaze/attention in natural behavior.

Gaze allocation when walking in a real environment

Things to do: control direction, avoid obstacles, foot placement, characterize surroundings etc   Normal vision involves sets of behavioral goals, or sub-tasks – need to allocate attention effectively between sub-tasks.

Portable ASL eyetracker
Oval path around large room





pedestrians

(Jovancevic & Hayhoe, 2009 J Neurosci)

How are gaze targets chosen?

Dynamic environments are tricky – timing of fixations more critical than in static scenes.

Manipulation of behavioral relevance/reward weight

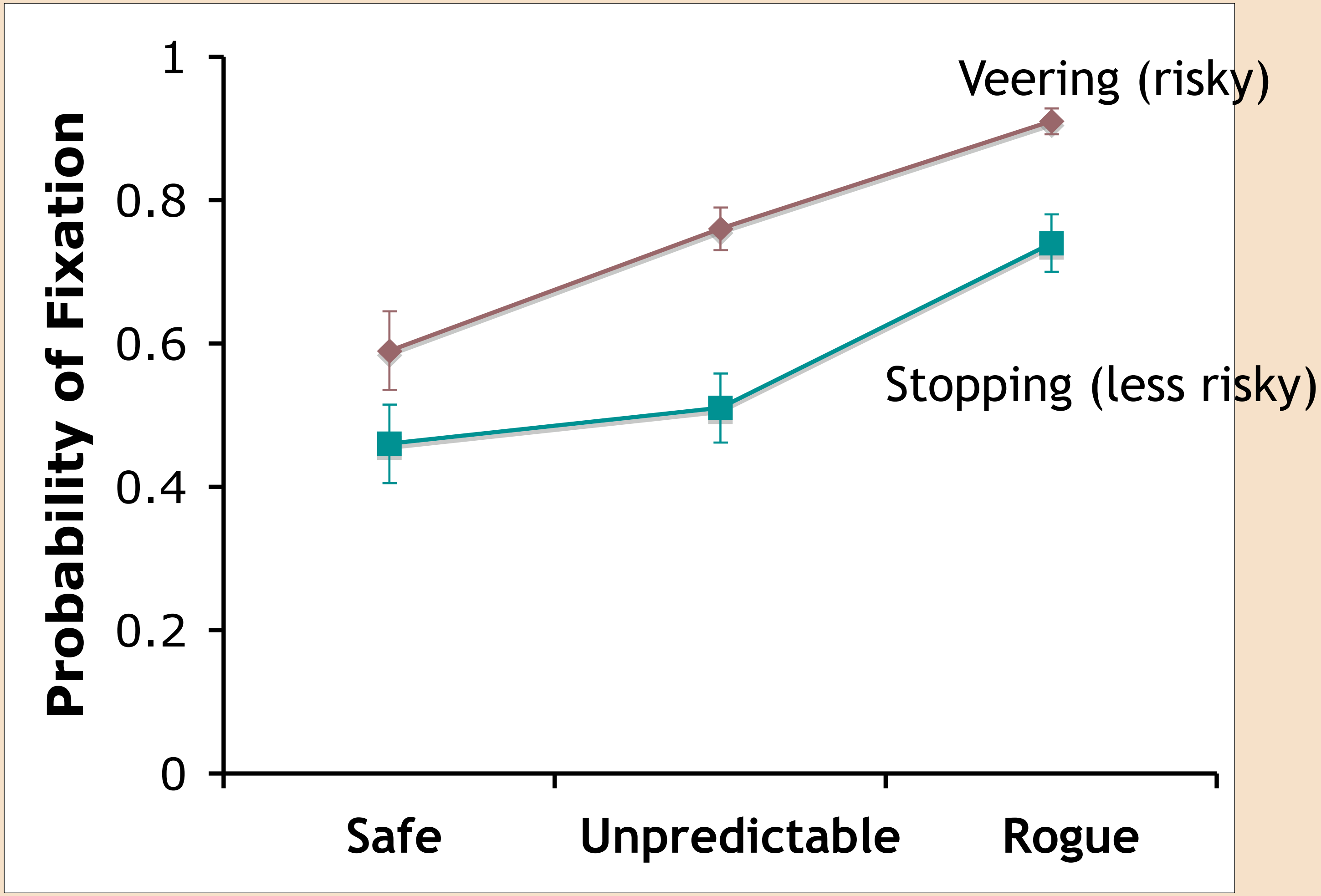
Occasionally some pedestrians either **stopped** for 1 sec or **veered** on a collision course with the subject

3 pedestrians behaved in characteristic ways:

Rogue pedestrian – always stops/veers
Safe pedestrian – never stops/ veers
Unpredictable pedestrian – stops/veers 50% of time

Reward (negative) = potential collision

Fixation probability depends on behavioral relevance (subjective value) and probability of veering/stopping



(Probability is computed during period in the field of view, not just collision interval.)

Almost all of the fixations on the Rogue were made **before** the veering onset (92%).

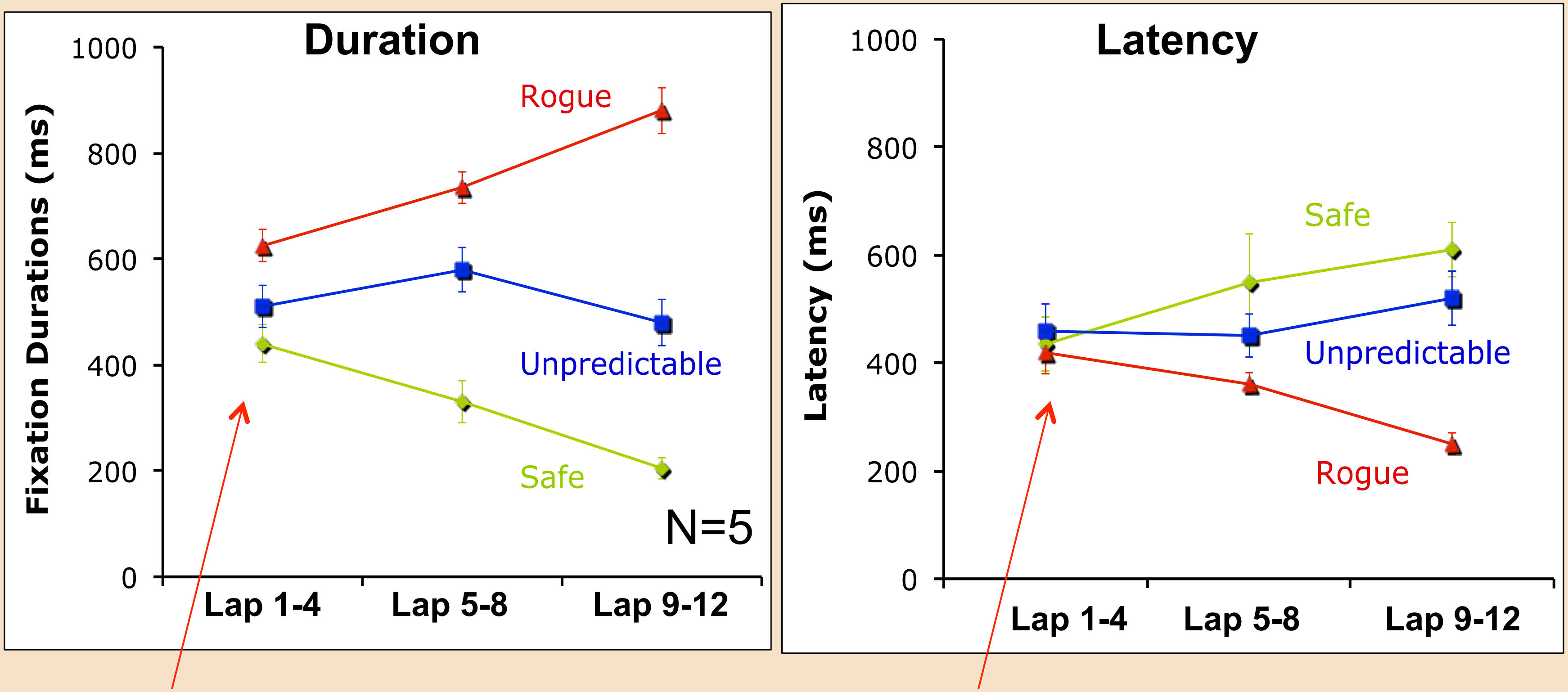
Thus gaze, and attention are anticipatory, based on history of events, not a result of what the pedestrian is actually doing.

Gaze behavior based on expectation, not on veering event.

Probability of fixating unpredictable pedestrian similar, whether or not pedestrian actually veers on that trial.

