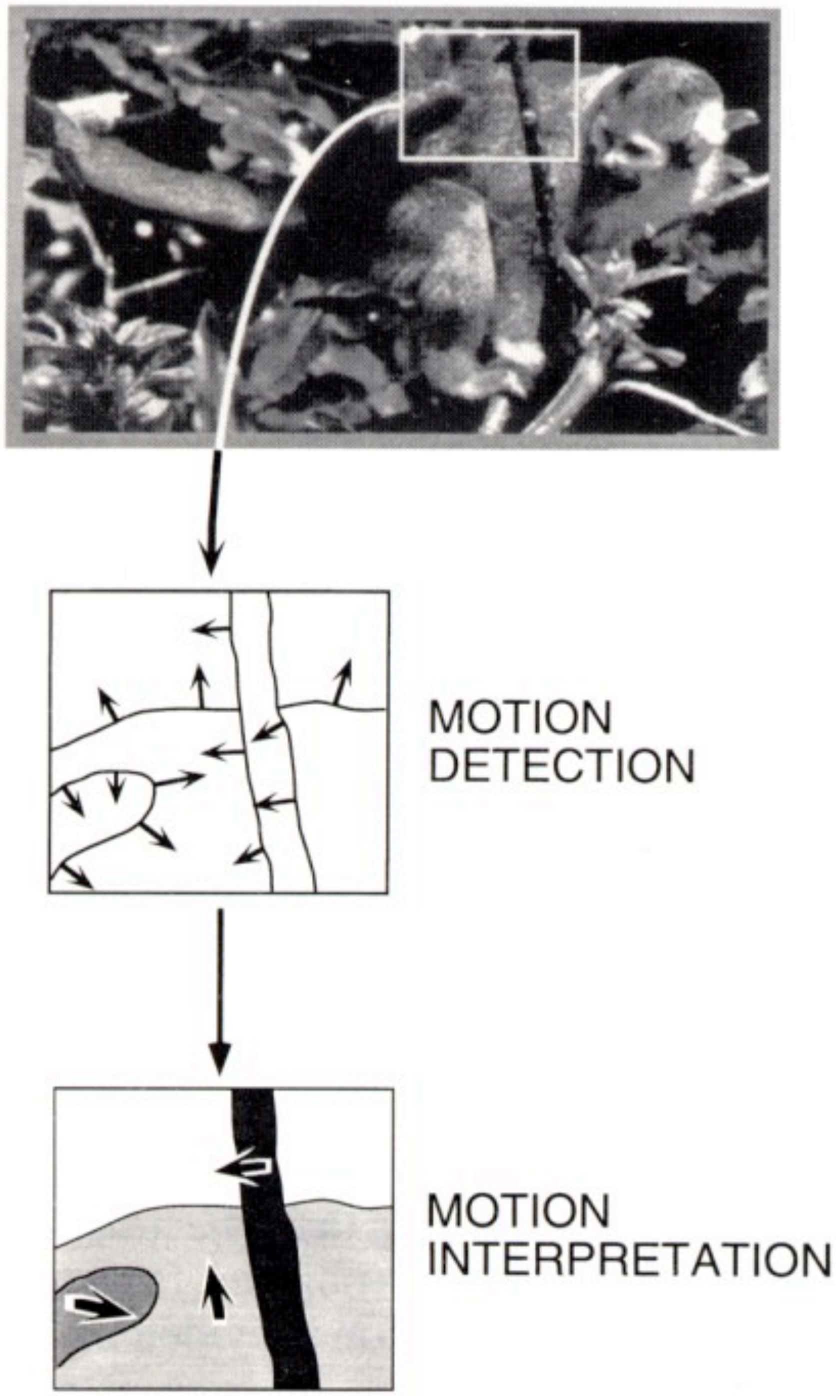
Computational Perception

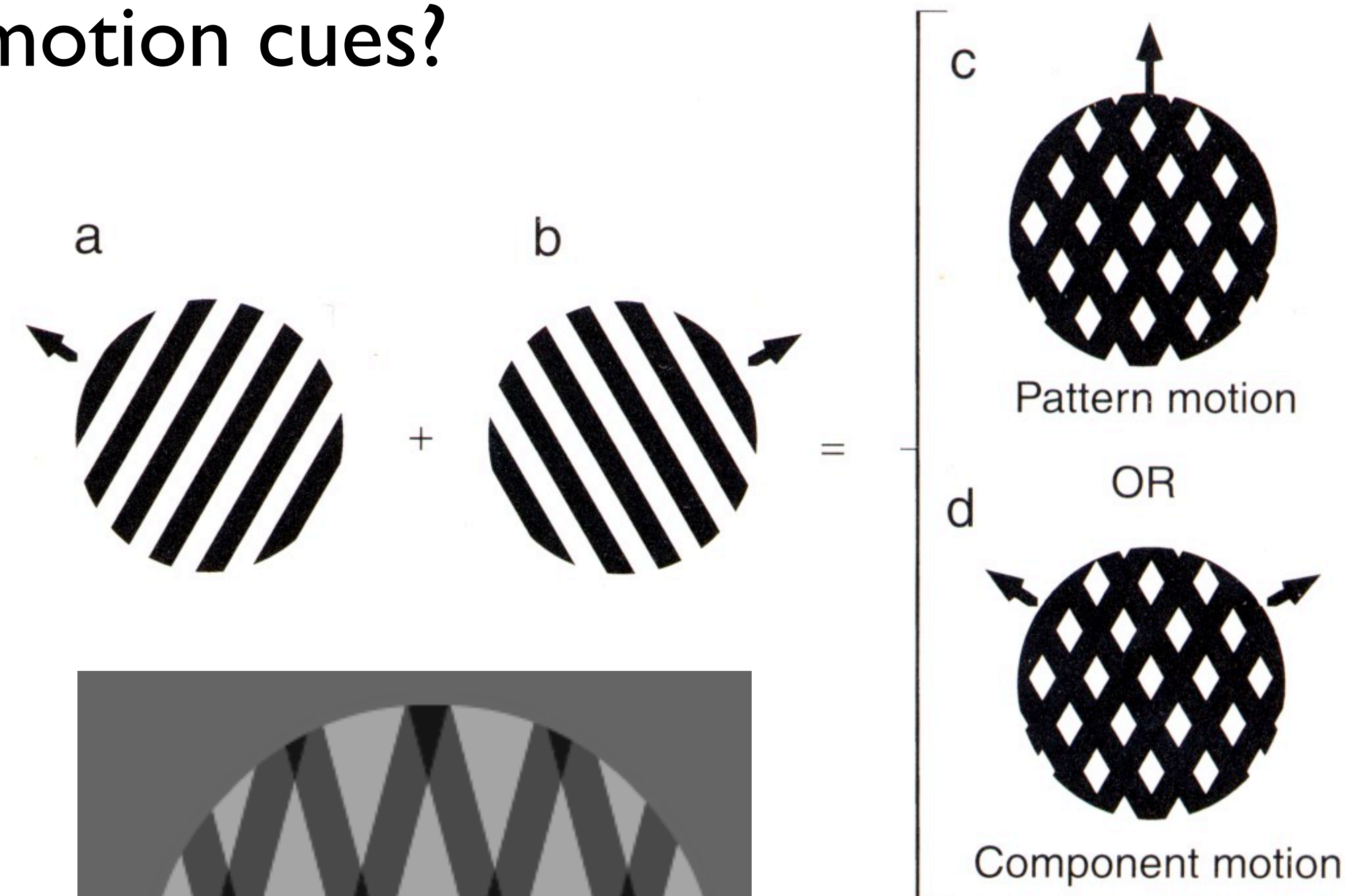
15-485/785

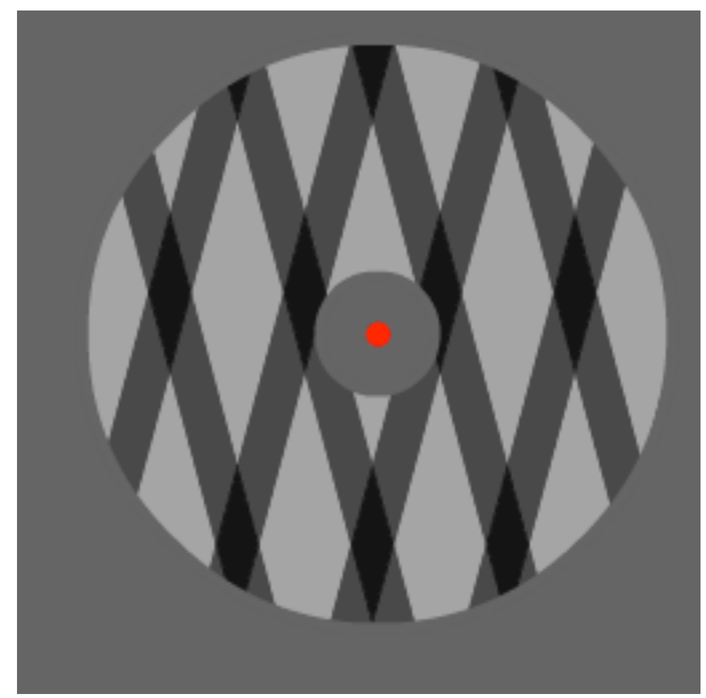
Motion 2

How do we integrating local motion cues?



from Albright and Stoner, 1995

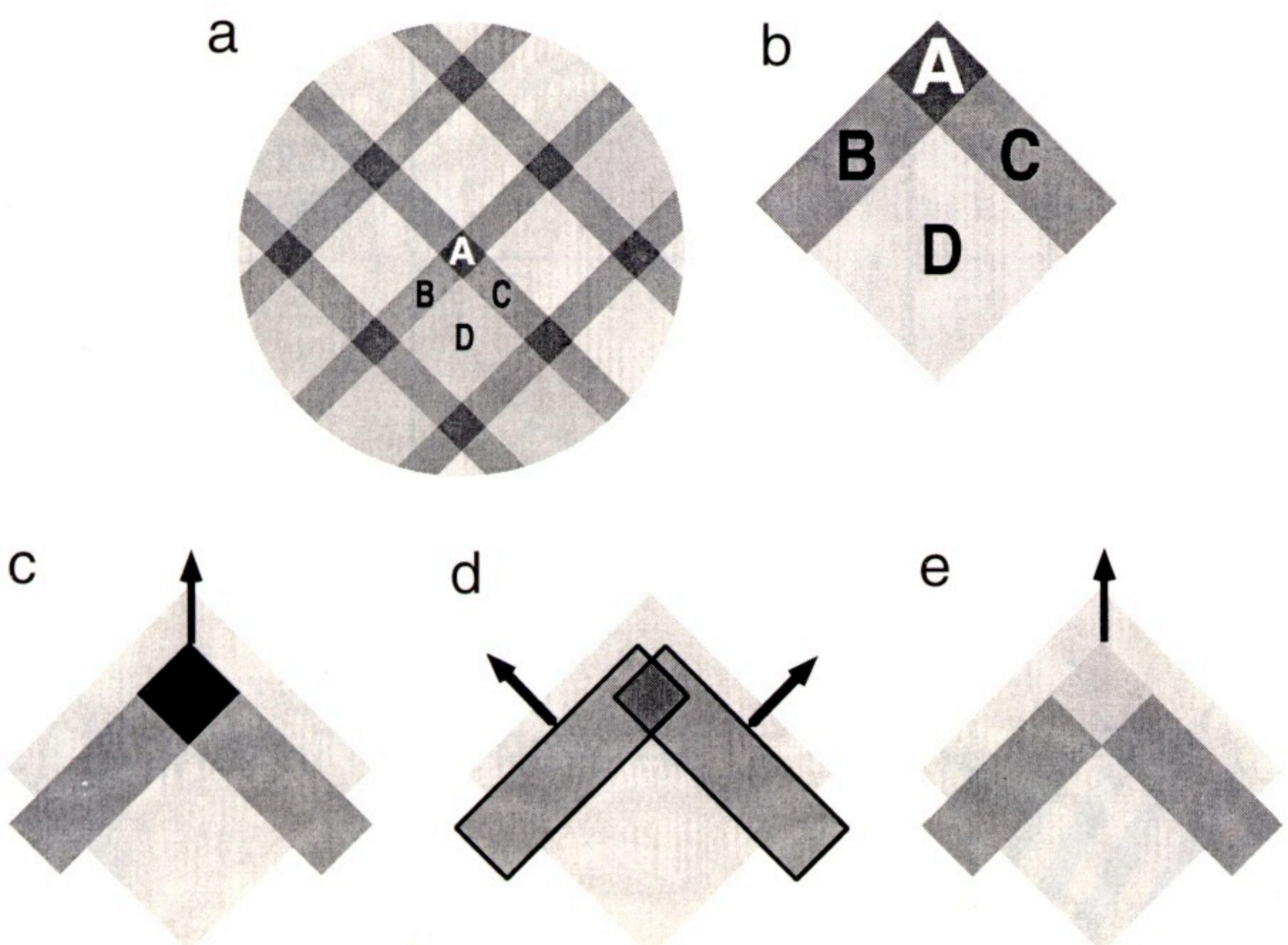


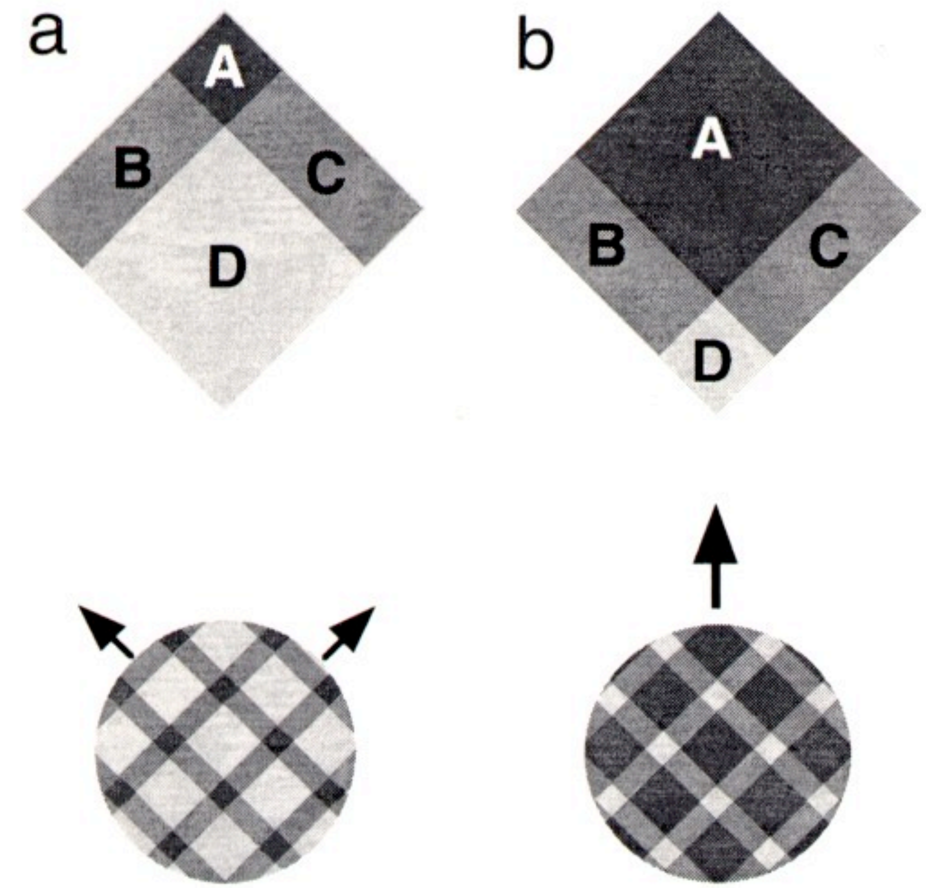


from Jean-Michel Hupé, NYU

Local motion cues do not necessarily correspond to motion of object.

Some cues that influence the percept





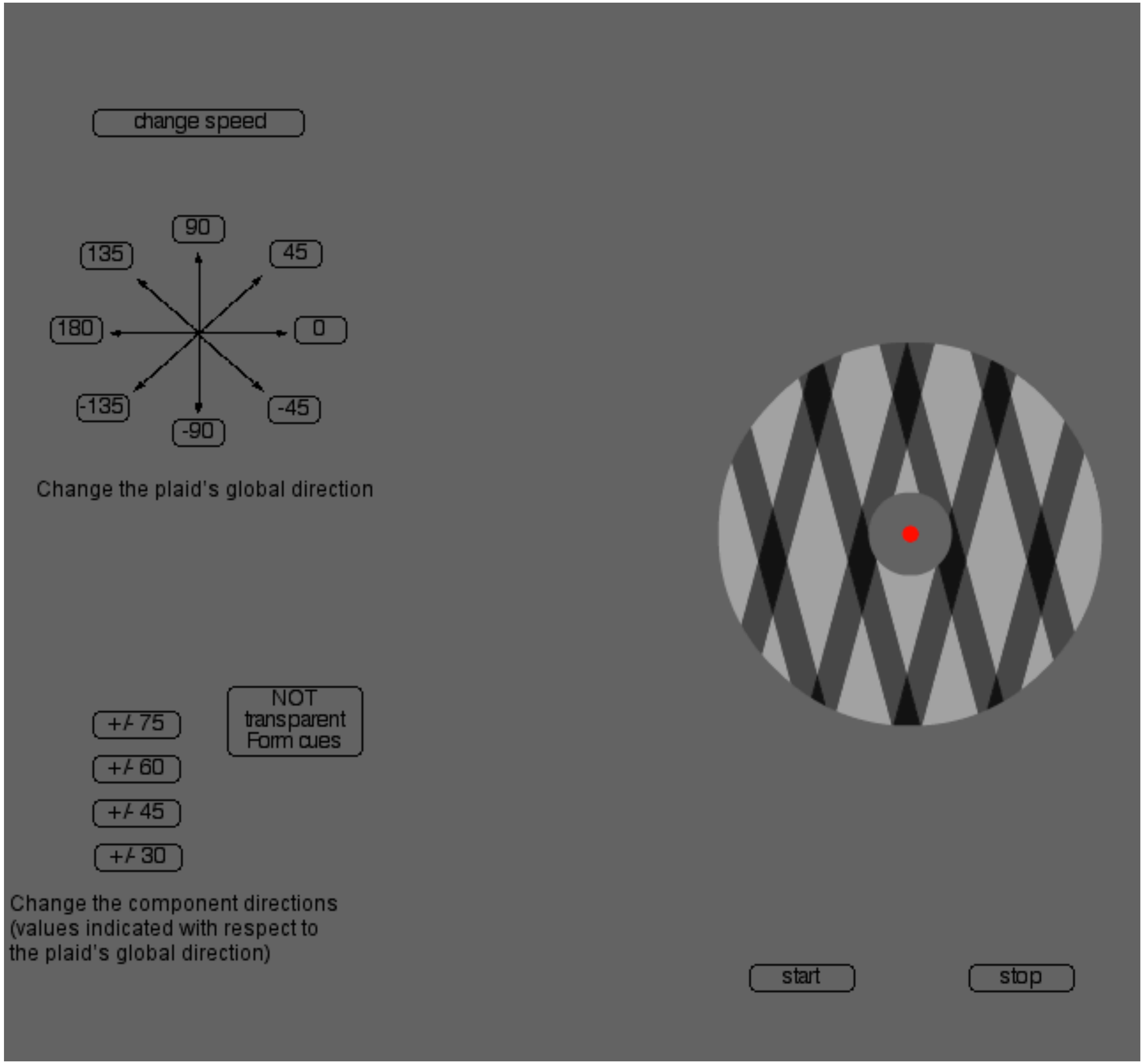
pattern motion

component motion

pattern motion

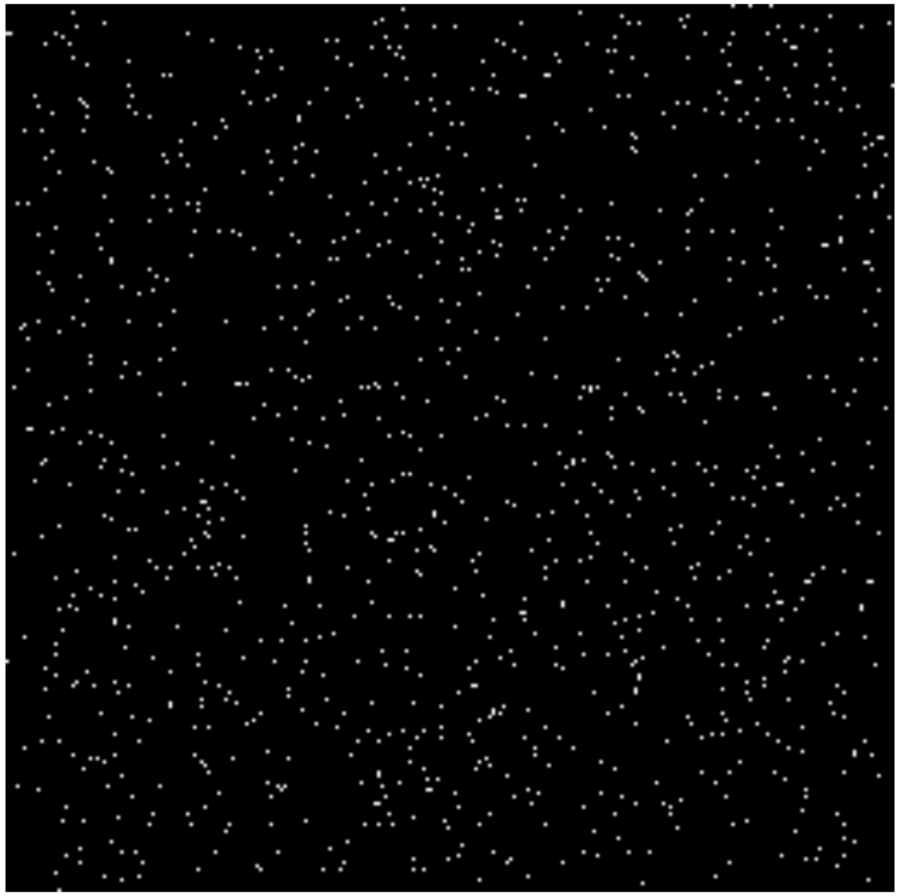
from Albright and Stoner, 1995

Transparent plaid motion demo (Jean-Michel Hupé)



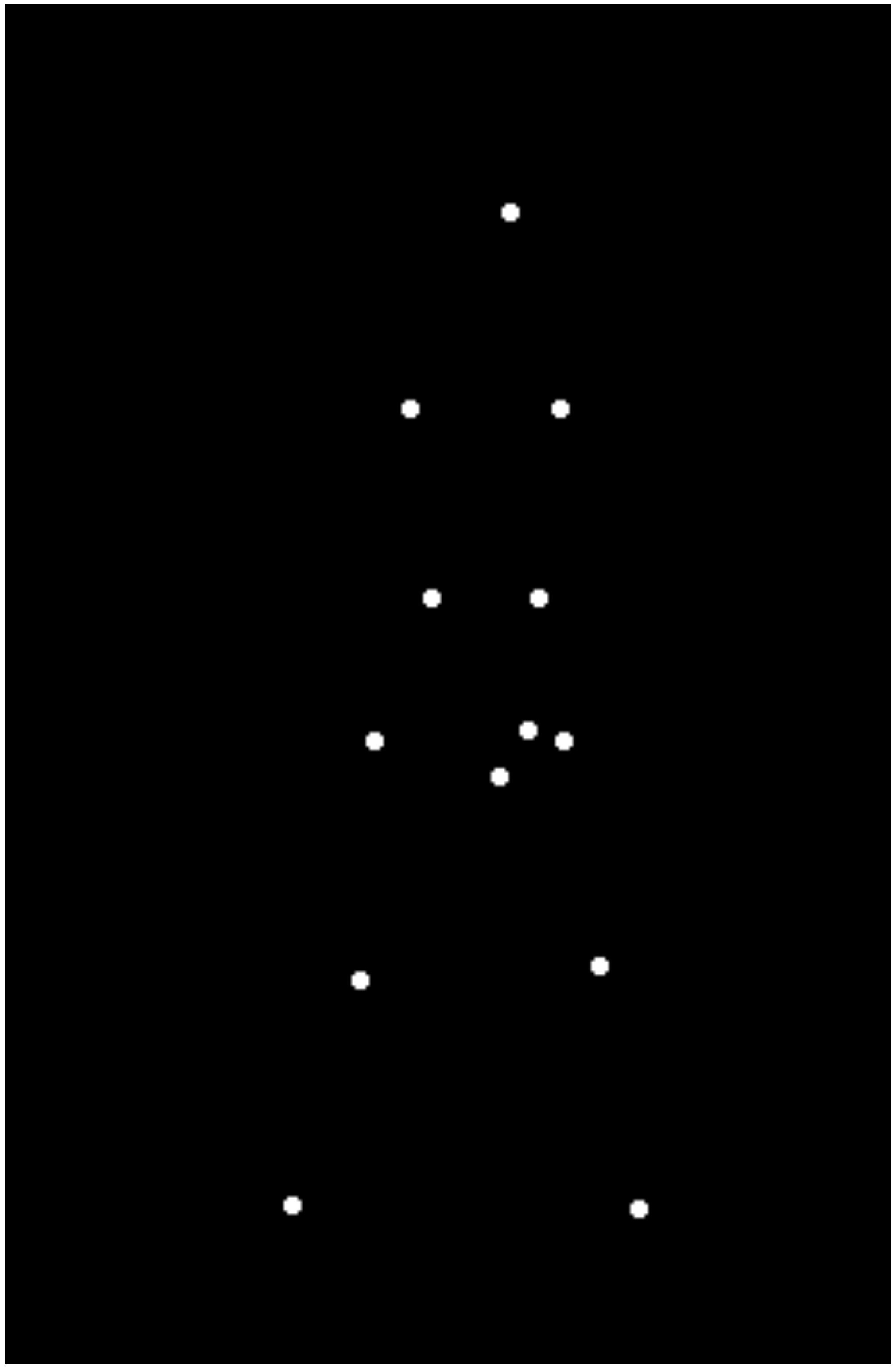
http://www.cns.nyu.edu/~hupe/plaid_demo/demo_plaids.html