Attention depends on reward probability (expected value)

Gaze behavior changes rapidly with experience (4-5 encounters):
priorities re-allocated depending on behavioral relevance



prior **prior**

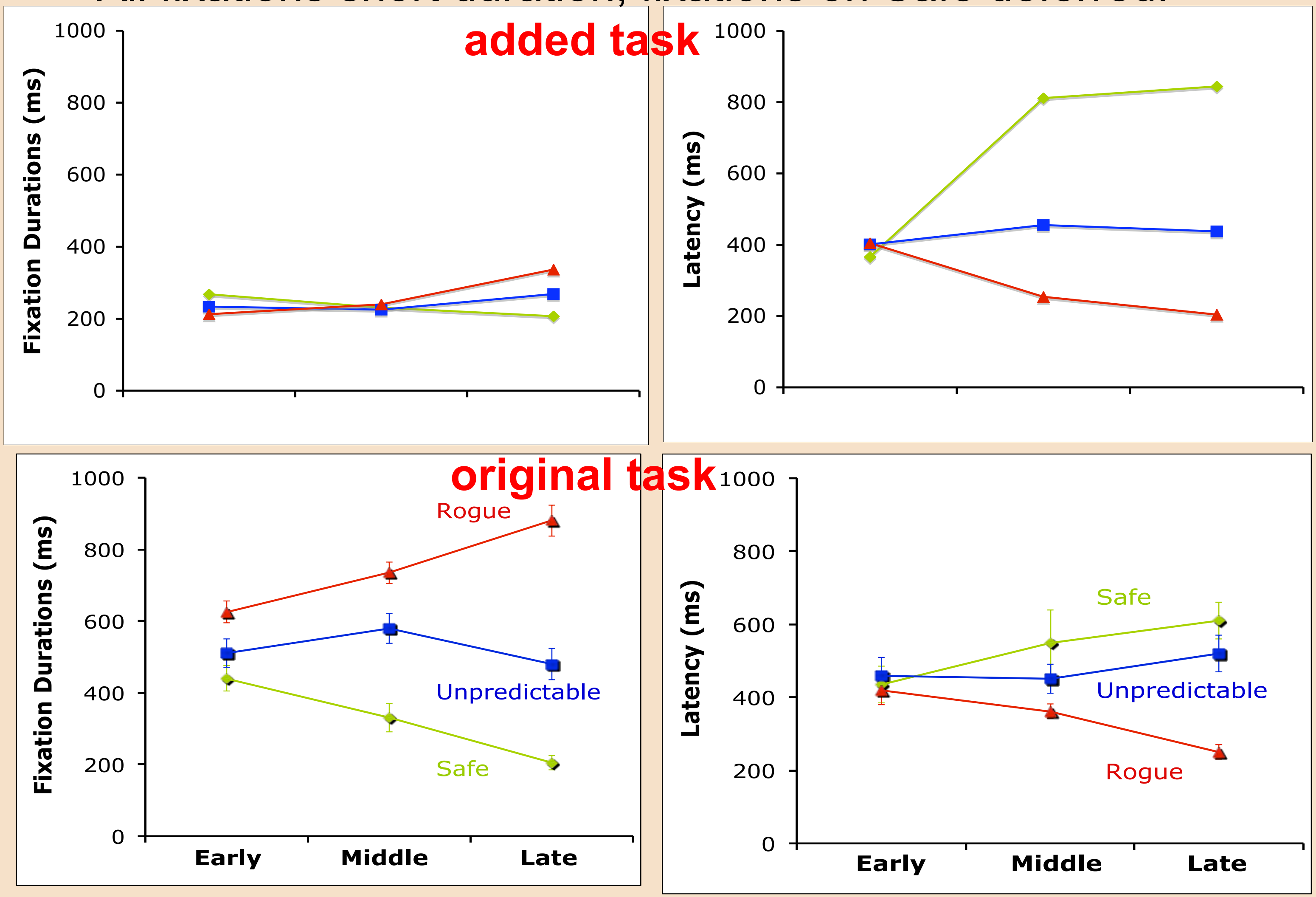
Fixations on Rogue get longer/earlier, on Safe shorter/later

Note – 5 subjects – similar behavior

SHARING ATTENTION BETWEEN TASKS

Gaze priorities change when another task is added.

All fixations short duration, fixations on Safe deferred.



Thus there must be some mechanism for determining how gaze is shared

Fixations modulated by task importance/value (reward, and probability of reward).

Subjects learn the statistical structure of the world and allocate attention and gaze accordingly.

Control of gaze, is proactive, not reactive ie based on estimated state, not on the current image.

Subjects behave very similarly despite unconstrained environment and absence of instructions.

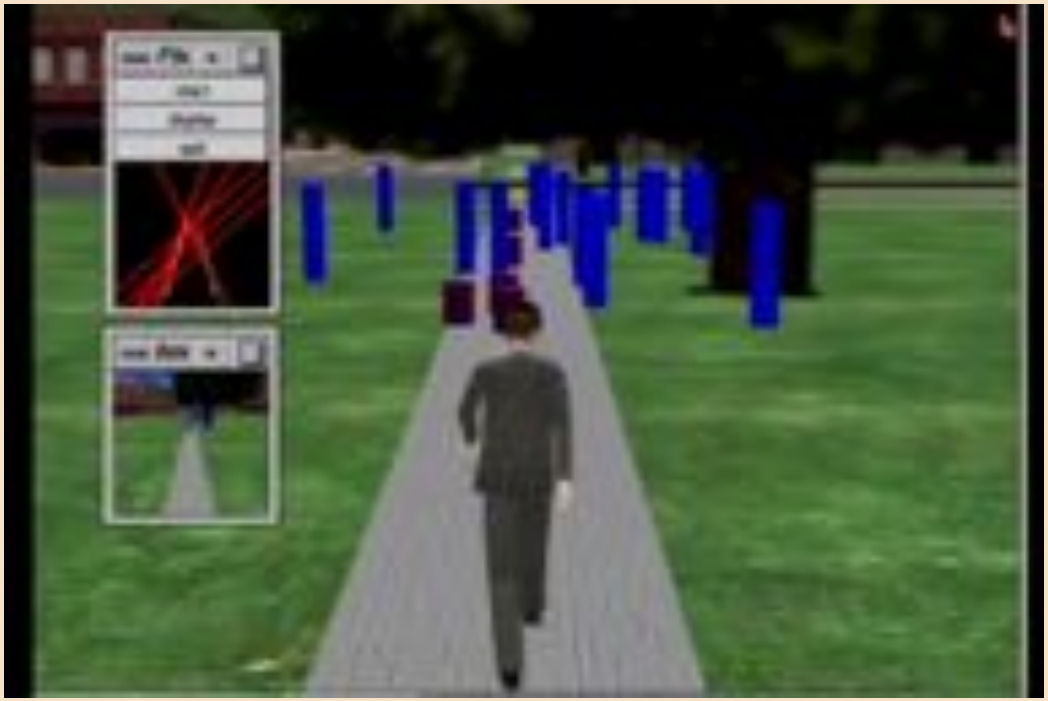
Reinforcement Learning

Neural reward machinery provides a basis for RL models.

RL models might explain how tasks influence human gaze.

R L Modeling of Gaze Control

Walter the Virtual Humanoid



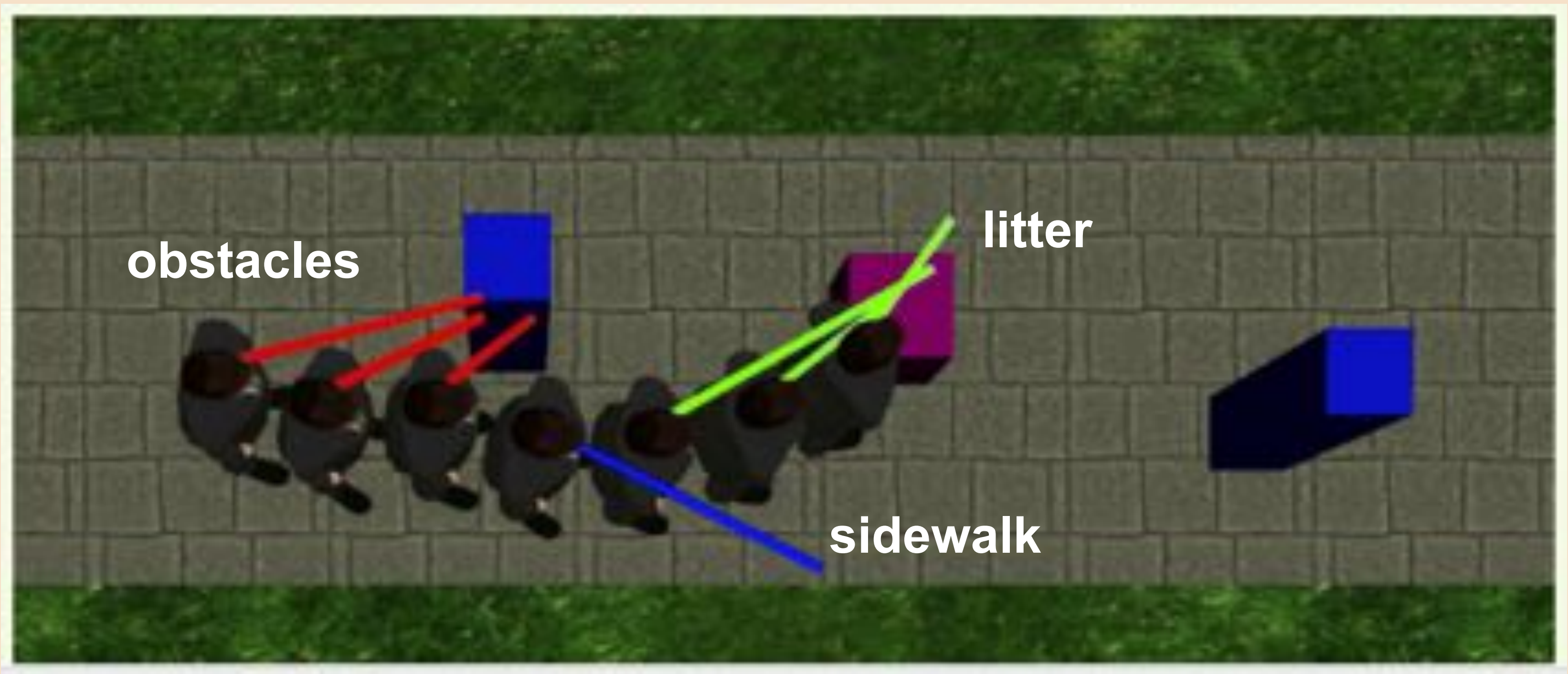
Sprague, Ballard, & Robinson TAP (2007)

Virtual Humanoid has a small library of simple visual behaviors (modules):