Transparent motion with separate components



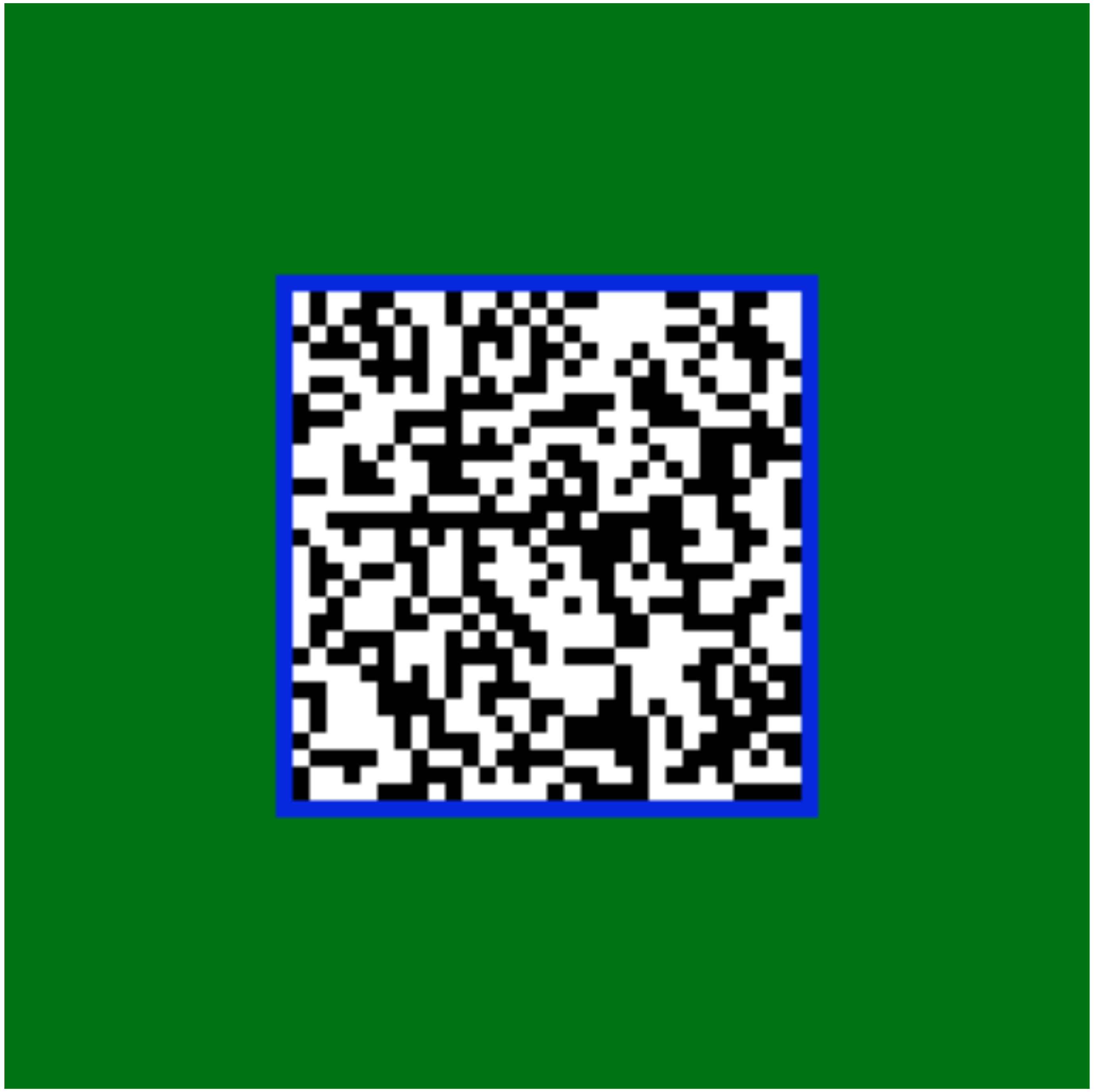
demo from George Mather

Biological motion



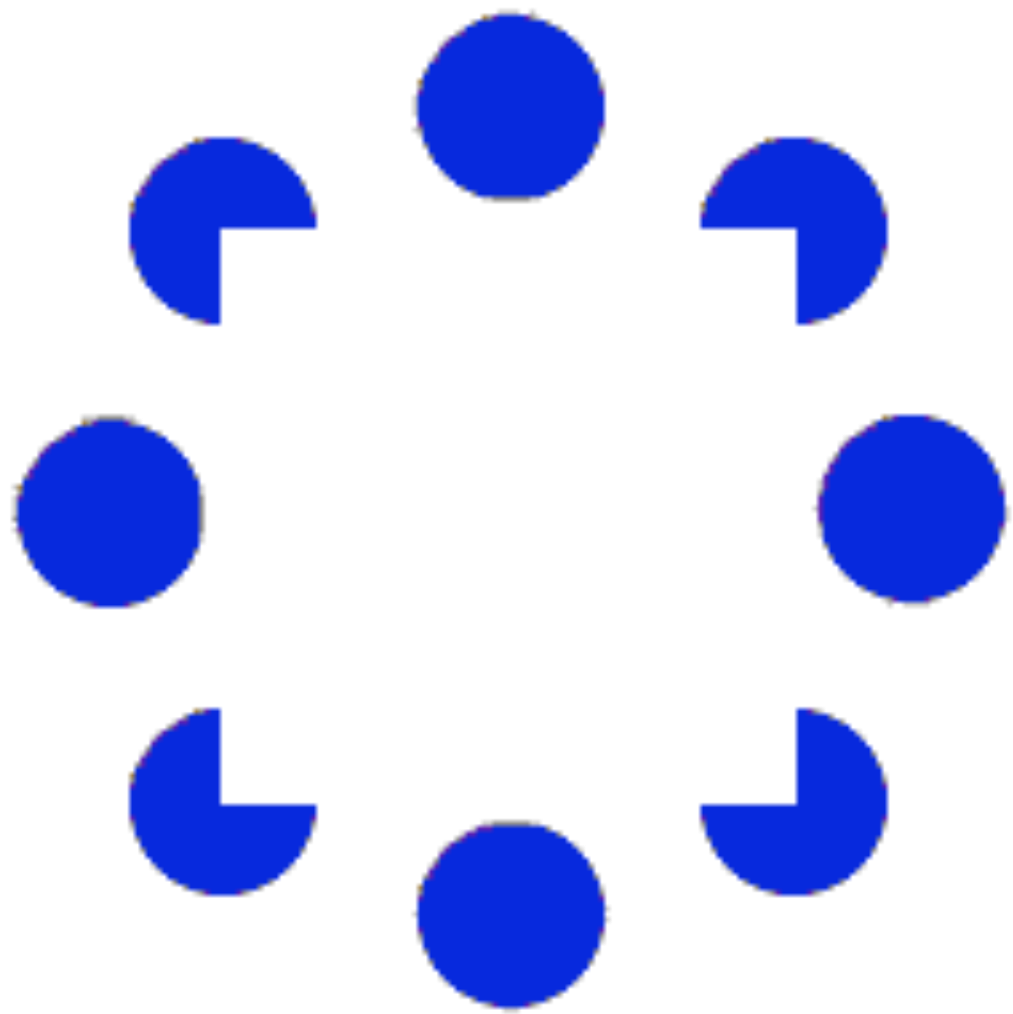
demo from George Mather

Second-order motion



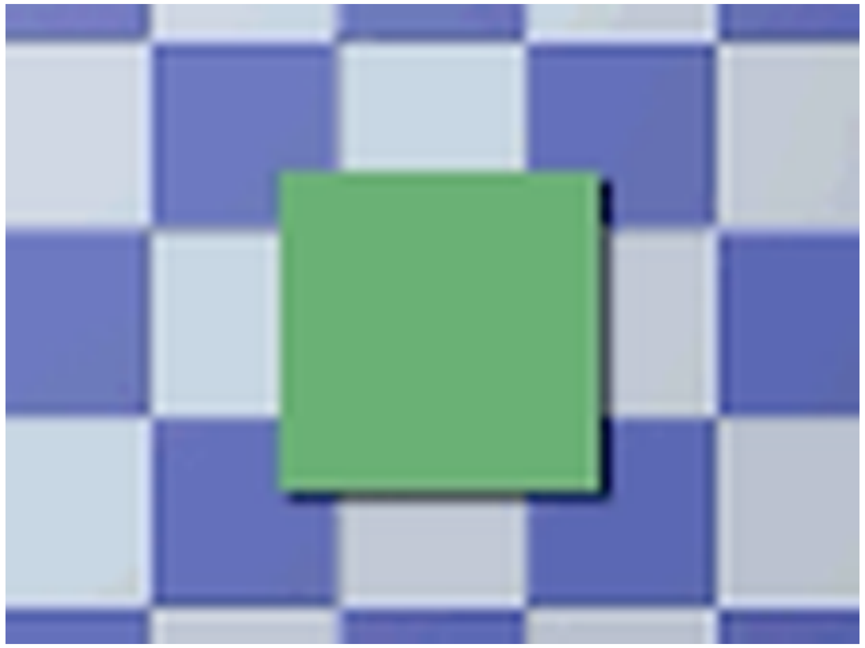
demo from George Mather

Implicit figure motion



demo from George Mather

Shadow motion: Rising square



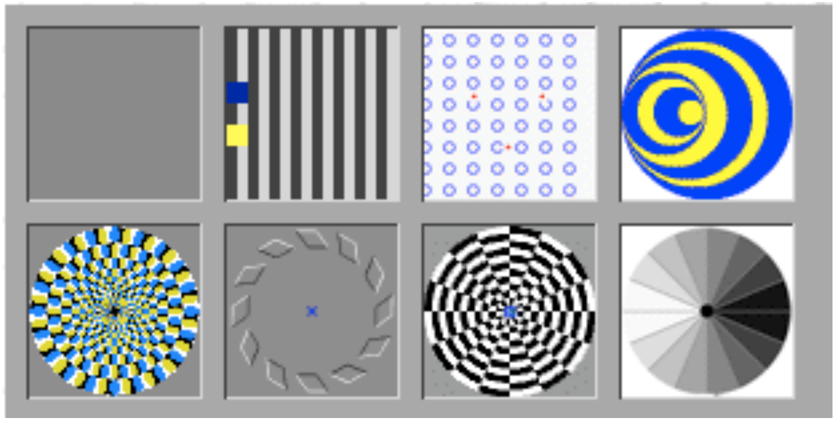
demo from Dan Kersten

Shadow motion: Ball in a box

demo from Dan Kersten

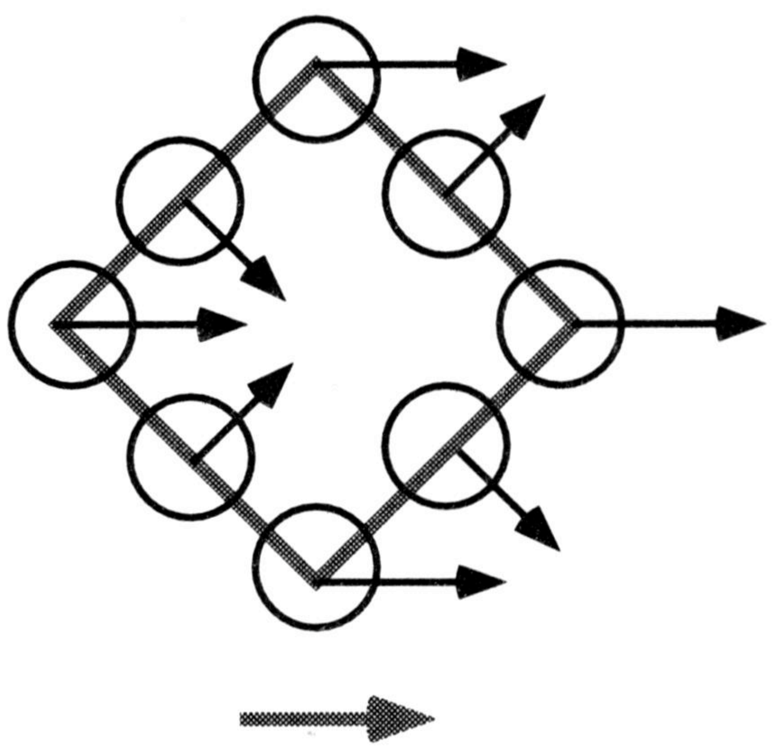
More optical illusions

Michael Bach's Illusion Site:
http://www.michaelbach.de/ot/



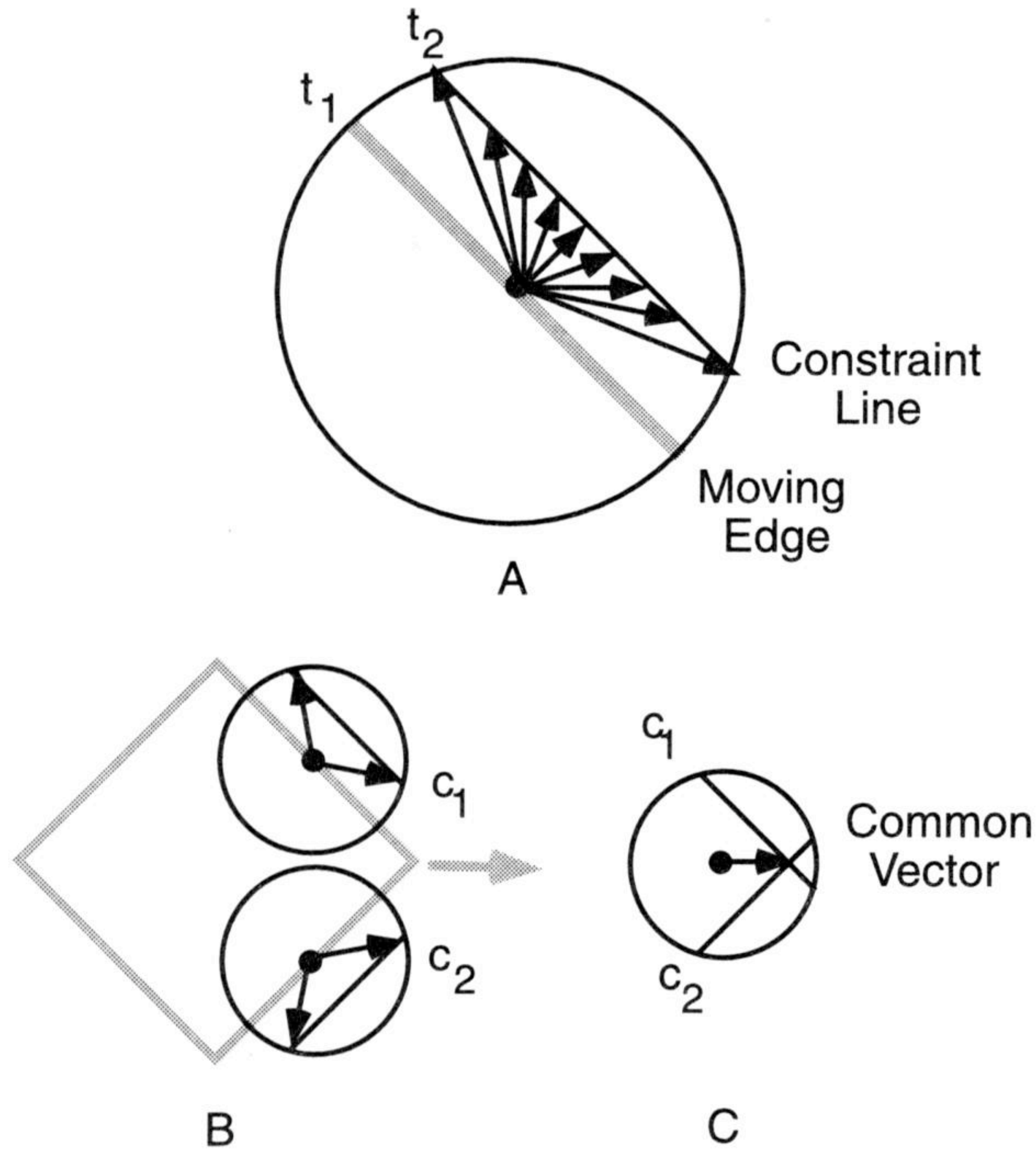
Resolving ambiguity in local motion estimates

Local estimates of motion



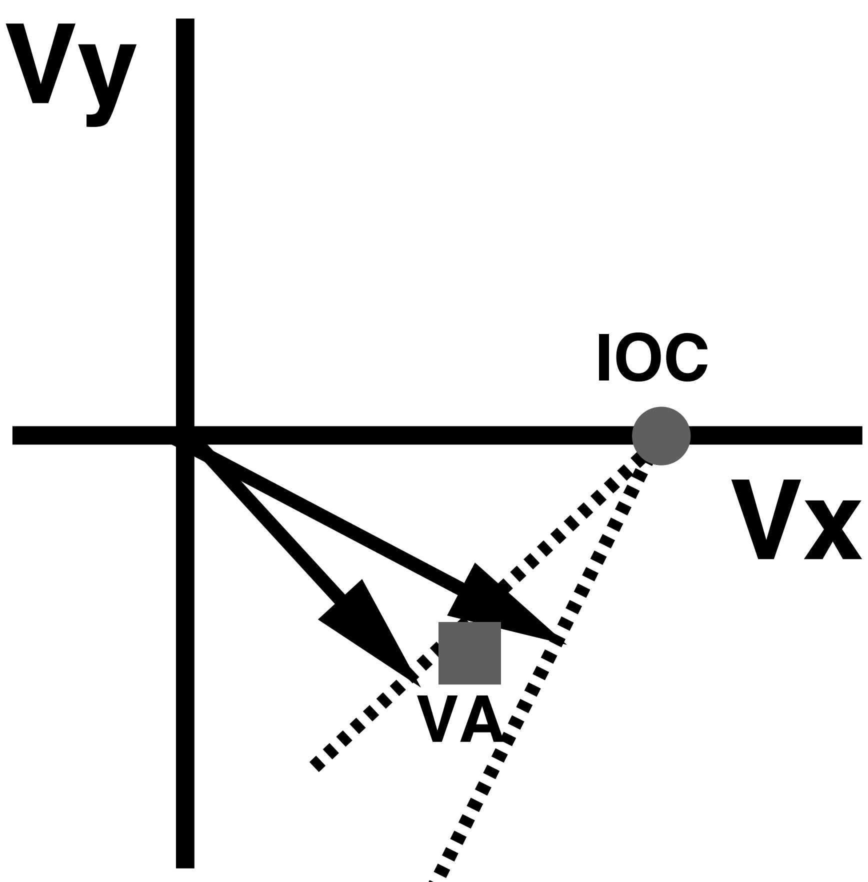
Actual motion

Intersection of constraints



Estimating pattern motion from components

Note axes are Vx and Vy



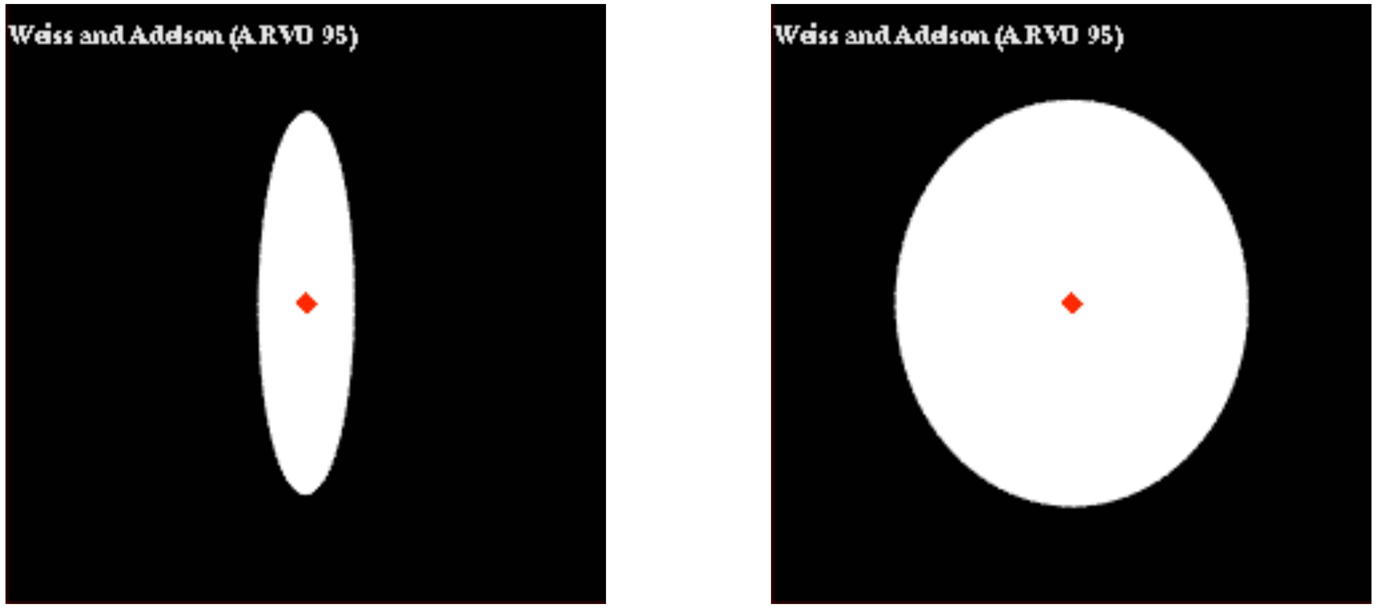
- How to integrate local cues?
  - intersection of constraints (IOC)
  - vector averaging (VA)
  - "blob" tracking
- perception consistent with all three
- influenced by many factors including:
  - orientation
  - contrast
  - presentation time
  - foveal location
- IOC and VA models make different predictions
  - Which is right?
  - Most motion perception models depend on post-hoc explanations

A Bayesian model of motion preception (Weiss and Edelson, 1998)

- Two key properties:
  1. combine different observations while taking into account their uncertainties
  2. estimates are obtained by combining observations with prior knowledge

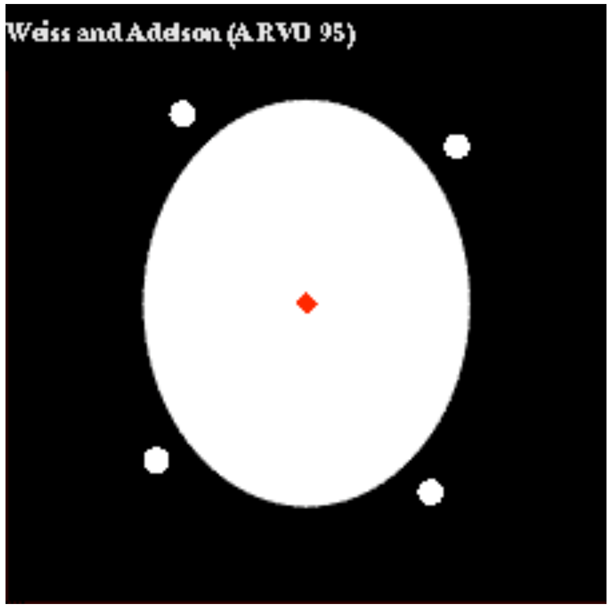
- Outline of Weiss and Adelson
  - a schematic explaination in (Vx,Vy) space.
  - a more general model
  - application to real patterns

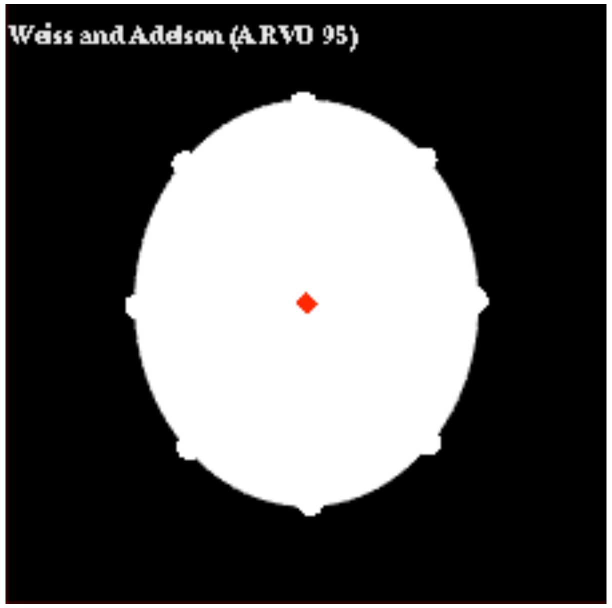
Motion illustion demonstrations on ellipses



- The rotation of the two ellipses is identical, however:
  - narrow ellipse appears rigid
  - fat ellipse seems to deform
- Why?

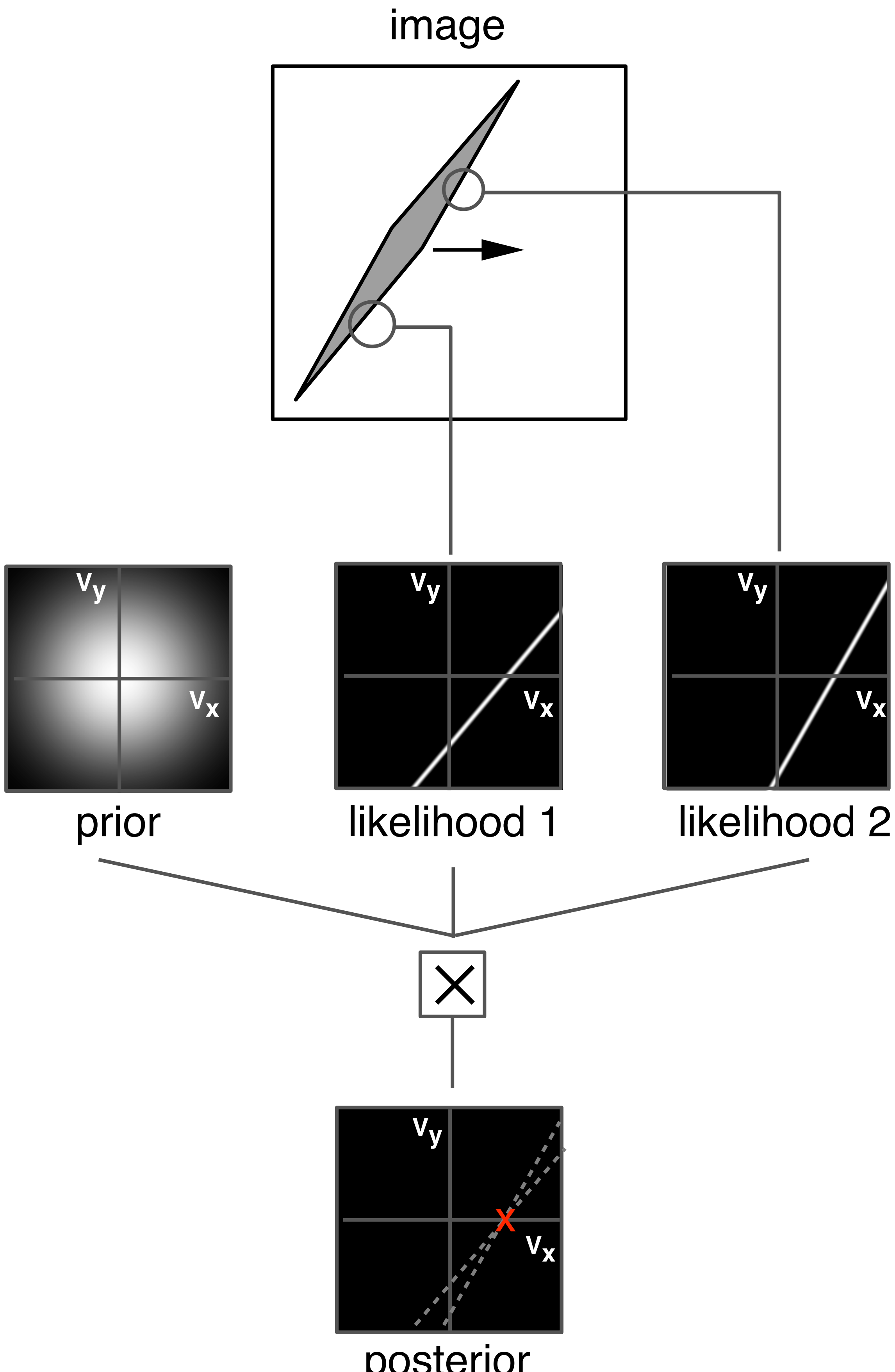
Adding texture changes percept





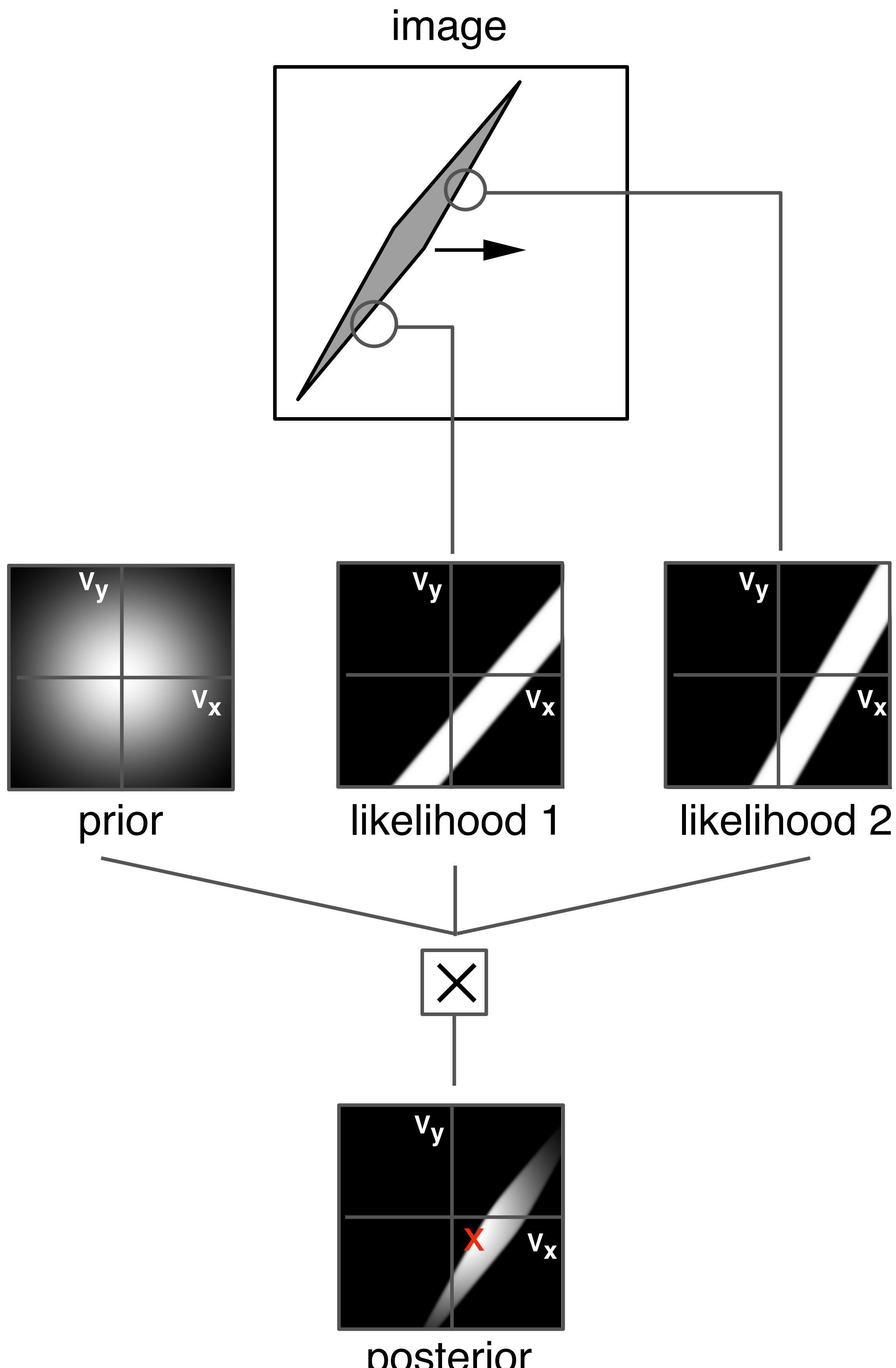
- In neither case does the shape of the ellipse change during the animation.
- Observation:
  - If dots move with ellipse, it now appears rigid.
- Why?