- Sidewalk Following
- Picking Up Blocks
- Avoiding Obstacles

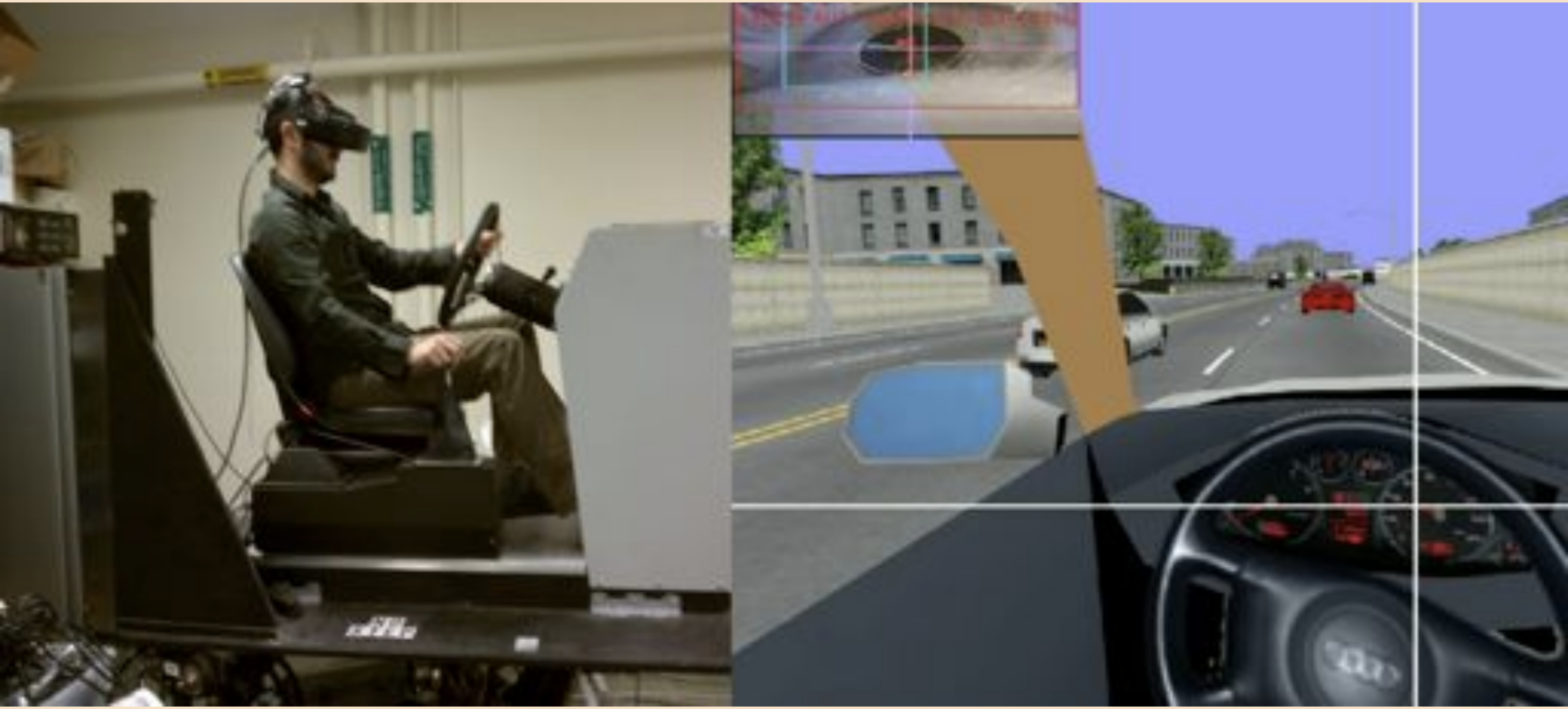
Each behavior uses a ***limited, task-relevant*** selection of visual information from scene.
Behaviors have different priority/ reward value.

Controlling the Sequence of fixations



Gaze target is chosen based on both reward and uncertainty.

Evidence that gaze scheduling depends on reward-weighted uncertainty.



Work by Brian Sullivan.

Instructions:

- Keep a constant speed

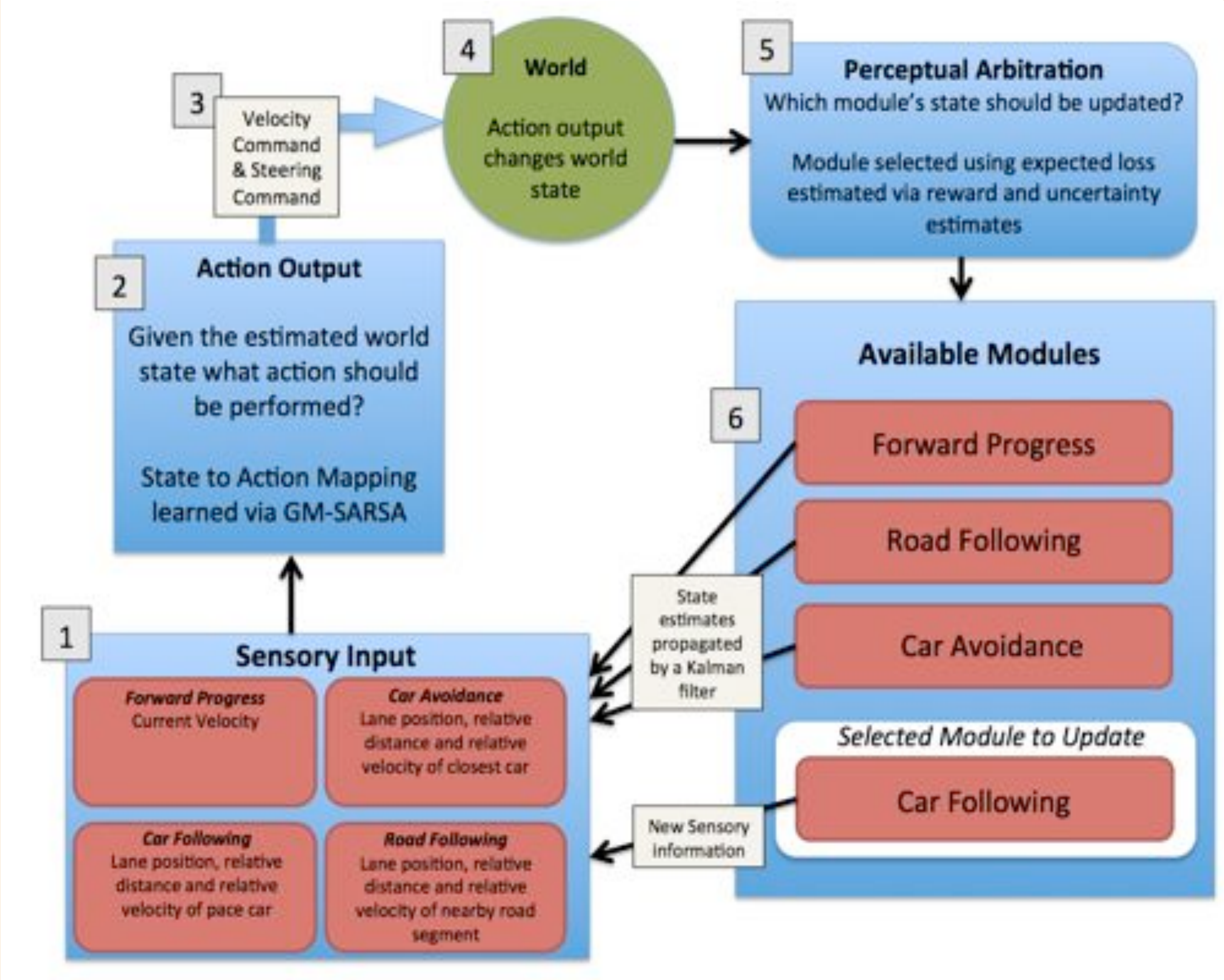
or

- Keep a constant distance behind a leader car

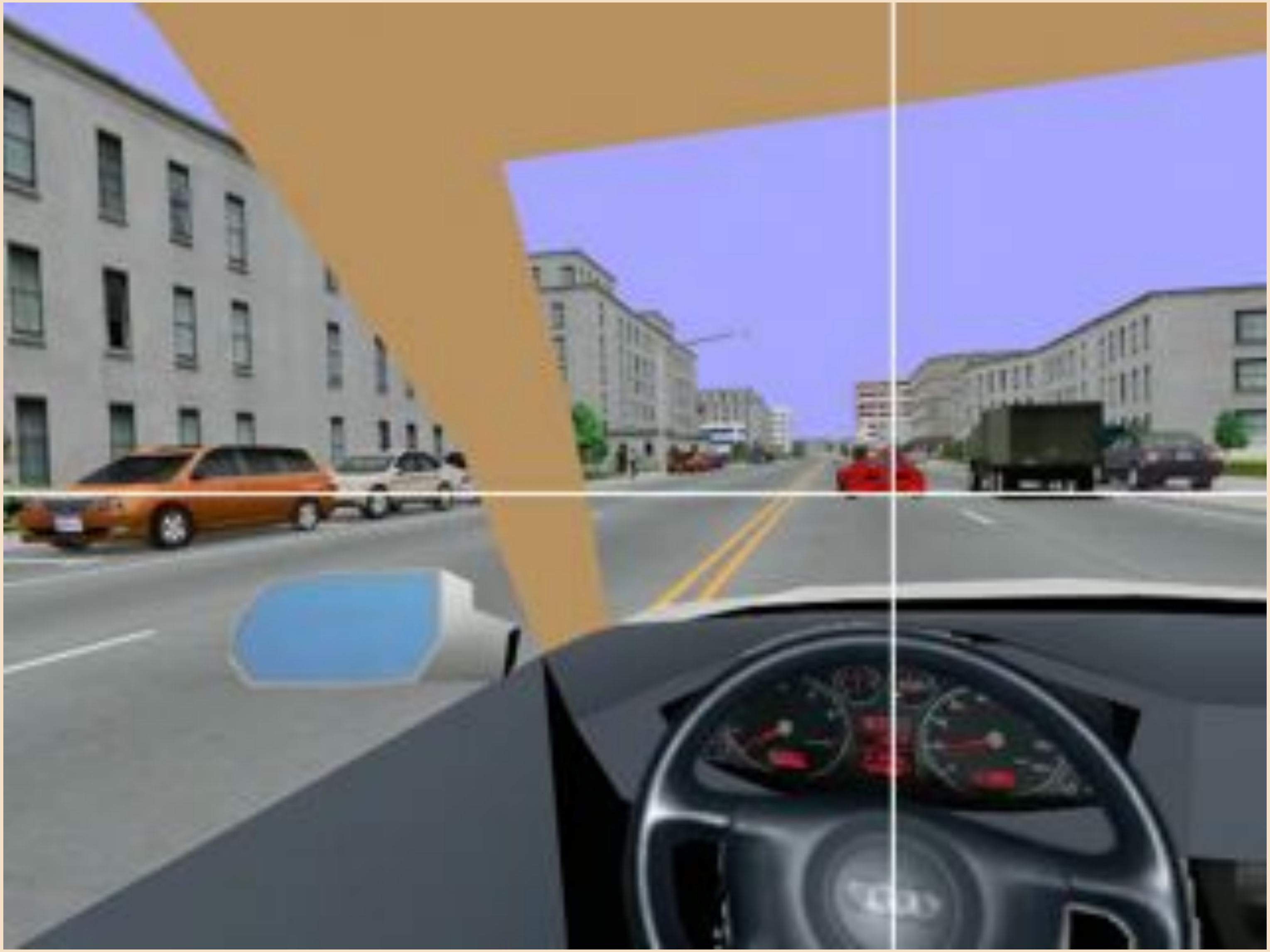
Noise was added to the gas pedal so that speed fluctuated. This should lead to greater uncertainty about speed, and lead to more frequent updates using gaze.

The **Task** manipulation varies implicit reward.

Schematic of the Sprague/Ballard model

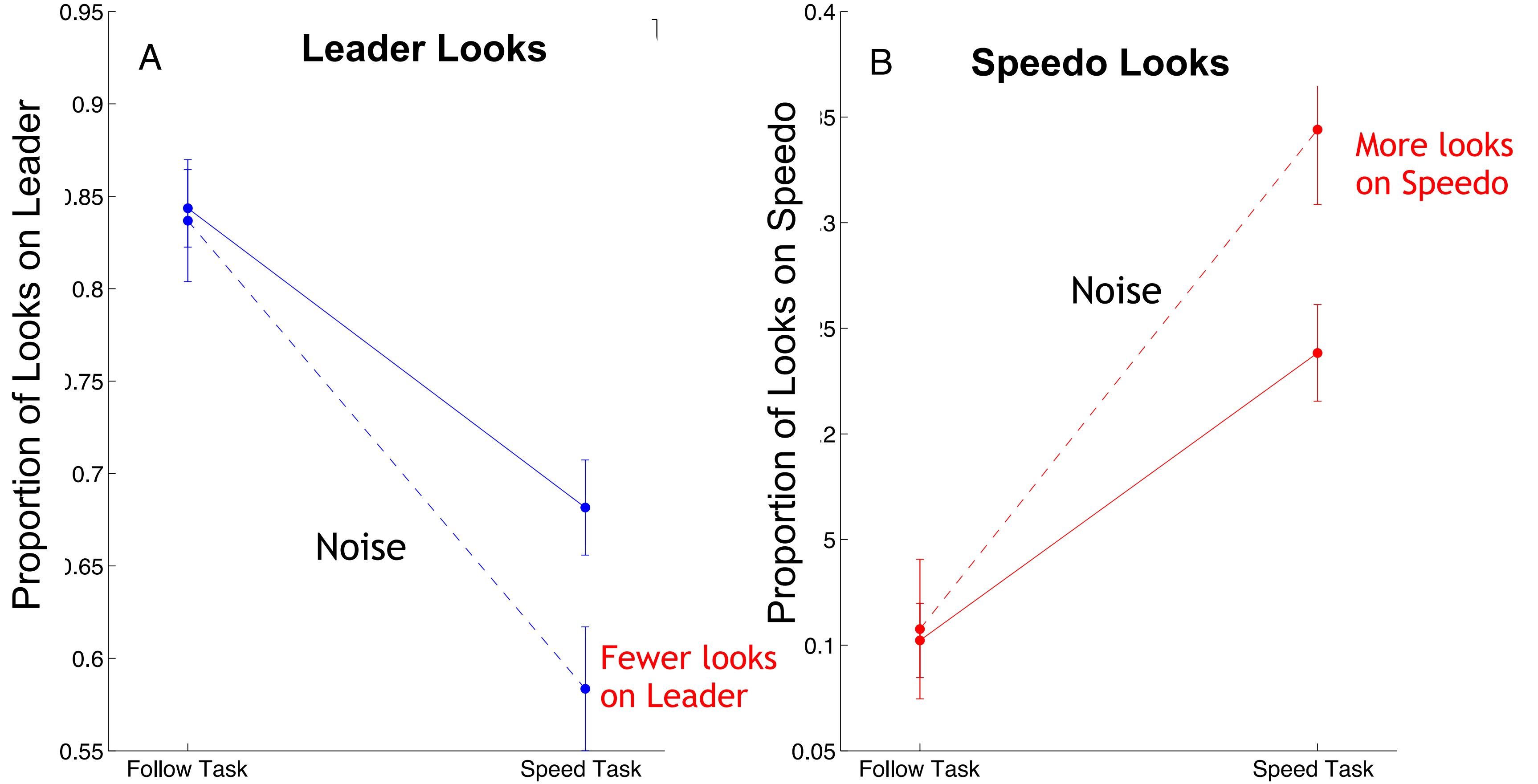


Sullivan, 2012



Subjects must schedule looks to leader and speedometer in order to maintain correct speed or distance.

How does task and noise affect gaze allocation?



Subjects look more at the Speedo in the Constant Speed task and more at the Lead Car in the Follow task.

Noise increases Speedo looks and reduces Leader looks in Constant Speed condition.

Tradeoff between task priority and uncertainty.

Qualitatively consistent with model that selects a gaze target on basis of reward-weighted uncertainty.

Reduction of visual uncertainty is gated by behavioral relevance .

FACTORS THAT CONTROL GAZE.

TASK

REWARDS

UNCERTAINTY

PRIOR BELIEFS/ Memory

IMAGE

In walking paradigm, gaze behavior was anticipatory. Fixations on Rogue almost always occurred before the relevant stimulus event.

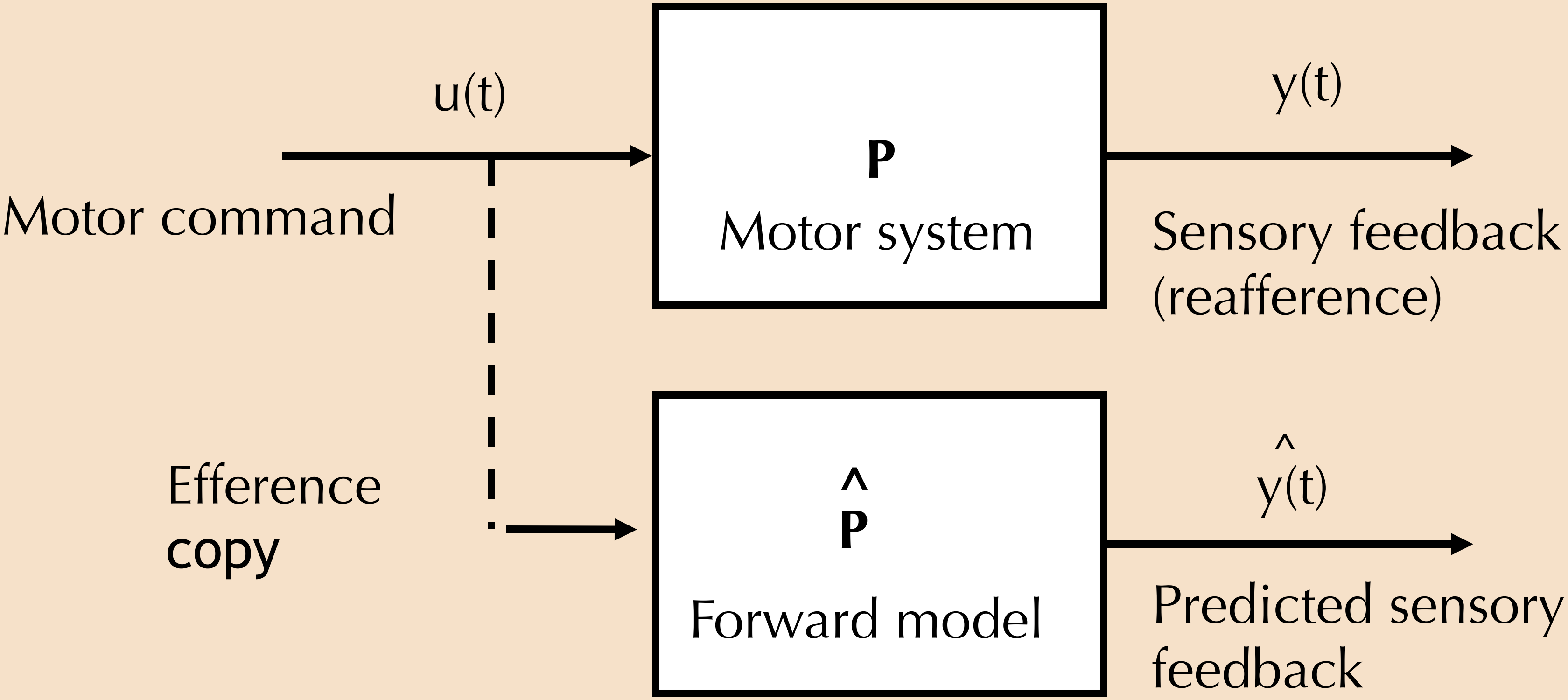
Fixations were based on predicted behavior.

Why?

Sensory delays make early planning of eye movements important/necessary.

What is the basis of prediction? Idea of Internal Models

In the case of body movements, forward models of body's dynamics predict somatosensory consequences of movements (Wolpert et al, 1998).



Rapid comparison of actual with expected feedback circumvents delays

Internal models of visual world?

Evidence to the contrary..

Many actions can be controlled by the momentary visual signals in the image. (Warren, 2006)

For example, use looming information to compute "time-to-contact" to control interception/braking; "focus of expansion" to control heading.

That is, extract a "control variable"

Advantage: computational efficiency.