A simple Bayesian estimator for velocity



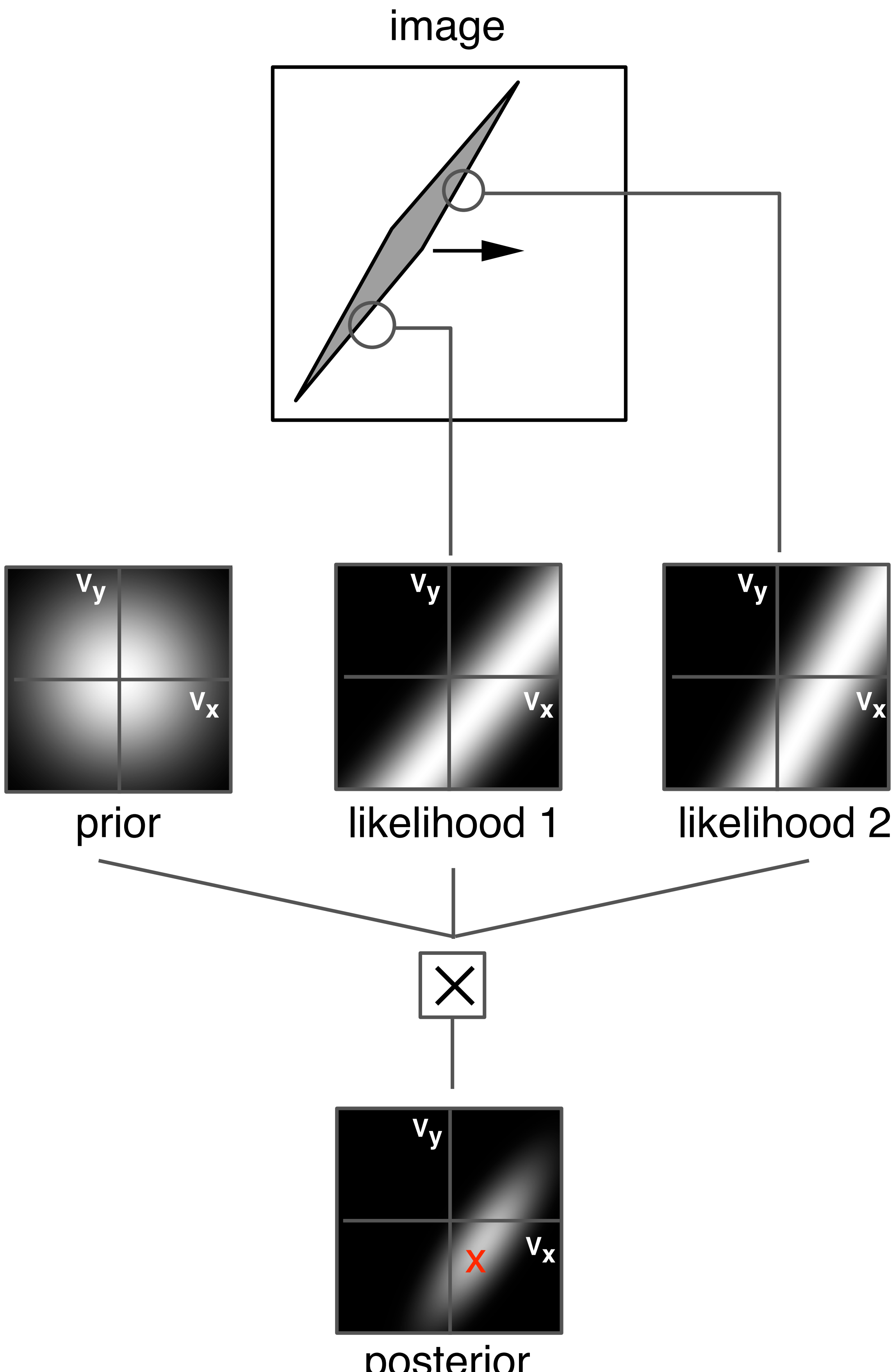
- Model integrates local motion estimates
- Likelihood is zero everywhere except on the constraint line.
- *What does the prior favor?*
- Maximum a posterior (MAP) estimate is IOC solution.
- Prior places no role. *Why?*

A simple Bayesian estimator for velocity



- Likelihoods non-zero within ε
- Note the effect of prior on MAP

A simple Bayesian estimator for velocity



- Likelihoods with Gaussian fall-off
- Posterior is also Gaussian

Limitations of simple model

- likelihoods are based on constraint lines

  ⇒ need to calculate likelihood from real images

- likehood functions consider only "1D" locations

  ⇒ need to define likelihoods for all images regions including "2D" features

- estimator assumes rigid translation

  ⇒ need inference for general motions (e.g., rotations and deformations)

Likelihood for a more general model

- Use weighted motion gradient constraint:

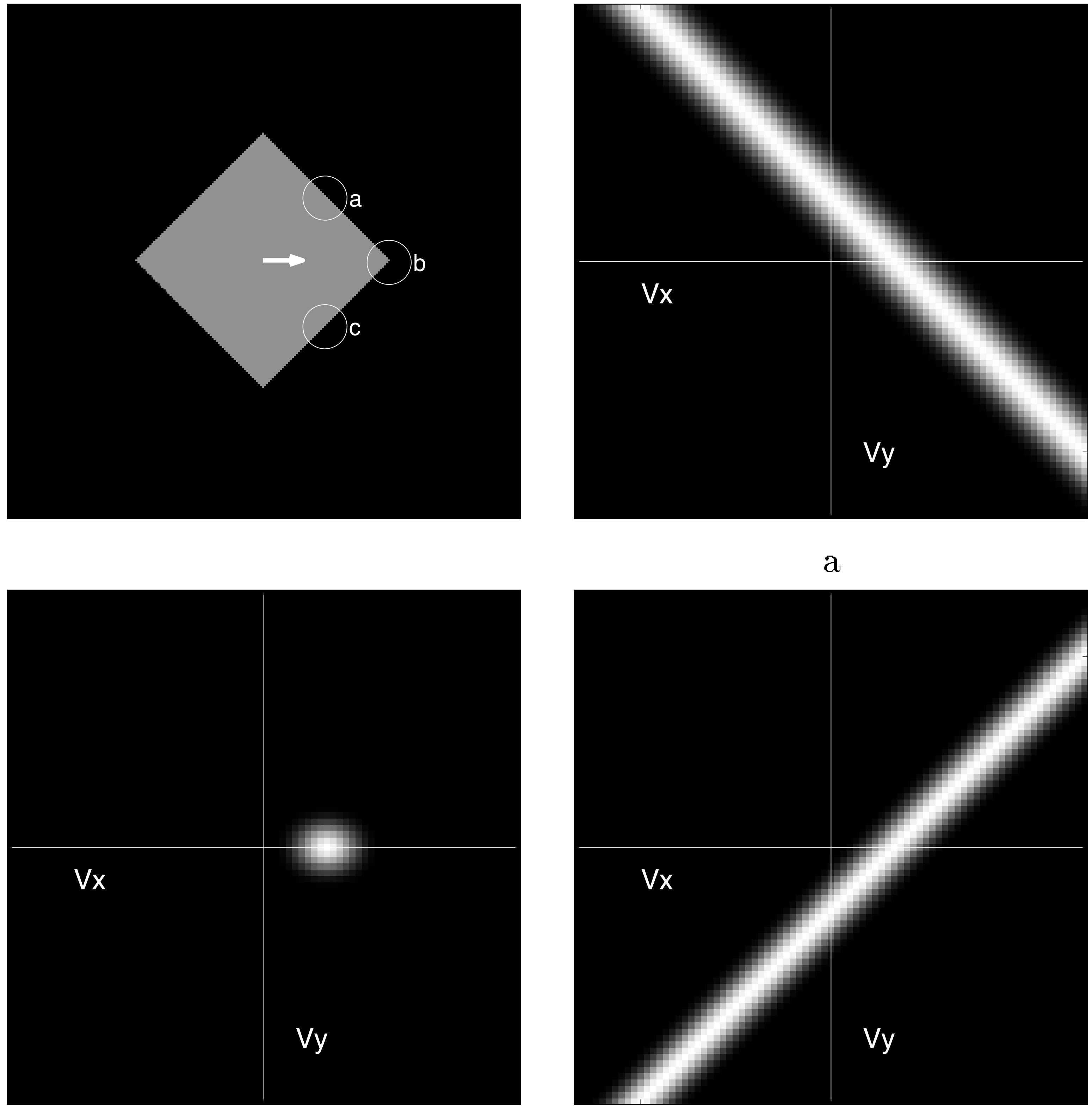
$$C(v_x, v_y) = \sum_{x,y,t} w(x,y,t)(I_x v_x + I_y v_y + I_t)^2$$

- Like Lucas and Kanade (1981), but w is a spatio-temporal window.
- The likelihood is defined as

$$L(v_x, v_y) = P(I_x, I_y | v_x, v_y) \propto e^{-C(v_x,v_y)/2\sigma^2}$$

- computes *probabilities* about local motion
- σ represents uncertainty in velecity estimate (e.g. from noise, low contrast, short time window, etc.)
- better satisfaction of constraint ⇒ higher likelihood
- higher noise ⇒ allow greater deviations from constraint