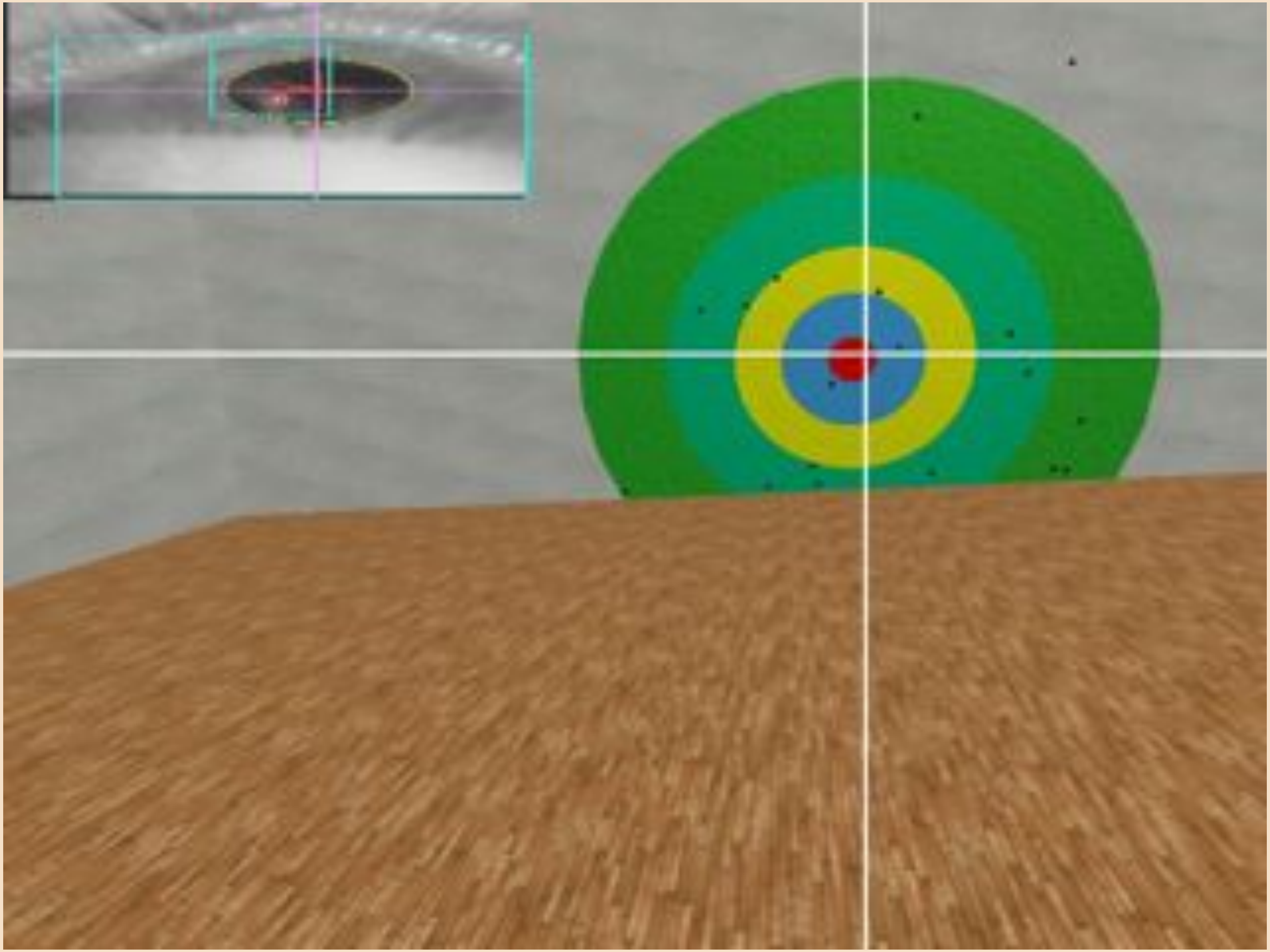
In natural movements, do we need internal models of visual state to generate eye movements in advance of events in the image?

What is the nature of these internal models?

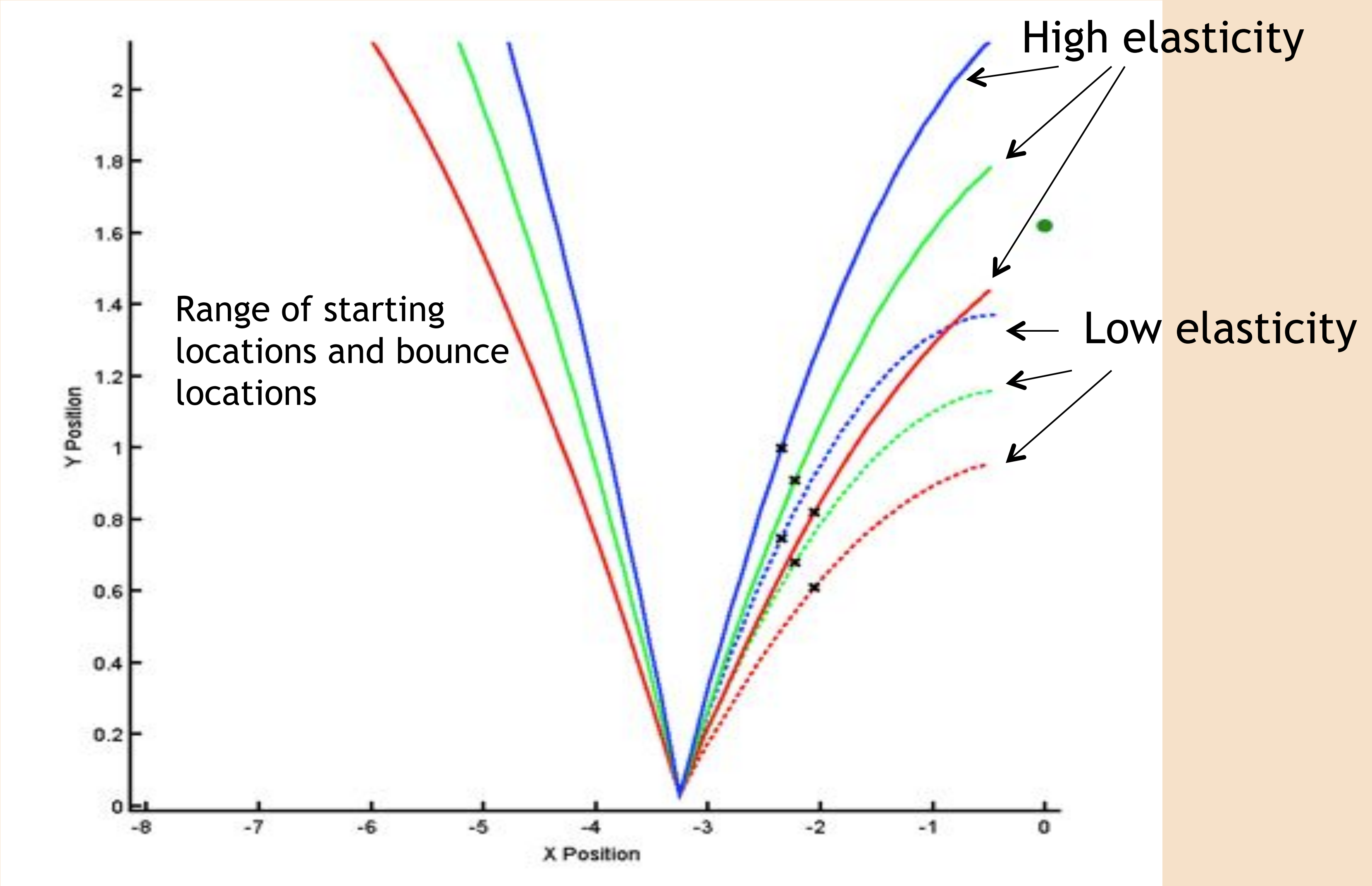
Virtual racquetball:

Work by Gabe Diaz

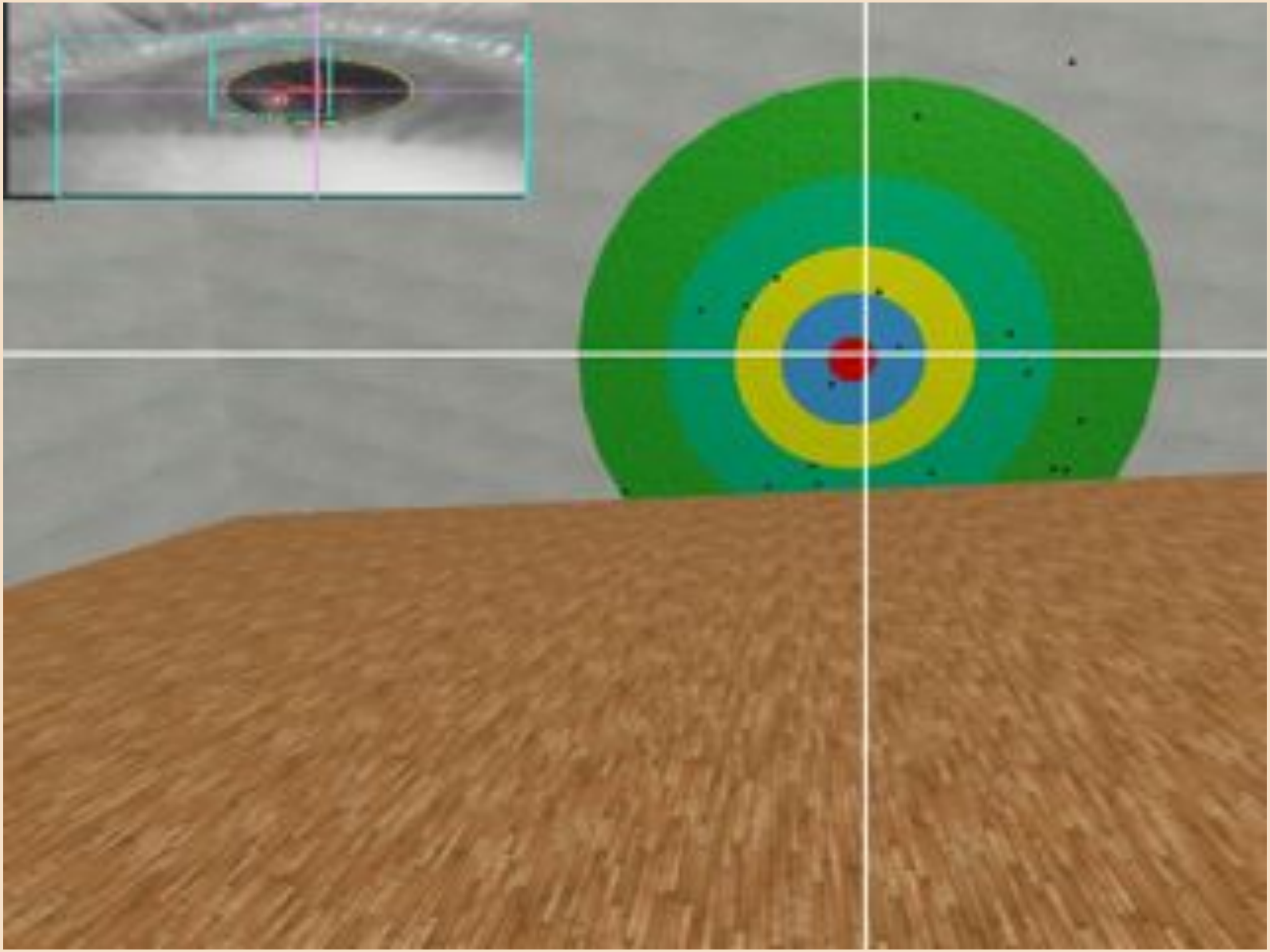
Nvis helmet, Arrington eye-tracker, PhaseSpace head/hand/racquet tracking, ODE to control ball and racquet interactions