Illustration of motion likelihoods

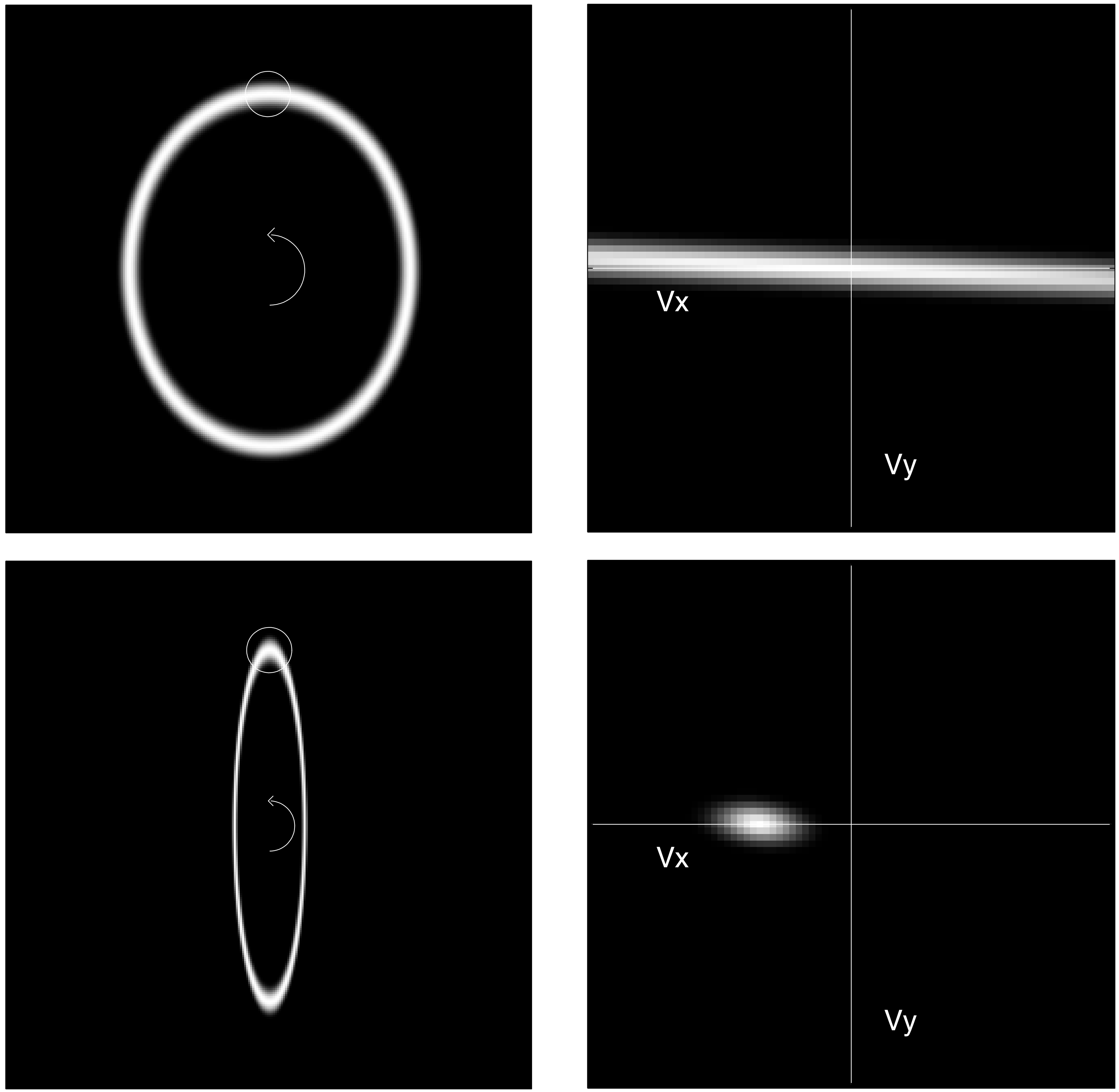


a

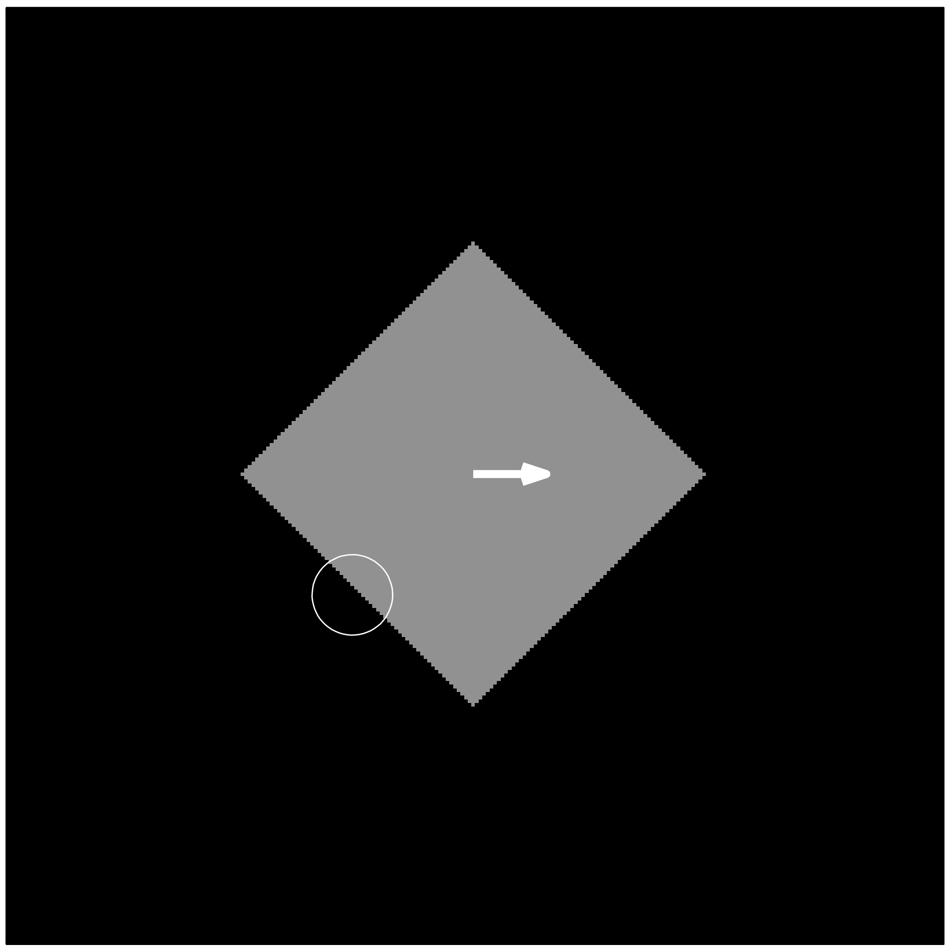
b

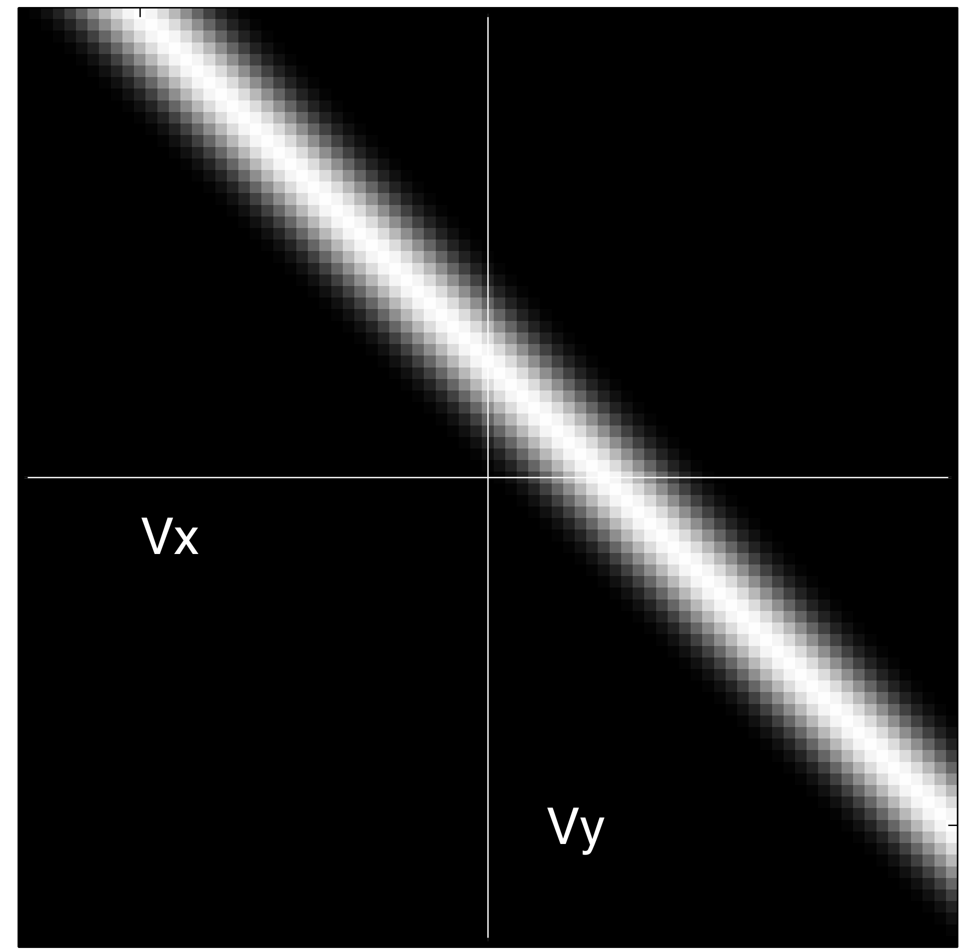
c

Likelihood for ellipses

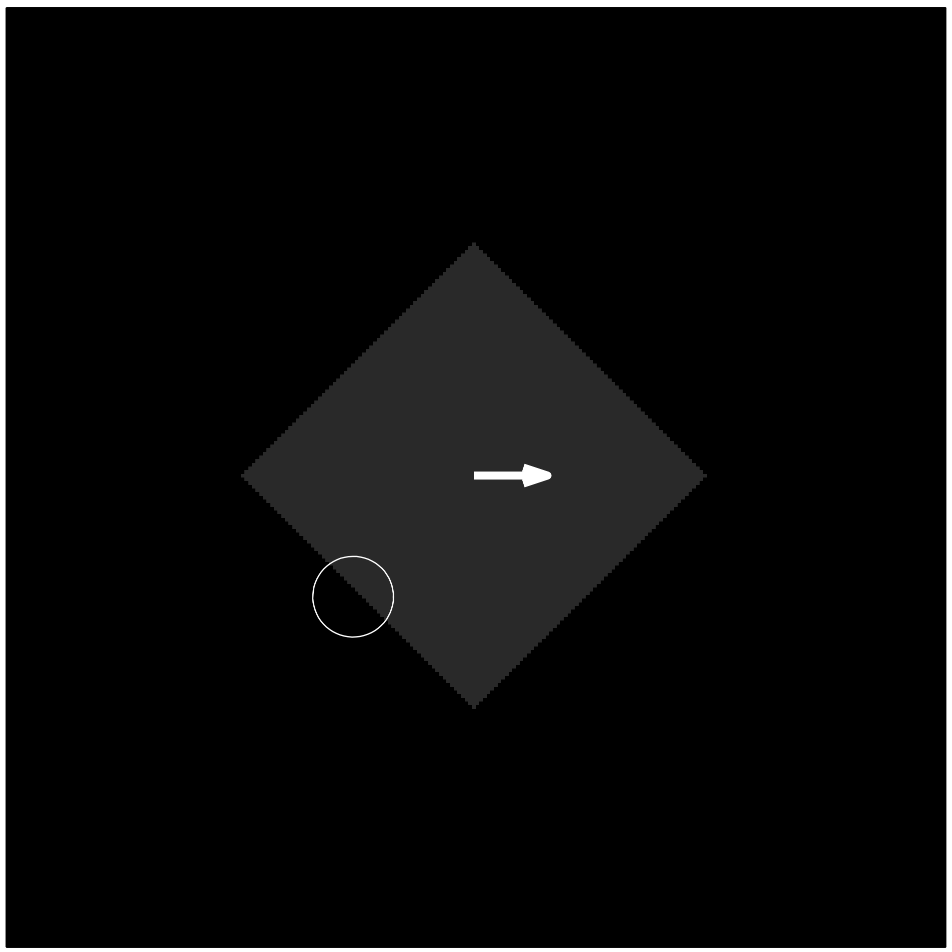


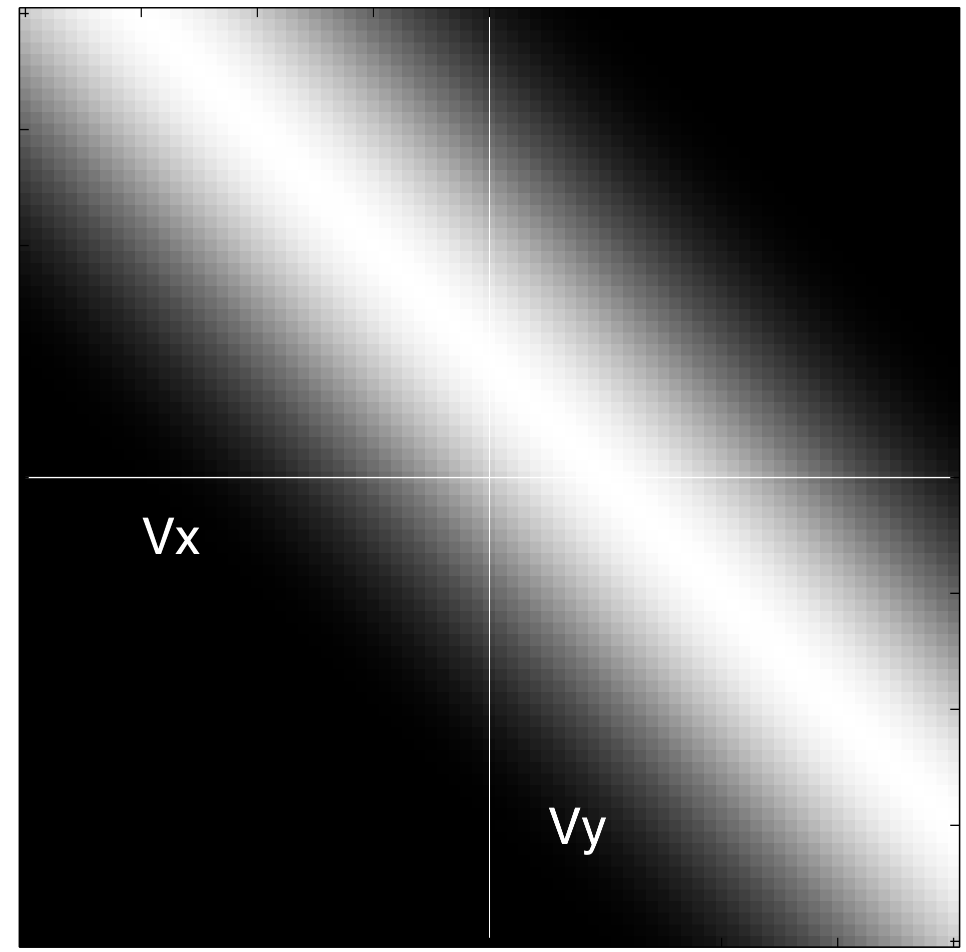
Likelihood for different contrasts



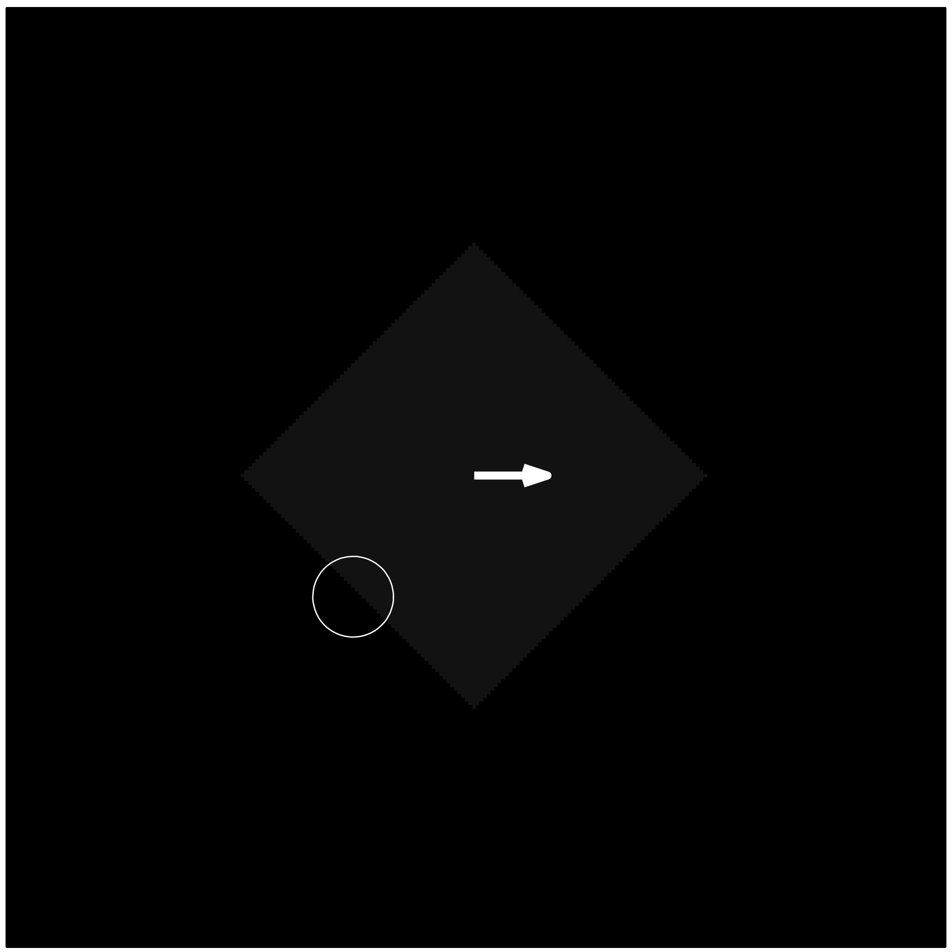


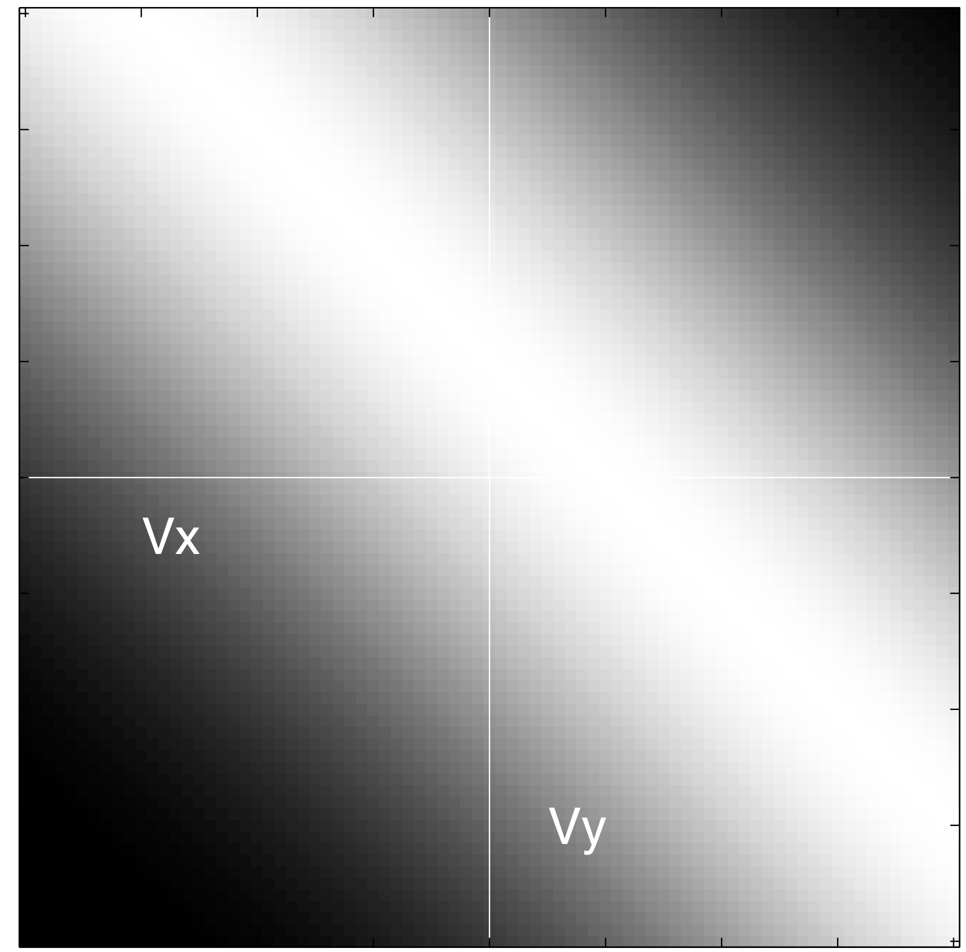
Likelihood for different contrasts



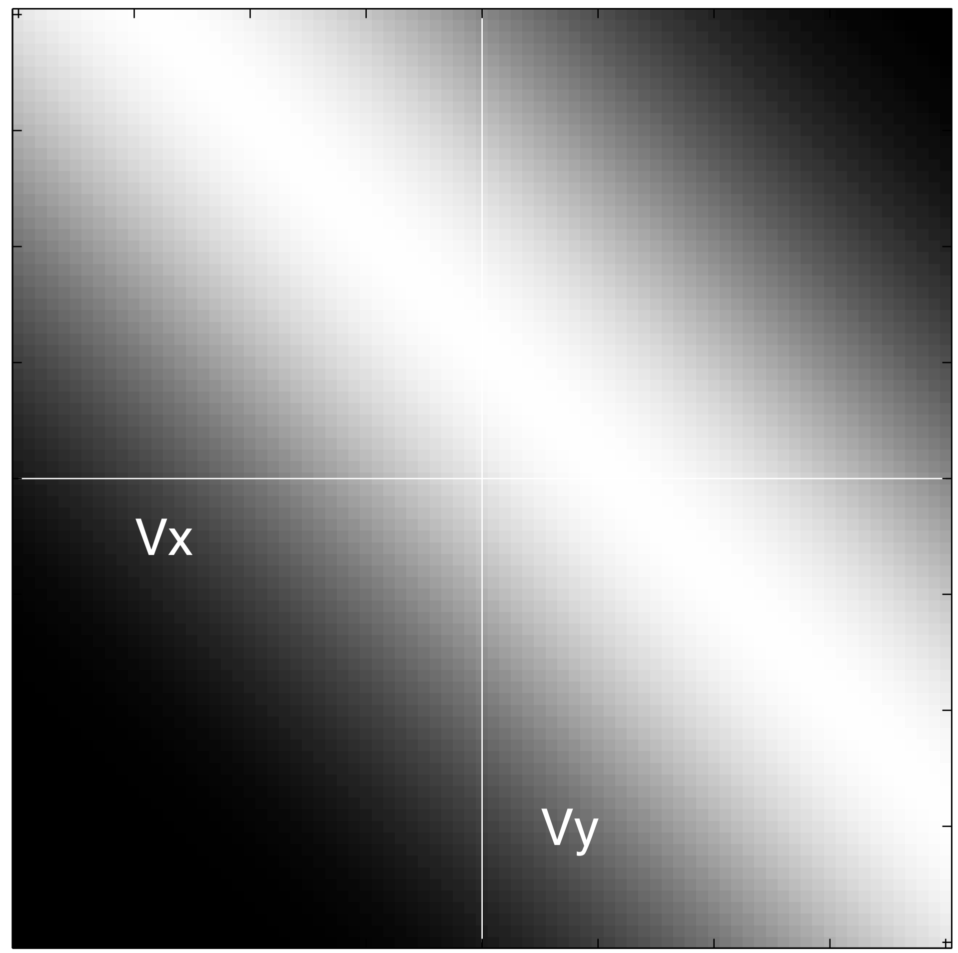


Likelihood for different contrasts



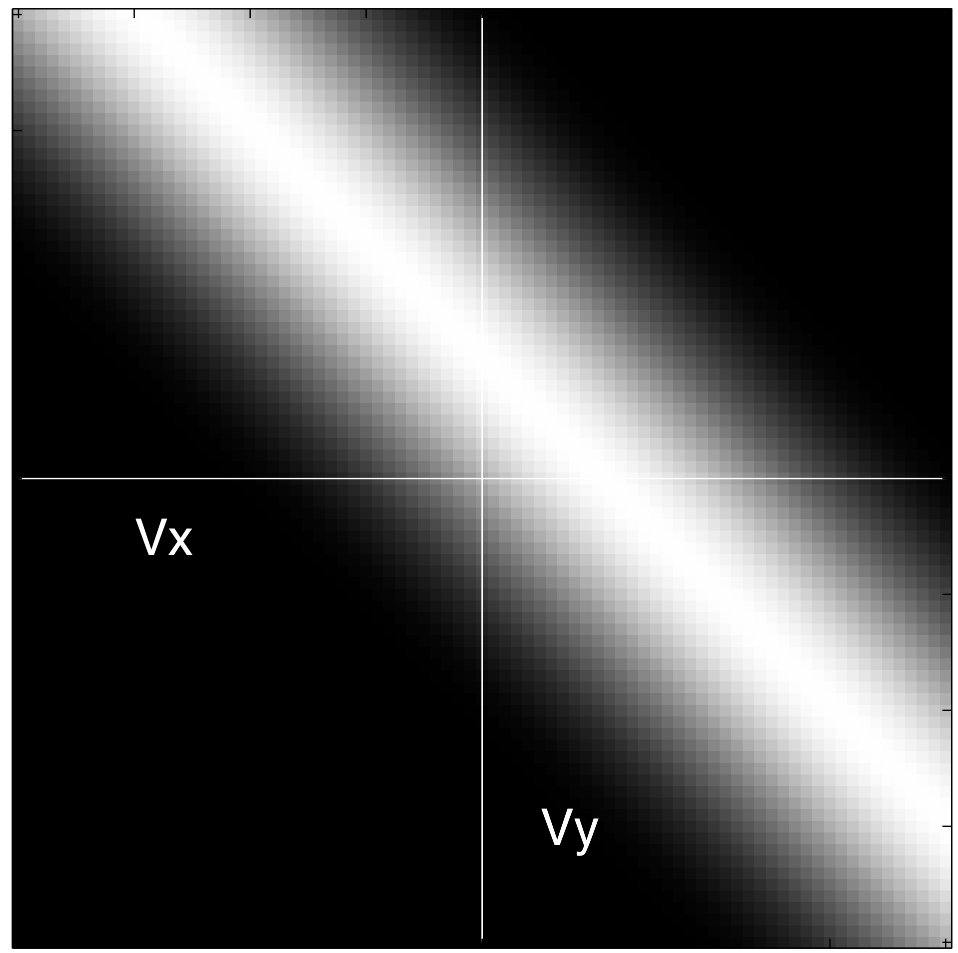


Likelihood for different durations



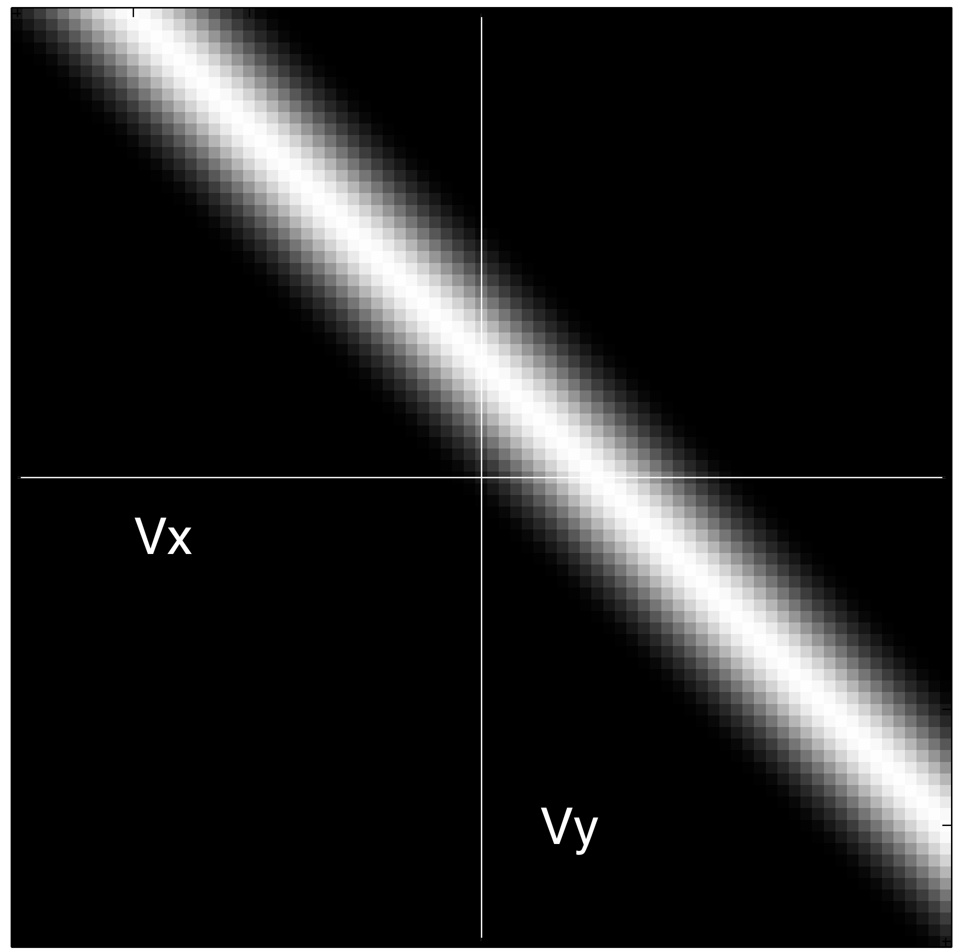
t=1

Likelihood for different durations



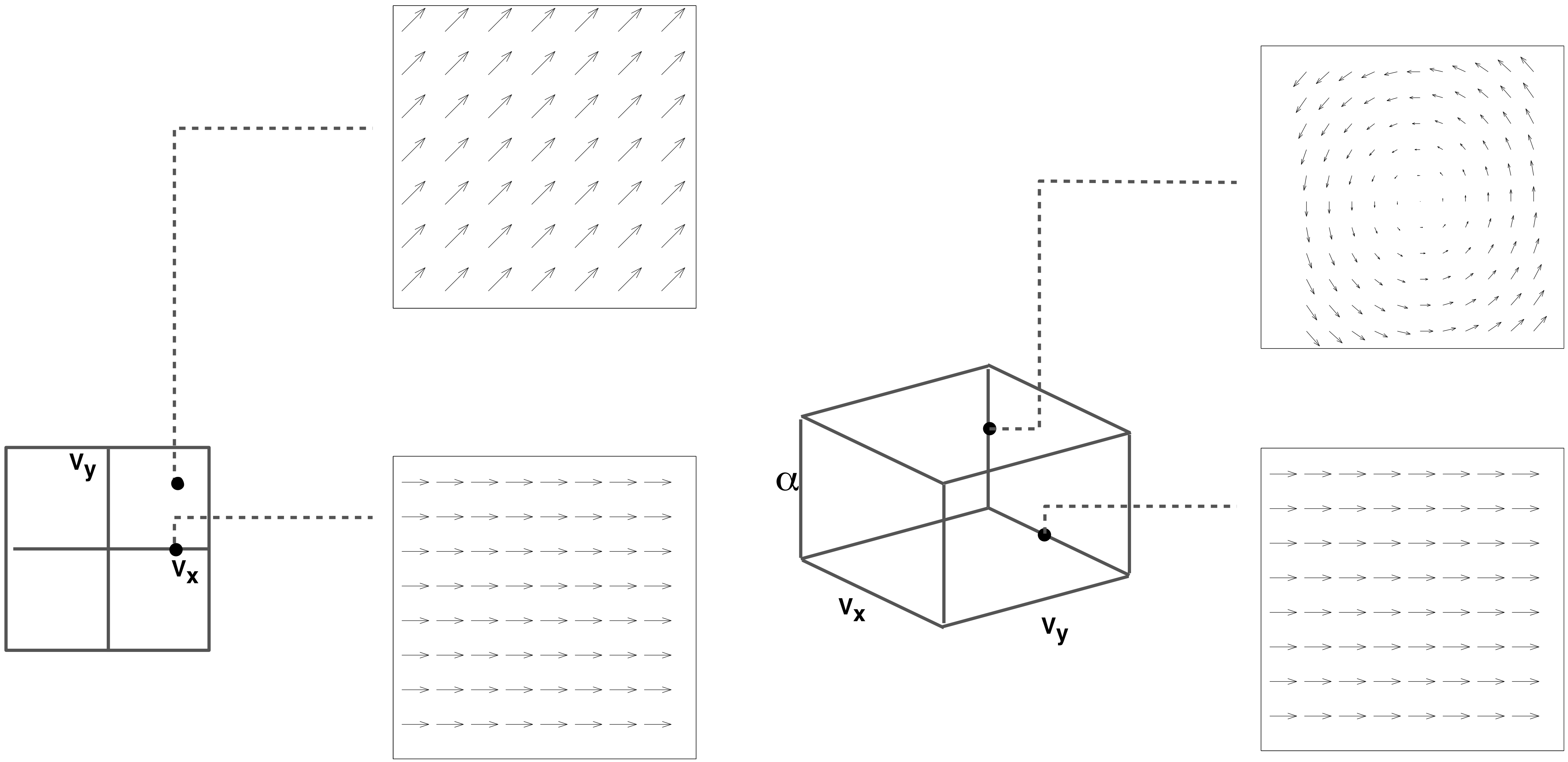
t=2

Likelihood for different durations



t=4

Limitations of the velocity space representation



- A 2D velocity field can't represent surface rotation.
- Rotation is possible by adding a third dimension, but need more so use standard flow field.

Bayesian representation of local velocity fields

- For a Bayesian model, need to define:
  - likelihood
  - prior
- Each point in the flow field integrates 25 local velocity estimates on a 5 × 5 grid

$$
\begin{aligned}
v_x(x, y) &= \sum_{i=1}^{25} \theta_i G(x - x_i, y - y_i) \\
v_y(x, y) &= \sum_{i=26}^{50} \theta_i G(x - x_i, y - y_i)
\end{aligned}
$$

(Some) Details of the model likelihood

- Let vector *r* represent location (x, y). The velocity field at point (x, y) is represented by the superposition of basis functions Ψ:

$$v(r) = \sum_i \theta_i \Psi_i(r) = \Psi(r)\theta$$

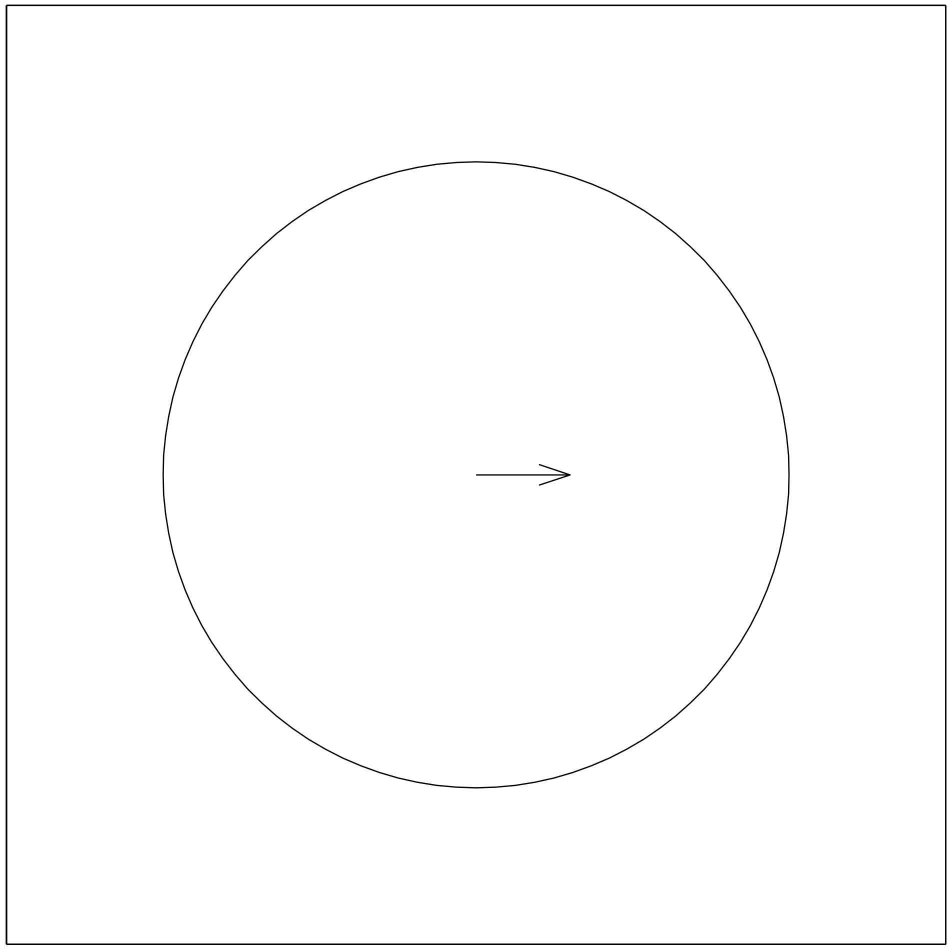
- Paper used 25 horizontal and 25 vertical basis functions, equally spaced on the image. The likelihood can be written in Gaussian form:

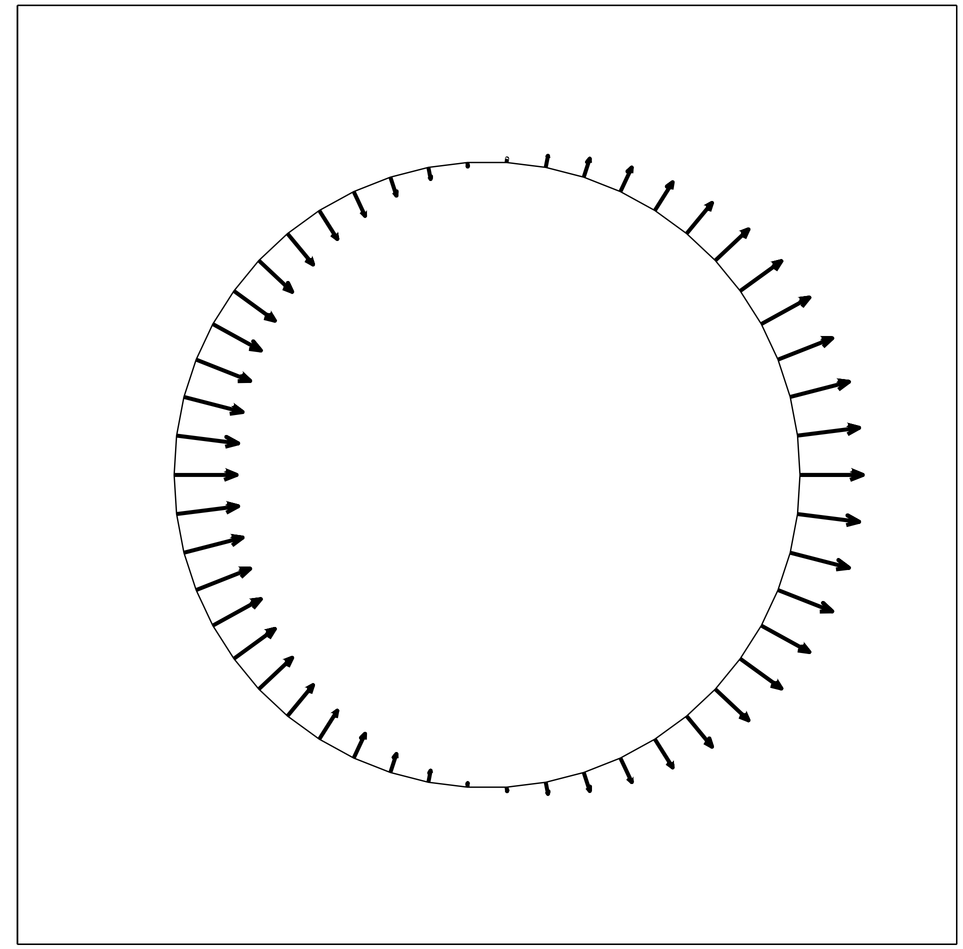
$$L_r(\theta) \propto \exp\left[-\frac{1}{2\sigma}(\Psi\theta - \mu)^T \Sigma^{-1} (\Psi\theta - \mu)\right]$$

  where Ψ, θ, μ, and σ are defined for location *r*.

- can implement computation of local likelihood by summing and squaring of outputs of spatiotemporal filters
- global likelihood is the product

Considerations of the prior





- Using prior that just prefers slow motion leads to non-rigid velocity fields (each local motion vector moves in direction of its normal).
- So, need to use a prior that is slow and smooth
  - penalize the speed of velocities
  - penalize the magnitude of the velocity derivatives, i.e. rigid translations have high probability

Details of the prior

- The prior is defined:

$$P(V) \propto \exp^{-J(V)}$$

where

$$J(V) = \sum_{xy} ||Dv(x,y)||^2$$

where *D* is an operator that penalizes strong derivatives:

$$Dv = \sum_{n=0}^{\infty} a_n \frac{\partial^n}{\partial x} v$$

- an weighted for convergence
- $n = 0$ penalizes fast flow fields ($a_0 = 1$)
- $n = 1$ penalizes non-smooth velocities ($a_1 = \lambda^2/2$)
- $n > 1$ has minimal effect ($a_n = 0$ for $n > 1$)

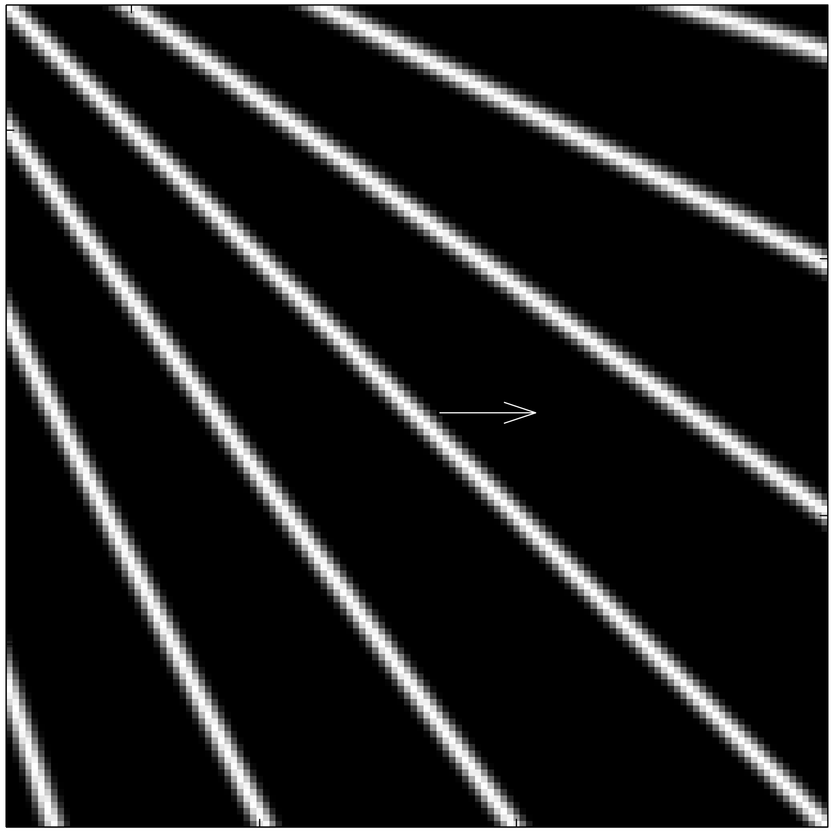
Posterior

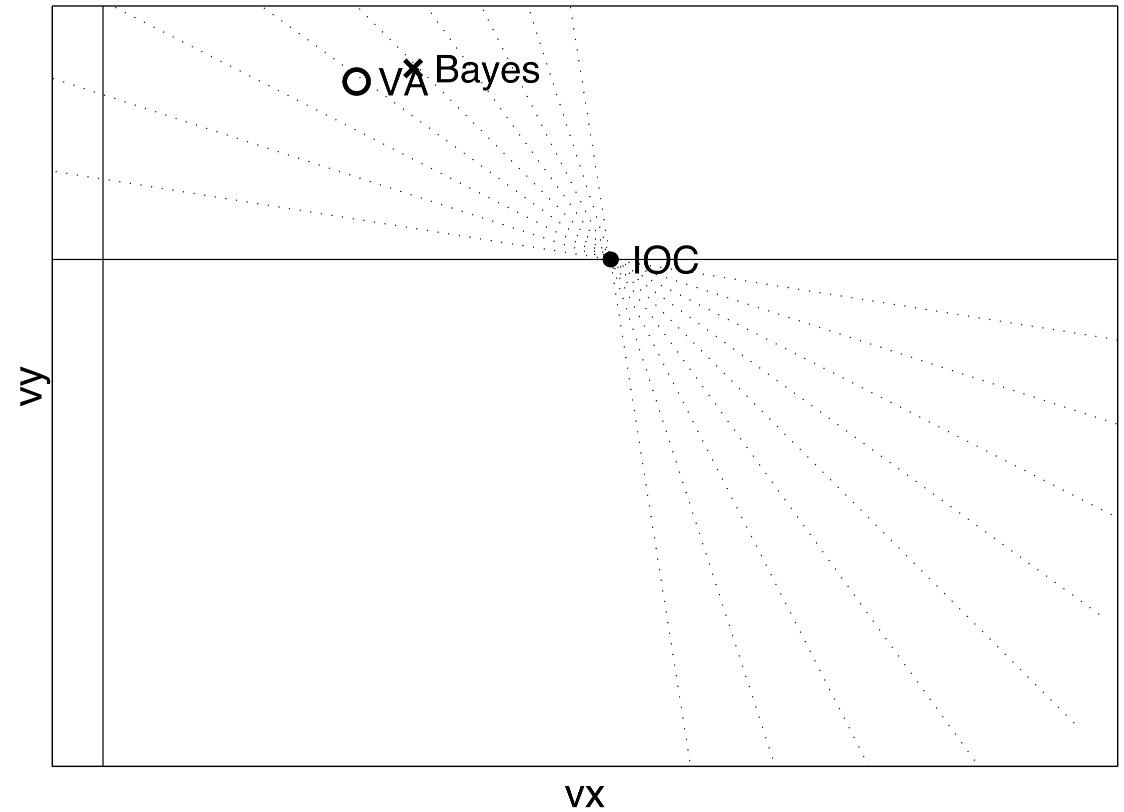
- Putting it all together, the posterior for the motion flow field is