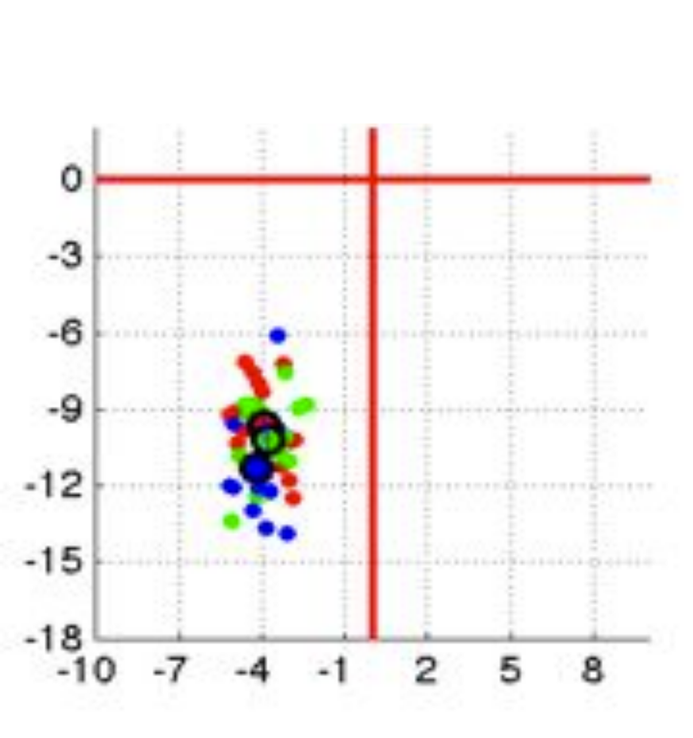
Balls varied in vertical **velocity** and **elasticity**.
Velocity varied from trial to trial, elasticity was constant within a block.
Height after bounce predictable from current trial and previous trials within a block.



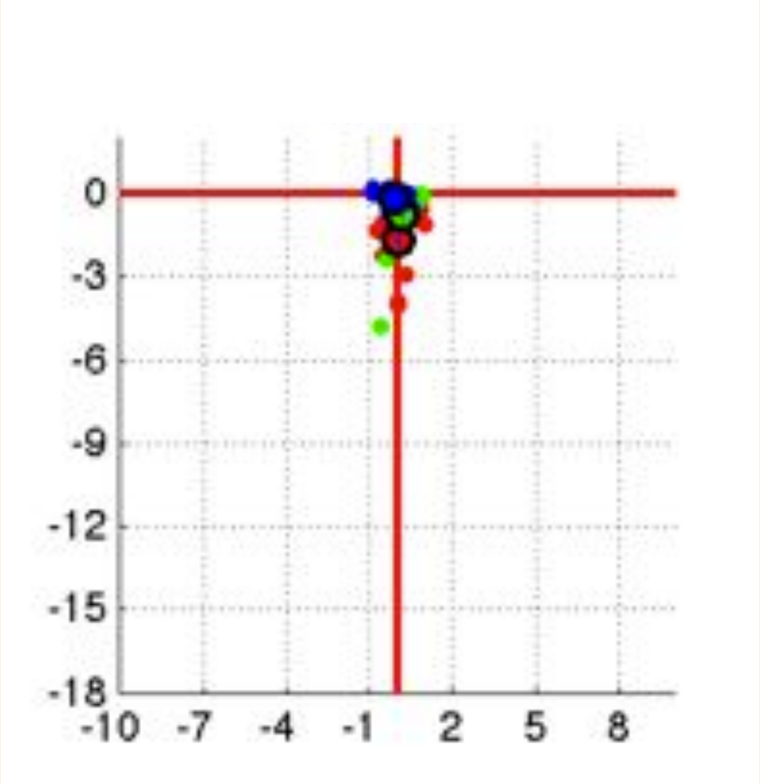
Saccade to a location ahead of the bounce



Subjects' gaze predicts location of ball after it bounces. Prediction is based on knowledge of elasticity, (based on history) plus velocity.



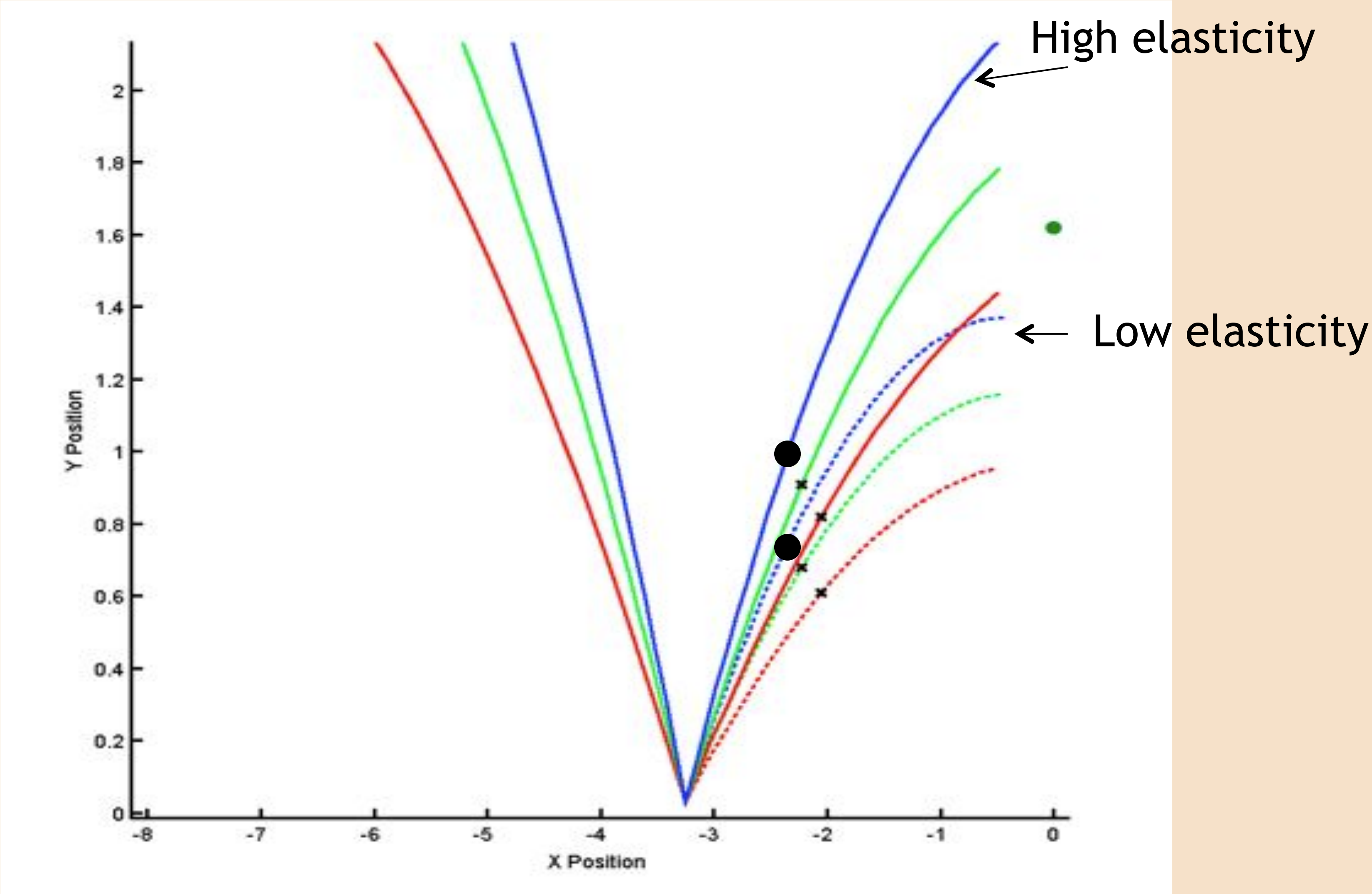
Ball location relative to gaze at time of bounce