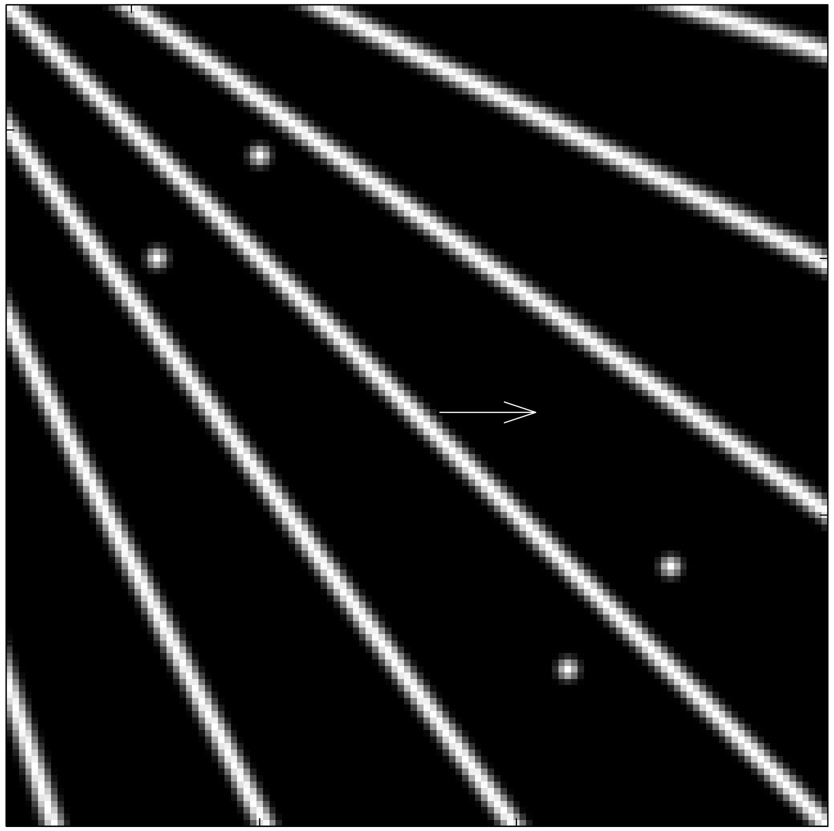
$$P(\theta|I) \propto P(\theta)P(I|\theta)$$

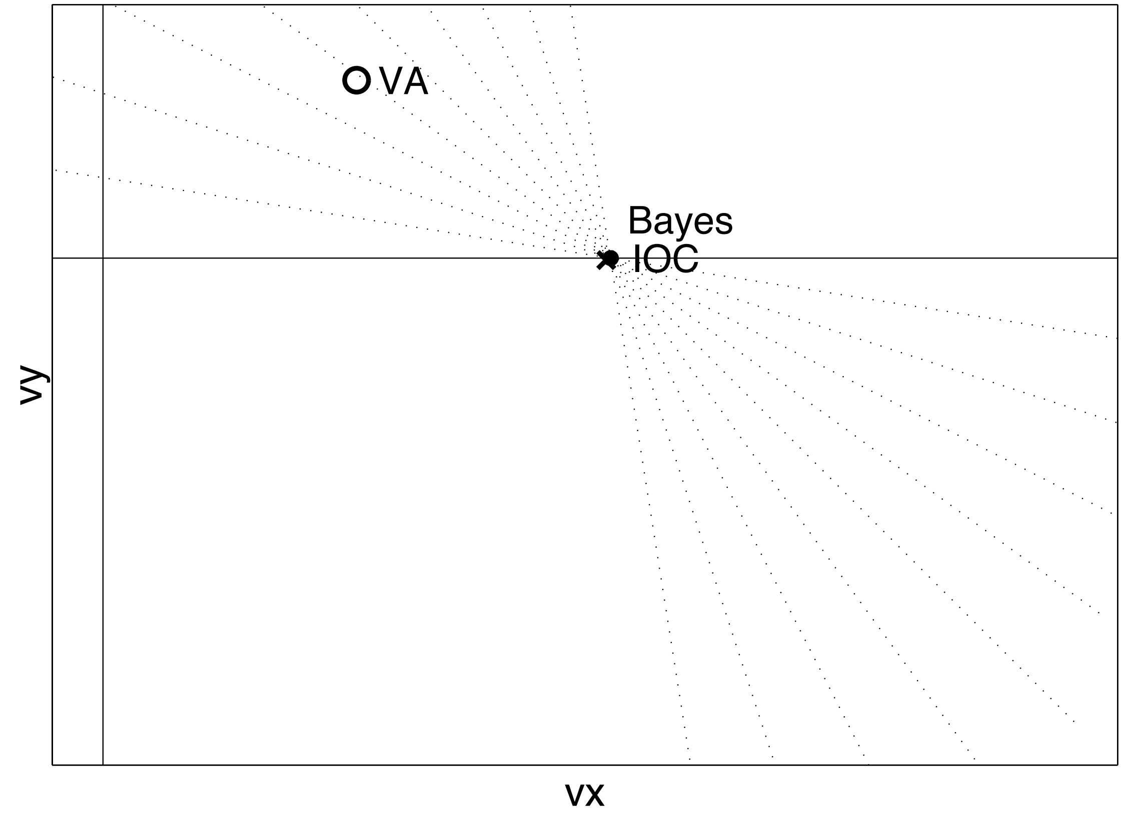
- Posterior has Gaussian form ⇒ most probable θ obtained by solving linear eqn.
- Applying the model to real images (5 frames, 128x128 grey-scale pixels)
  - local velocity likelihoods are calculated at every image location using $5 \times 5 \times 5$ spatiotemporal filters
  - combine local estimates to calculate most probable velocity field using prior that favors slow and smooth motions