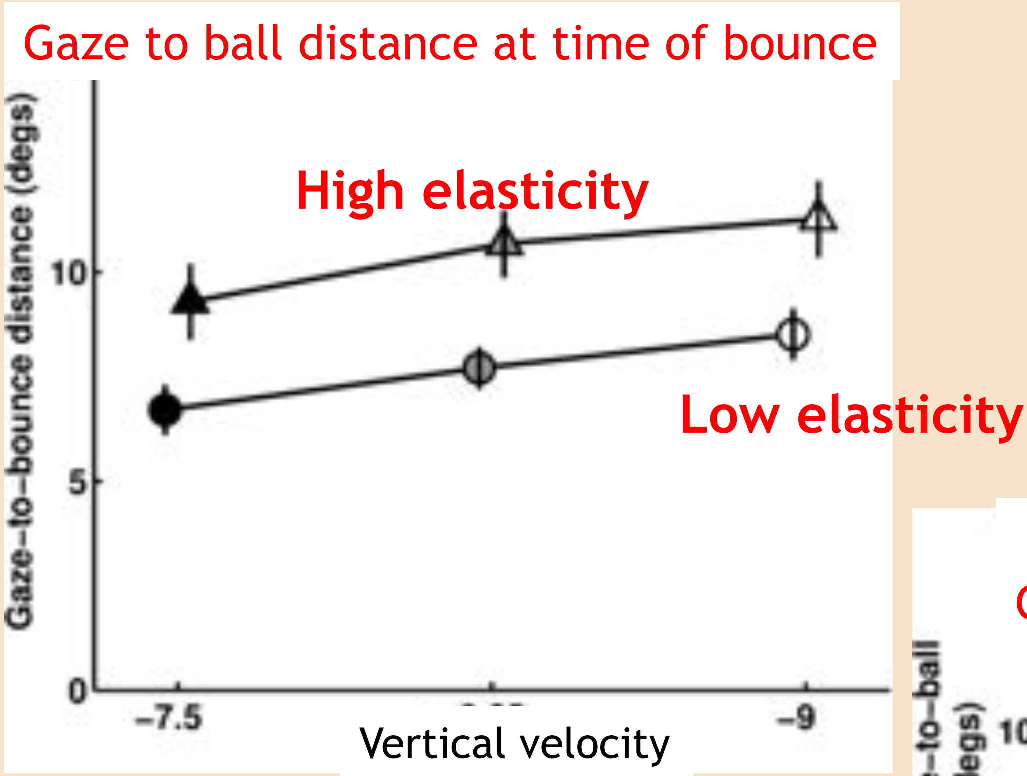
Ball location relative to gaze 150 msec later.

Ss adjust predicted gaze point for elasticity and velocity.

Location of saccade precisely adjusted to compensate for elasticity and pre-bounce velocity

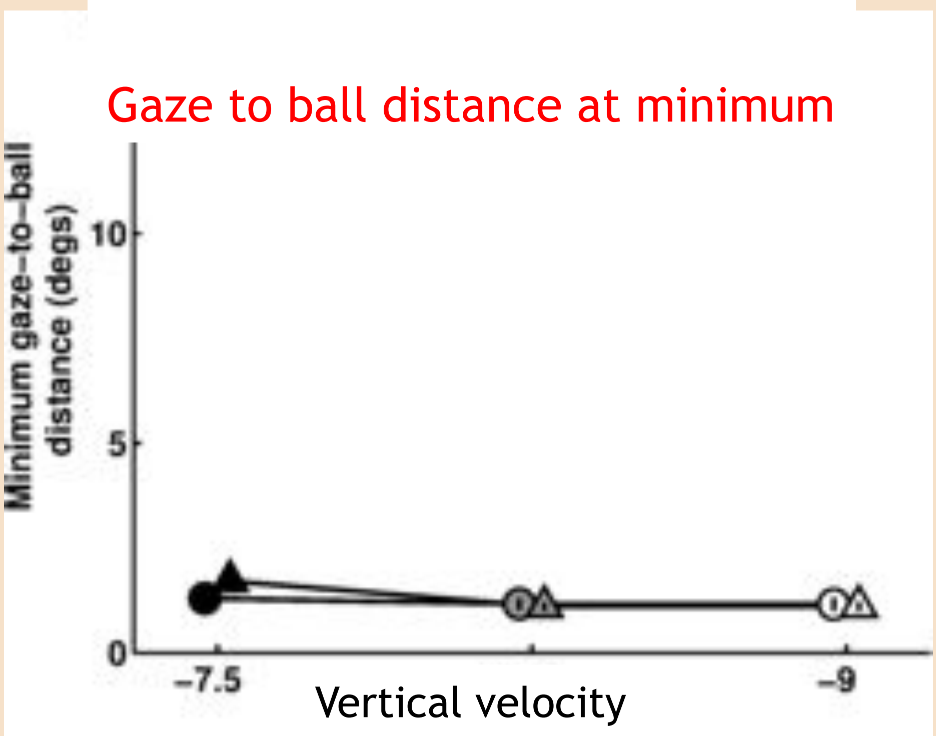


Predictive Saccades: Location

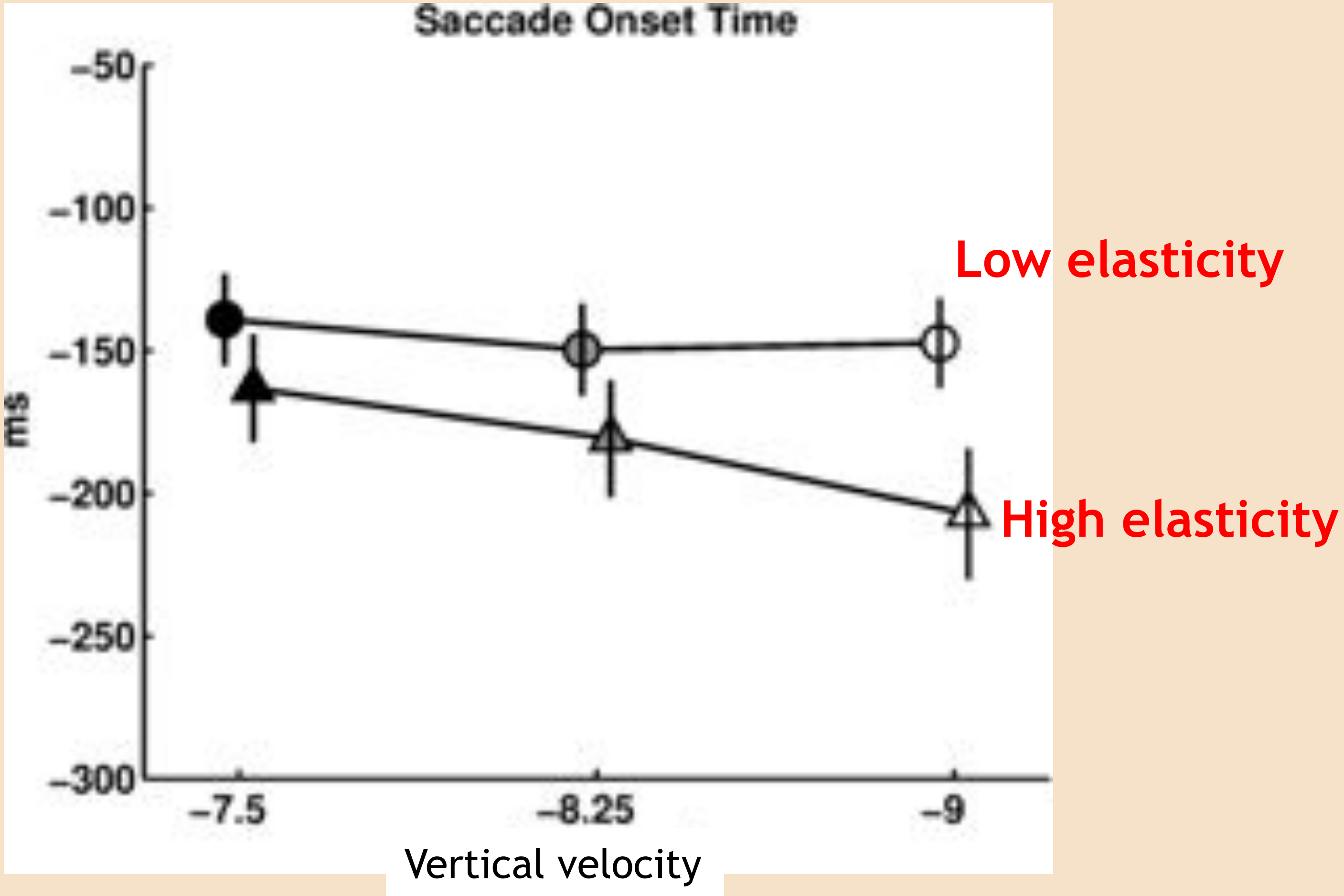


Subjects saccade to location above the bounce point.

Ball then passes close to gaze. Location of saccade precisely adjusted to compensate for elasticity and pre-bounce velocity



Predictive Saccades: Timing



Earlier saccade for more elastic balls (prior trials).
Earlier saccade for high velocity balls (current trial).

Internal Visual Models Allow Prediction

Anticipatory saccades reveal that gaze is planned for a predicted state of the world.

Predictions must be based on some kind of internal model/prior of visual events.

Internal Models

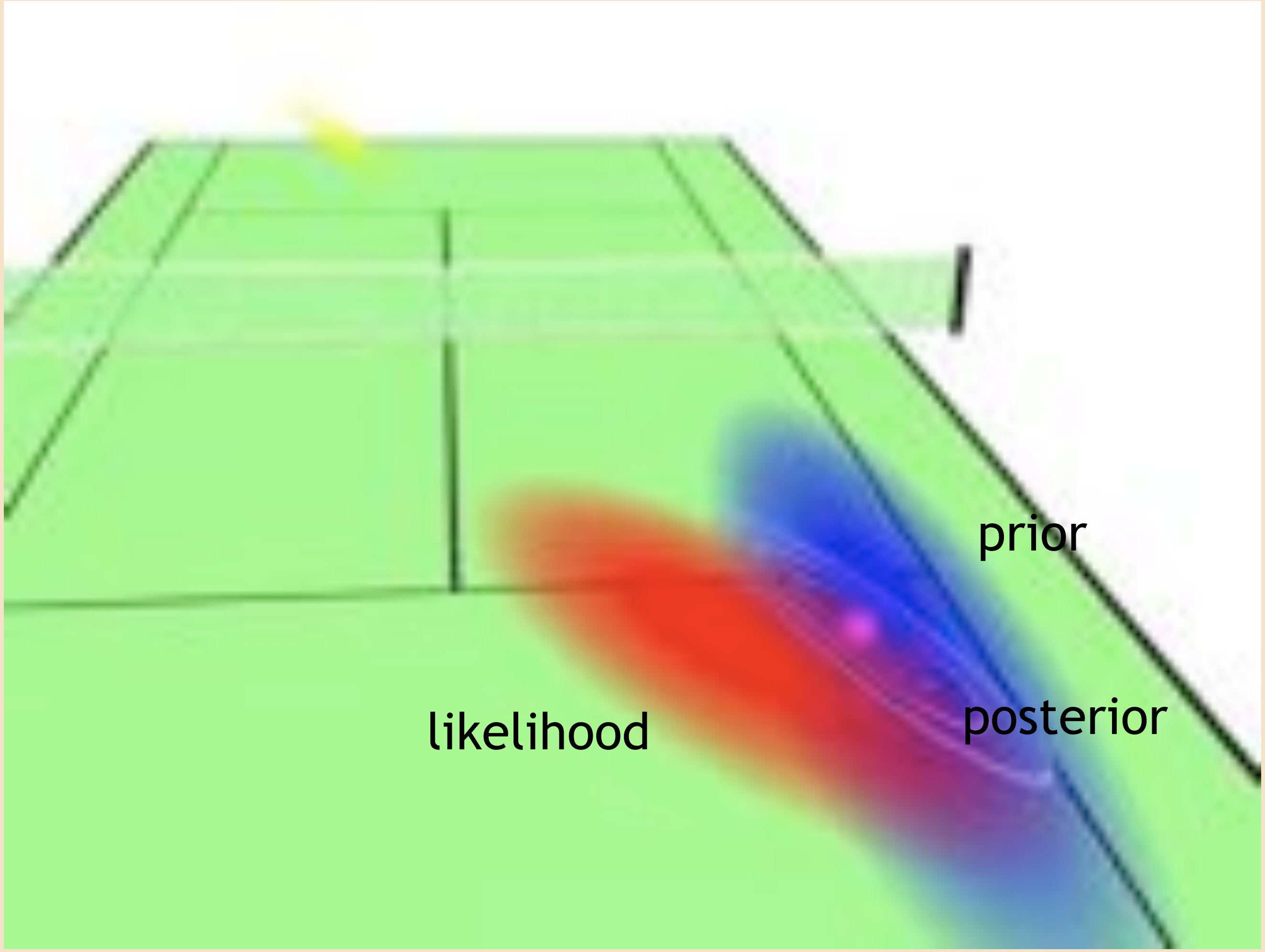
What do we know about the internal model? Evidence suggests it is high level and complex (angle, speed, elasticity, 3D, gravity).

In addition to mitigating the problem of visual delays, another value of experience-based internal visual models is that it allows better coordination of eye, head, arm, and body movements.

In reaching, evidence for the optimal Bayesian integration of current visual information with visual priors, (Koerding & Wolpert, 2004; Brouwer & Knill, 2007; Tassinari et al, 2006)

Perhaps a similar optimal weighting occurs with saccadic eye movements. Ie targeting has an image based and a memory based Component.

Hypothesis: Bayesian prediction of future state



Wolpert, 2005

Summary so far

Complex behavior can be broken down into sub-tasks or modules. This is consistent with observations of natural behavior.

Execution of sub-tasks/modules is learned and is governed by reward and uncertainty about task-relevant information.
Supported by gaze allocation in walking and driving.

Learnt statistics/ priors about world state govern allocation of attention.
Supported by both walking and racquetball.

What is the role of the image?

Image properties eg contrast, edges, chromatic saliency can account for some fixations when viewing images of scenes (eg Itti & Koch, 2001; Parkhurst & Neibur, 2003). (Also attentional capture by sudden onsets etc Theeuwes et al 2001.)

How important is this in natural vision?

Important stimuli may not be salient.

Salient stimuli may not be important.

ie Image properties won’t necessarily serve behavior

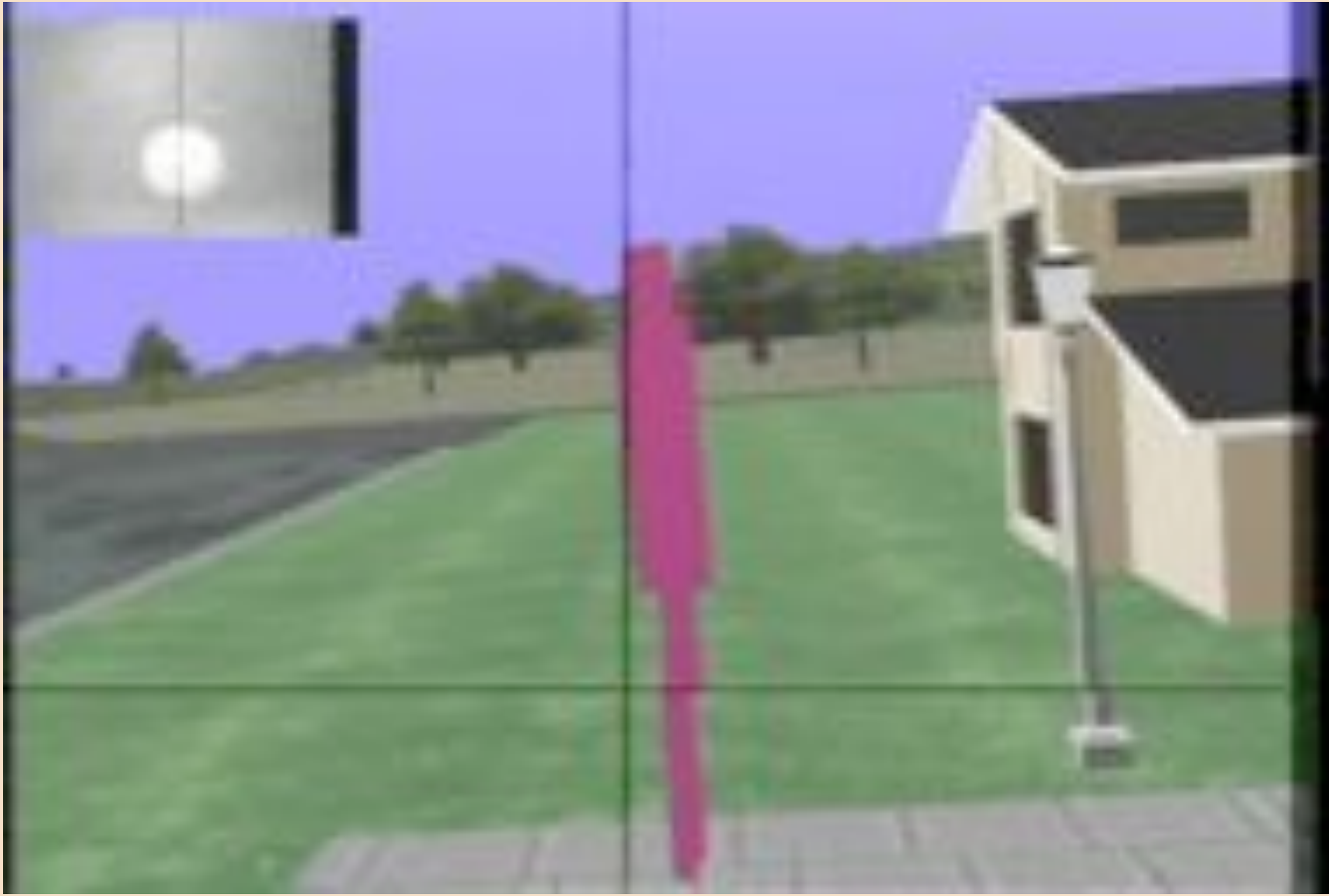
In free-viewing, the constraints are weak, and inferences about cognitive state are difficult.

However, need to be able to notice important stimuli that are not on the current task agenda.

Possible solution: attention may be attracted to deviations from expectations based on memory representation of scene.

Most scenes are highly familiar. Therefore extensive memory representations can be built up and used as a basis for expectations.

Hacer Uke-Karacan 2007



A mechanism where gaze is attracted by deviations from expectation may serve as a useful adjunct to task-driven fixations.

Such a mechanism will be more robust than image saliency.

How can information about gaze be used?

Provides a safe environment for testing effectiveness of rehabilitation strategies in clinical situations.

Hemianopic subject: blind in left visual field as a consequence of damage to right visual cortex

Subject detects a moving object in the good field.

Subject misses a moving object in the bad field.

By measuring the probability of detection both pre and post training one can evaluate the effectiveness of rehabilitation methods.

Importance of the Paradigm

“Free viewing” paradigm may not reflect what happens in natural behavior.

Example from schizophrenic patients

When viewing static displays, schizophrenic patients make fewer saccades around the display and have longer fixations.

Work by Boucart & Delerue shows much more normal patterns in the context of behavior.

Gaze may be tightly orchestrated by the task – eg racquetball

In other cases, gaze is less well constrained, but still regular
- eg driving, walking.

Other cases such as looking at pictures:

- optimality is not well defined

- fixations do not map very clearly onto cognitive operations

Gaze behavior is learned via reward machinery of brain.

This is consistent with RL models of gaze control that Incorporate uncertainty.

Scenes are learned also and are often very familiar.

How much can be learnt from gaze behavior will depend on how tightly the context determines optimal behavior
eg racquetball versus free-viewing.

In well-constrained contexts subjects behave similarly, and performance is stable.

Many potential applications eg clinical for eye tracking.

Acknowledgments:

Jelena Jovancevic – walking
Gabe Diaz – virtual racquetball
Brian Sullivan – driving
Hacer Uke-Karacan – memory for scenes
Krystel Huxlin – hemianopes