Examples

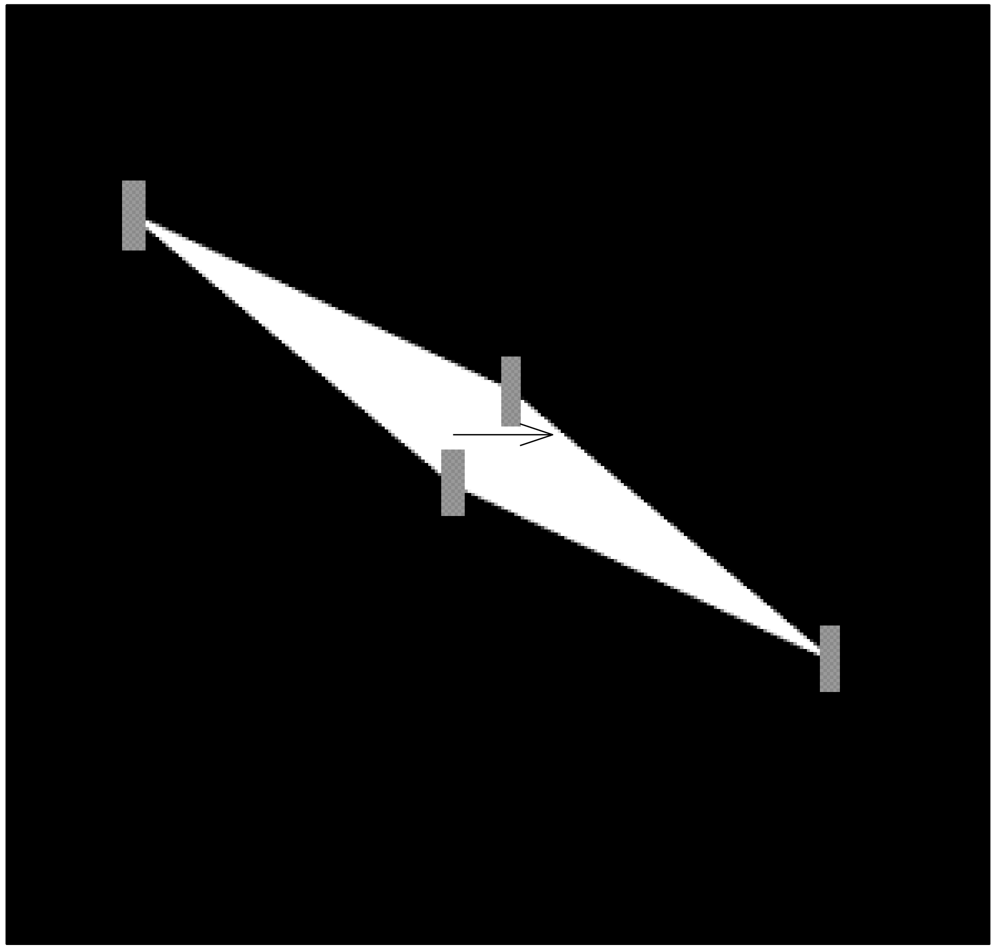


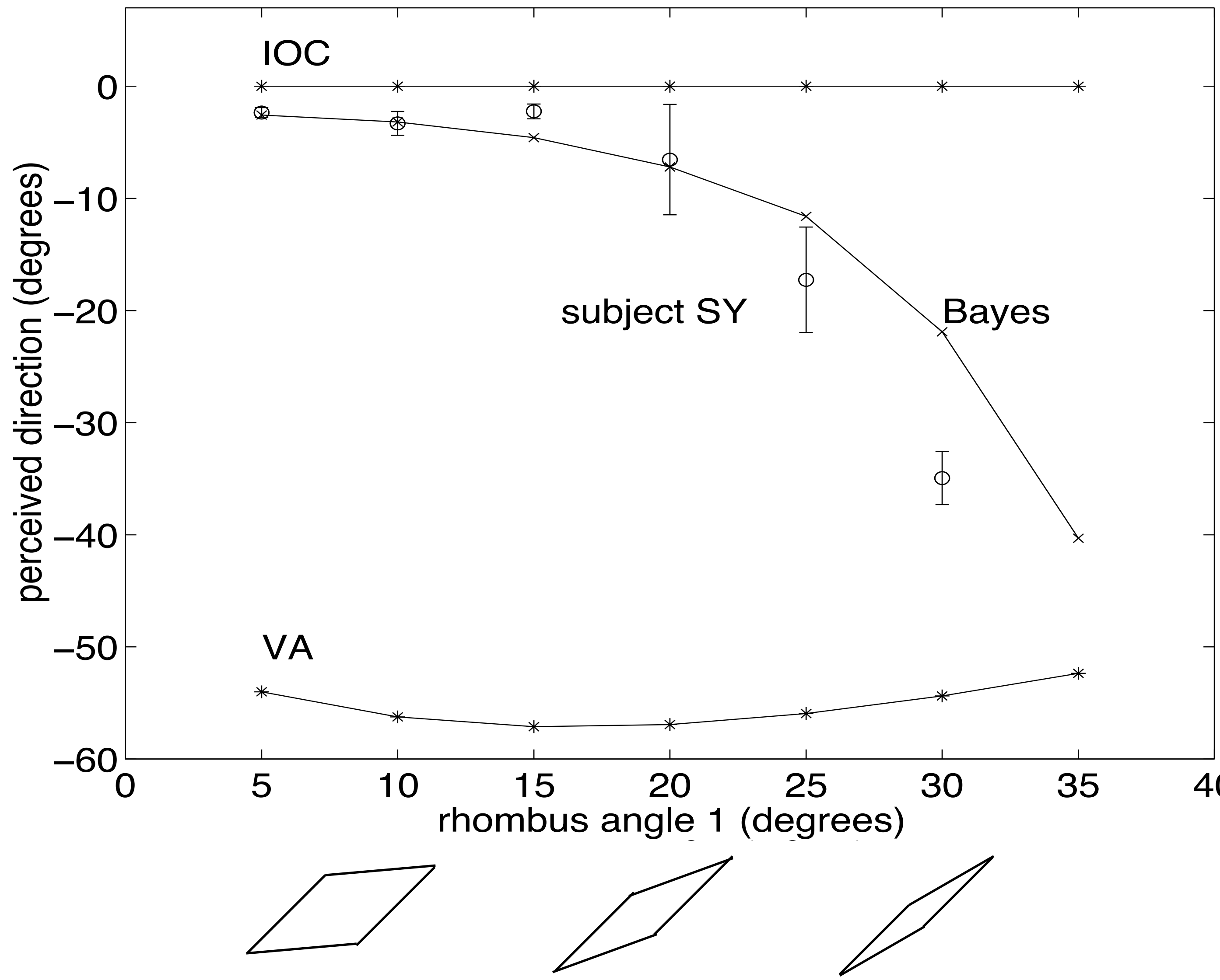




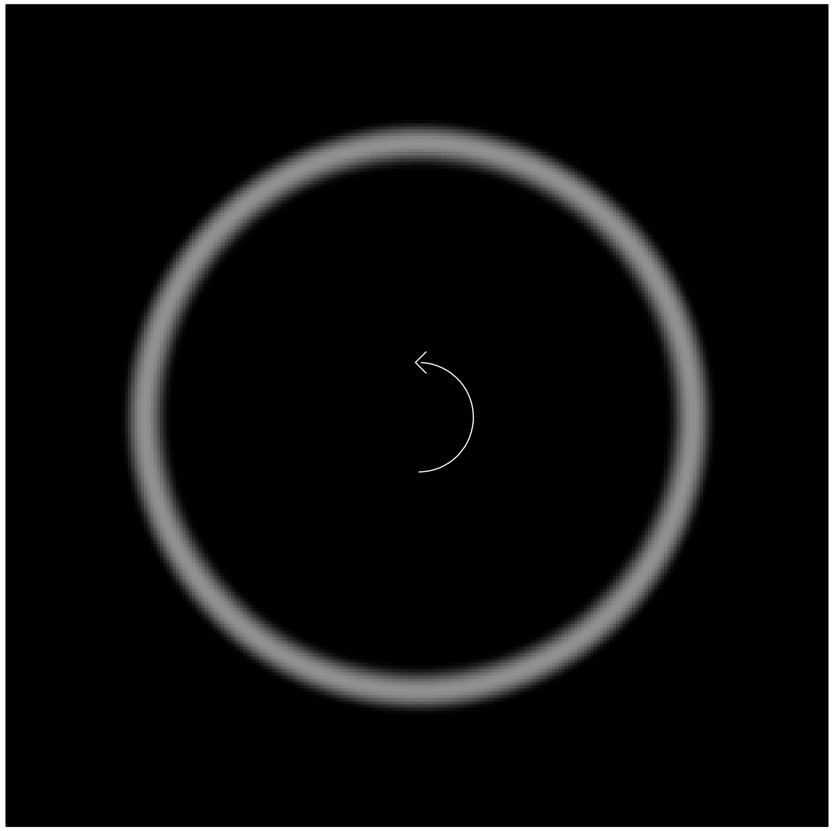


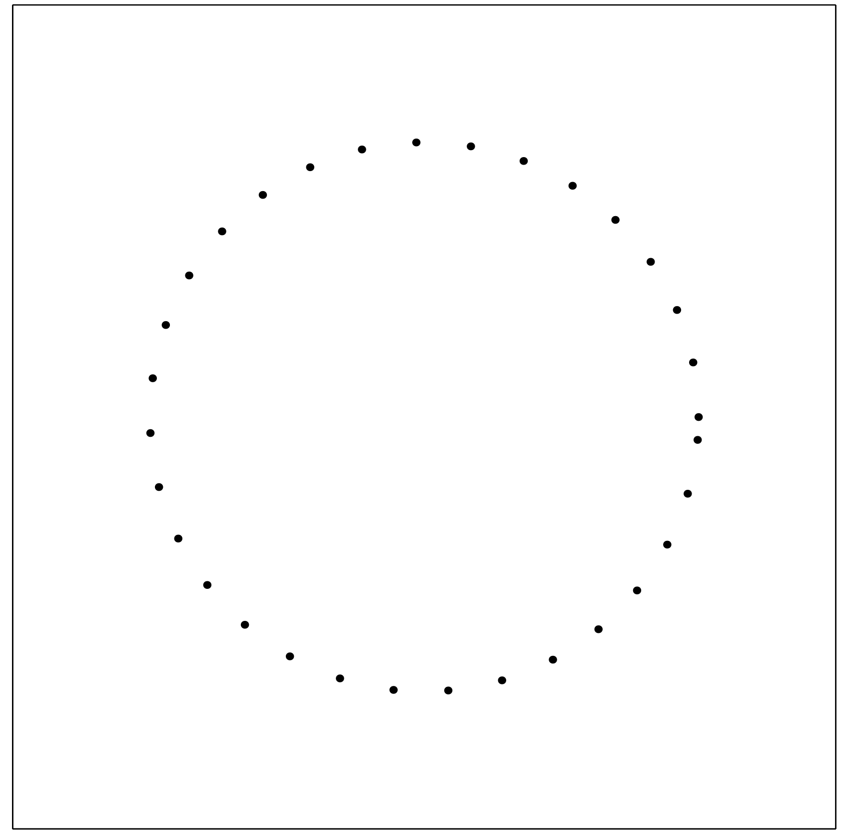
Model explains human perception

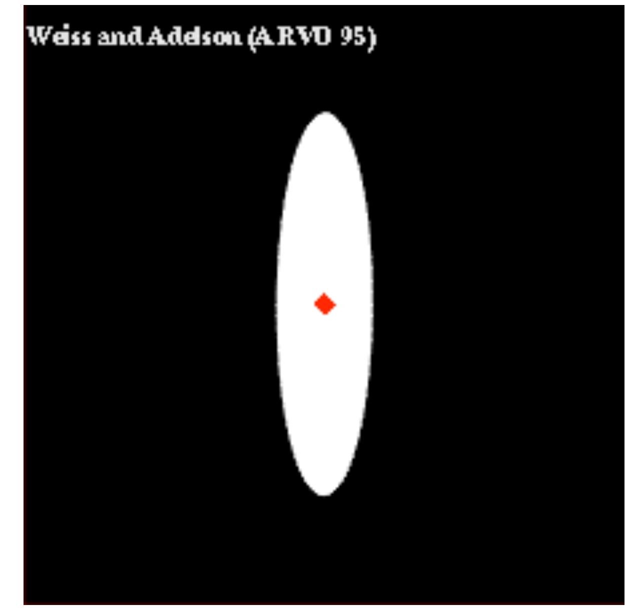


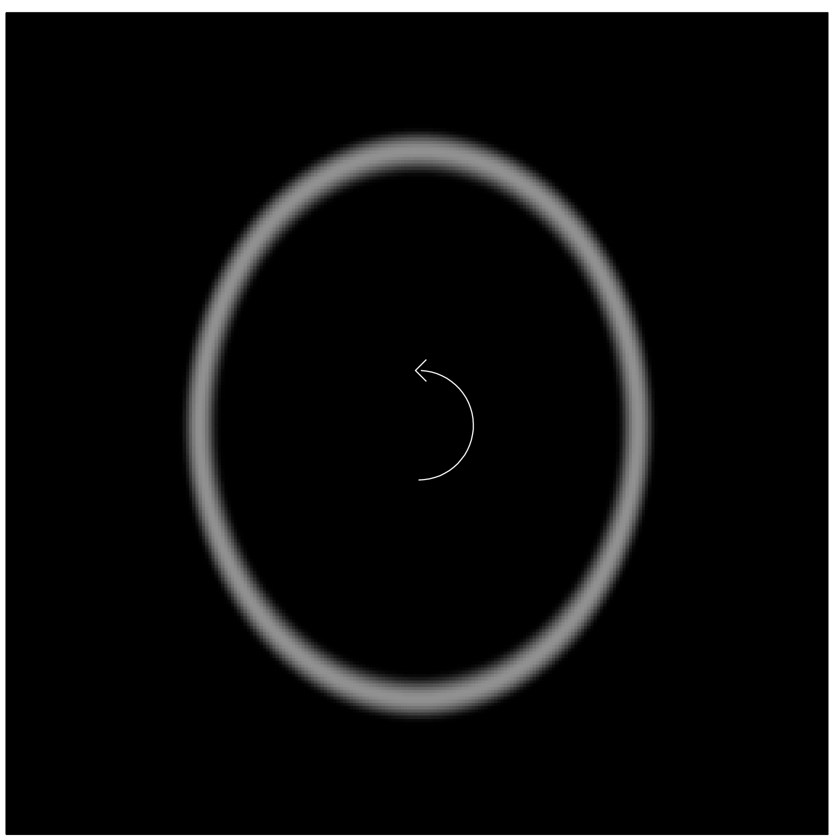


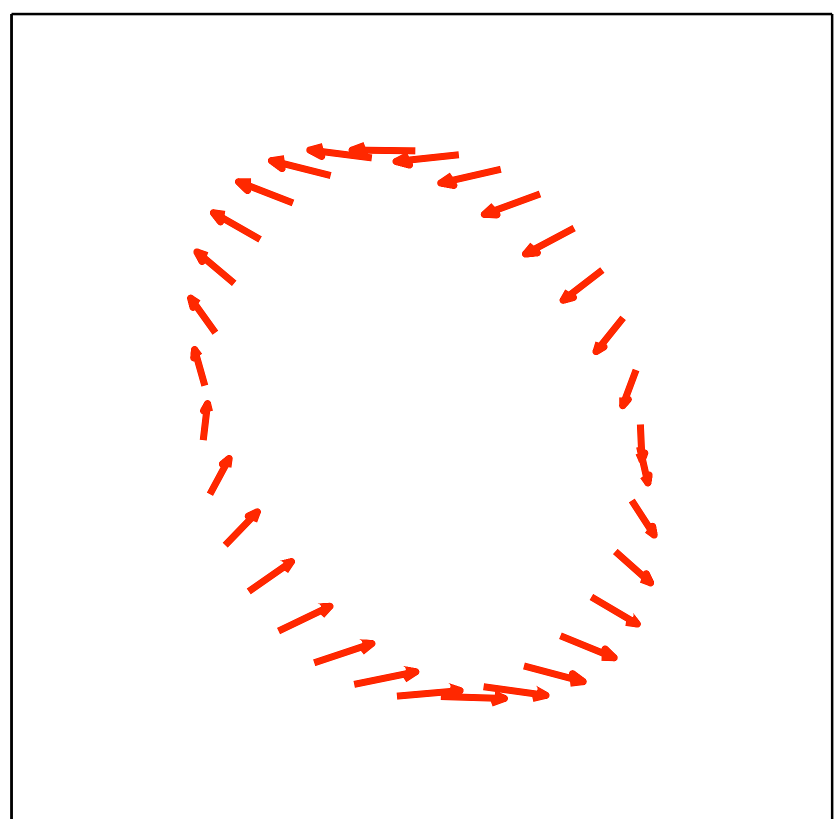
Examples

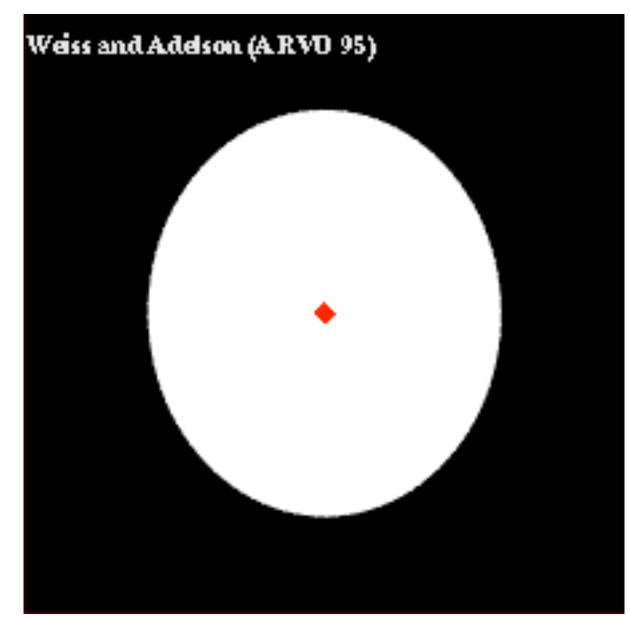




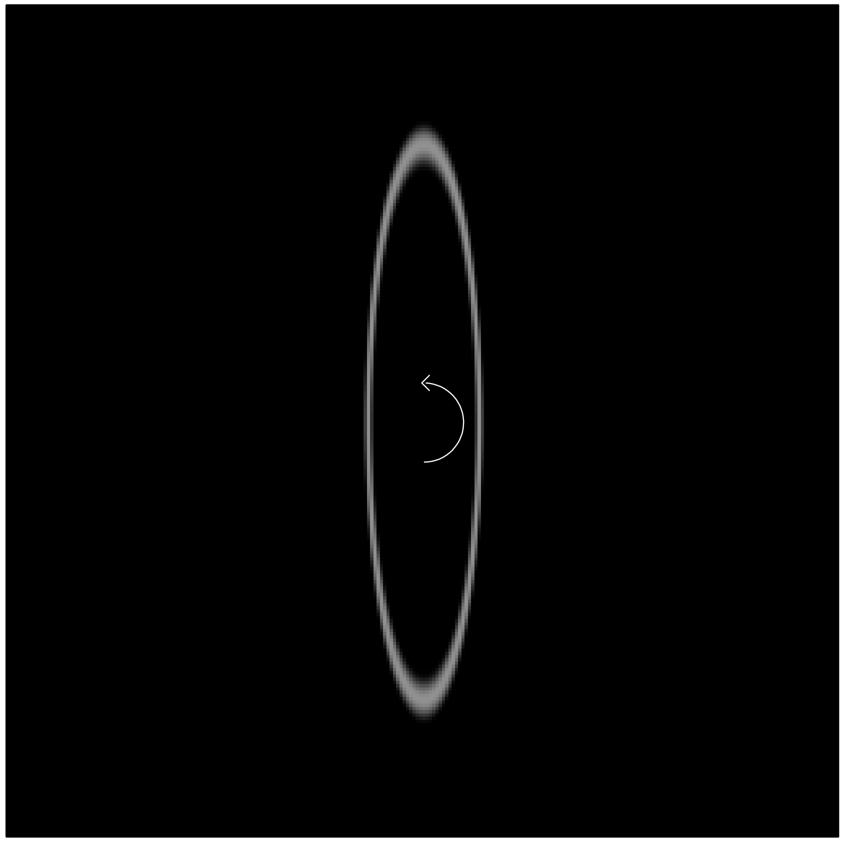


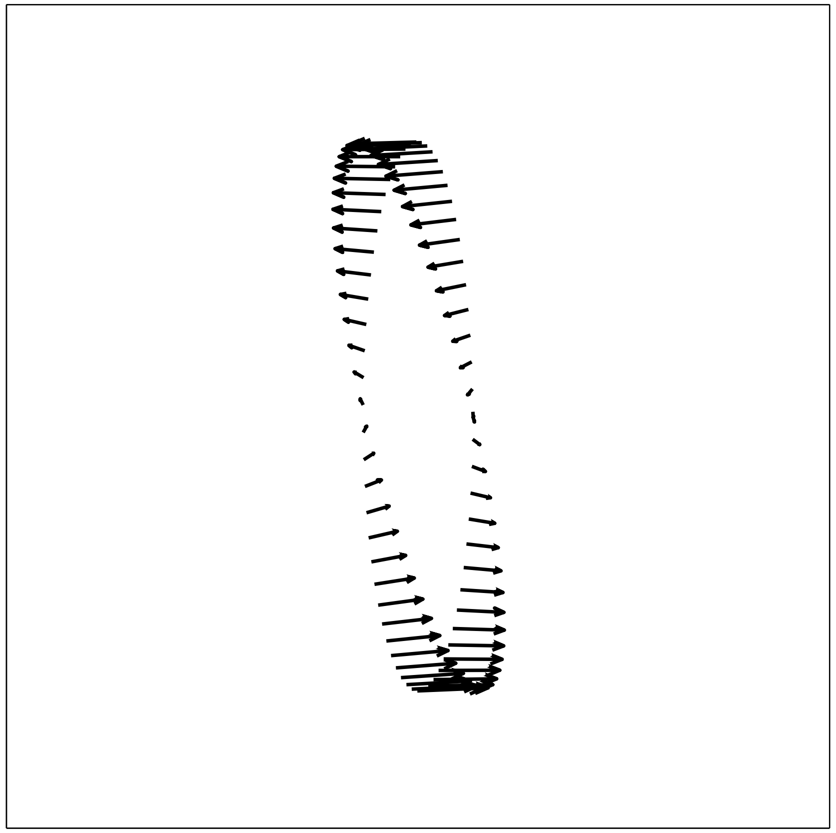


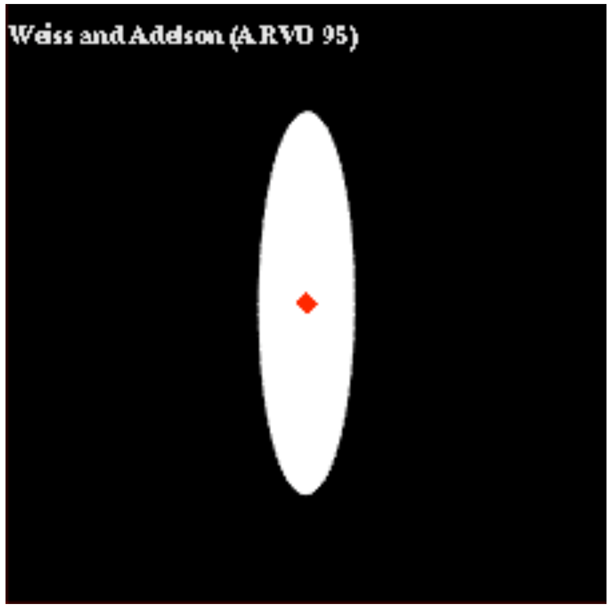


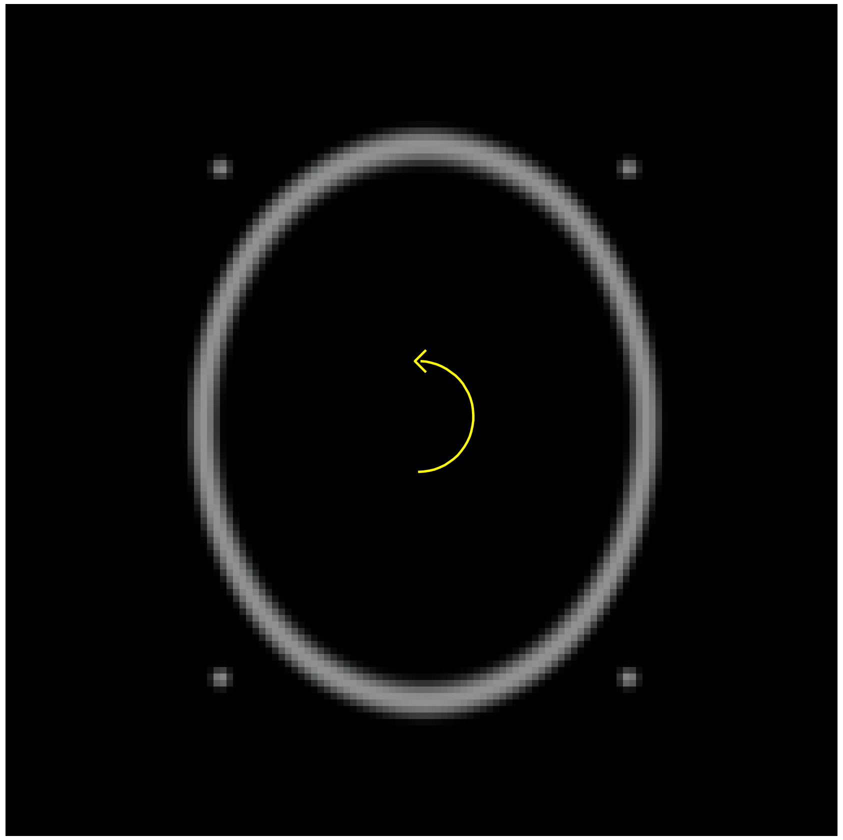


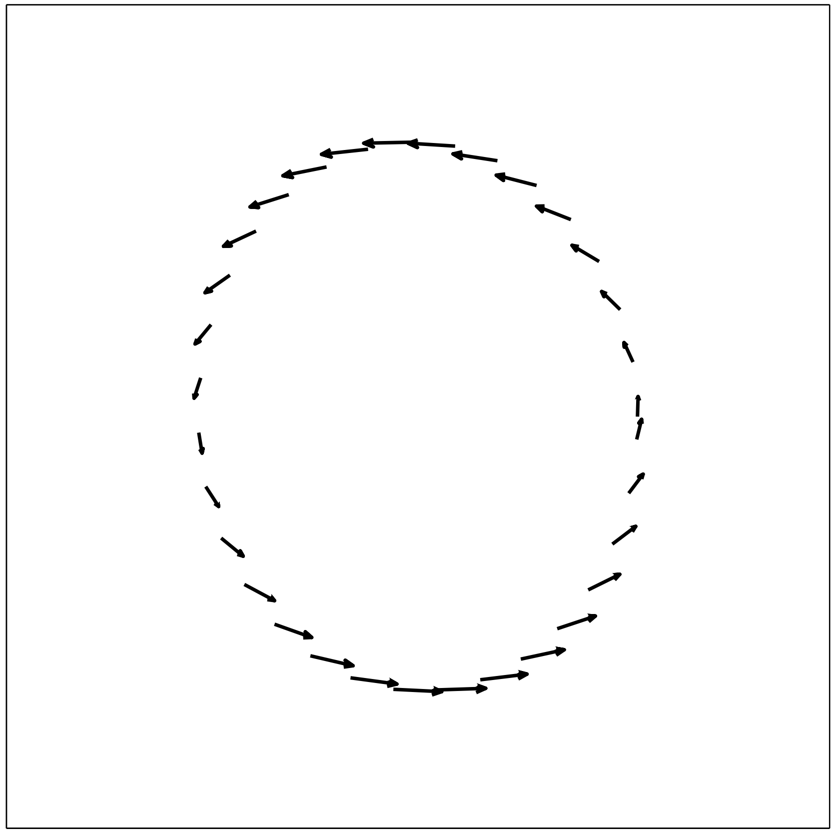
Examples

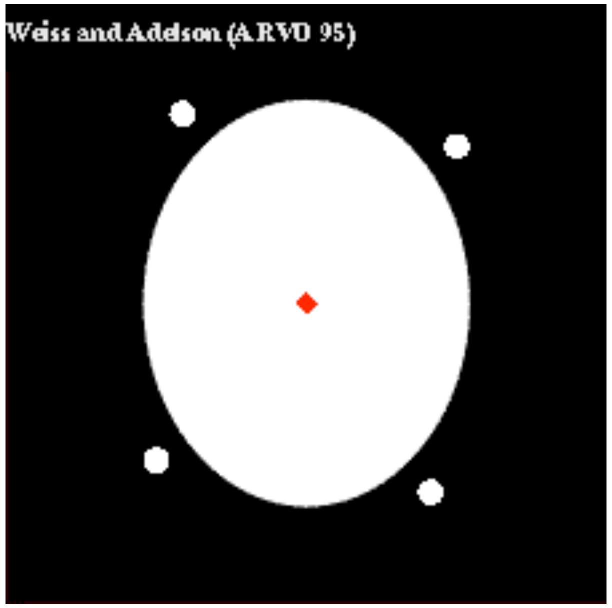




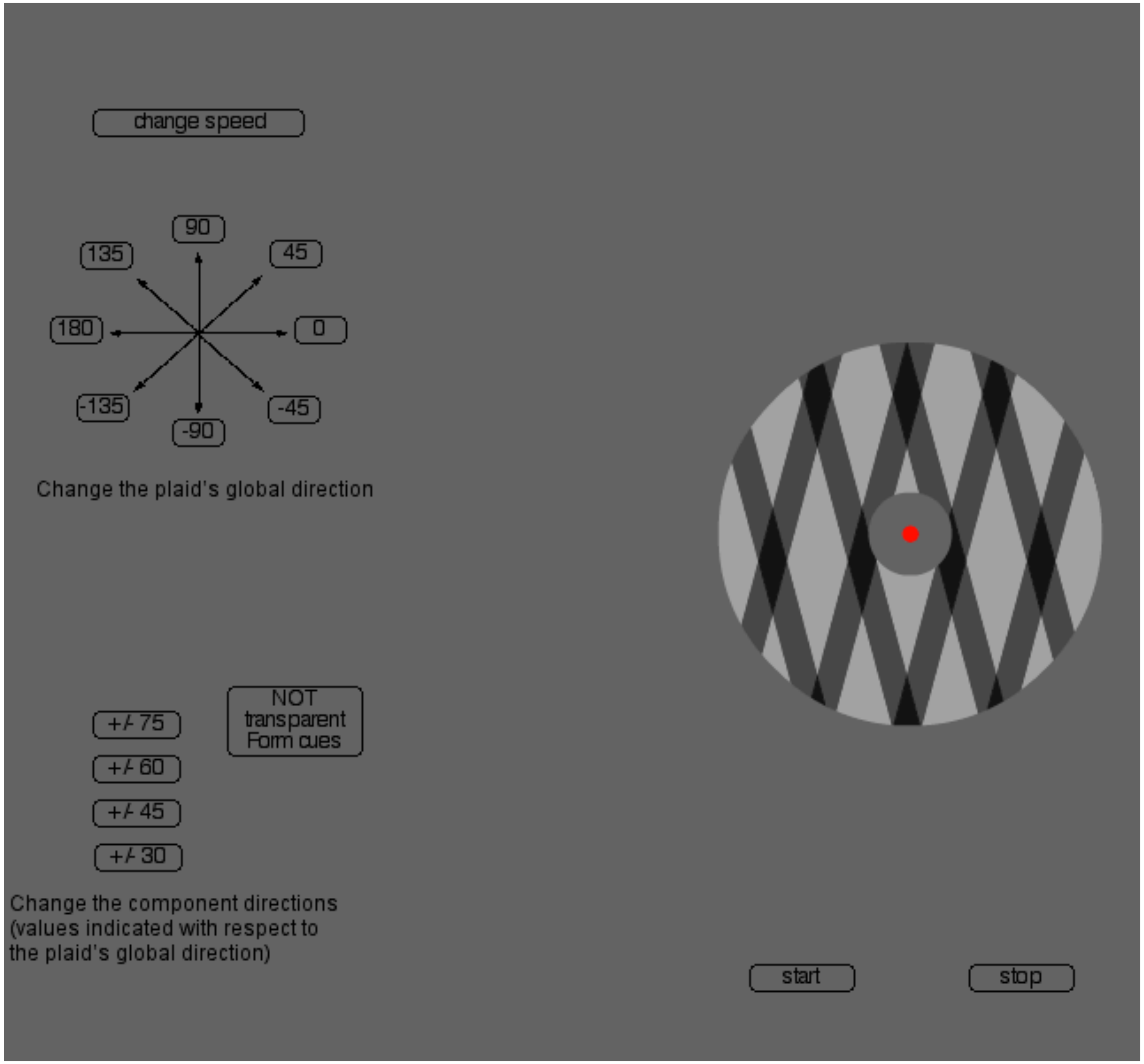








Does this give us any new insights?



http://www.cns.nyu.edu/~hupe/plaid_demo/demo_plaids.html

Summary

- Clean incorporation of asssumptions of motion cues (regularization)
- Accounts for a wide range of motion perceptual phenomena
- Simpler model: No need for specific mechanisms to handle edges, corners, plaids, blobs, etc.
- Makes different predictions from both IOC, VA, and smoothness-based methods
- Handles contrast and integration time
- Bayes estimate almost always corresponds to perceived velocity