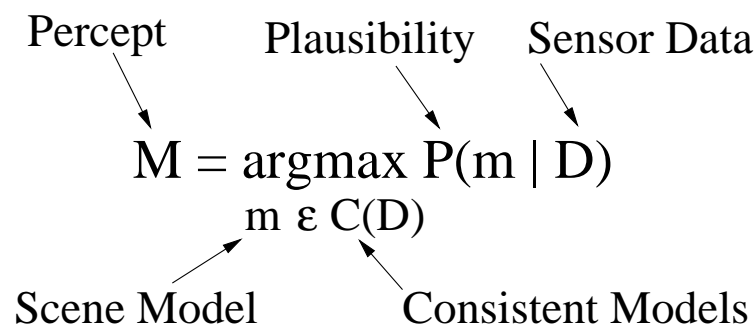


Perceivers

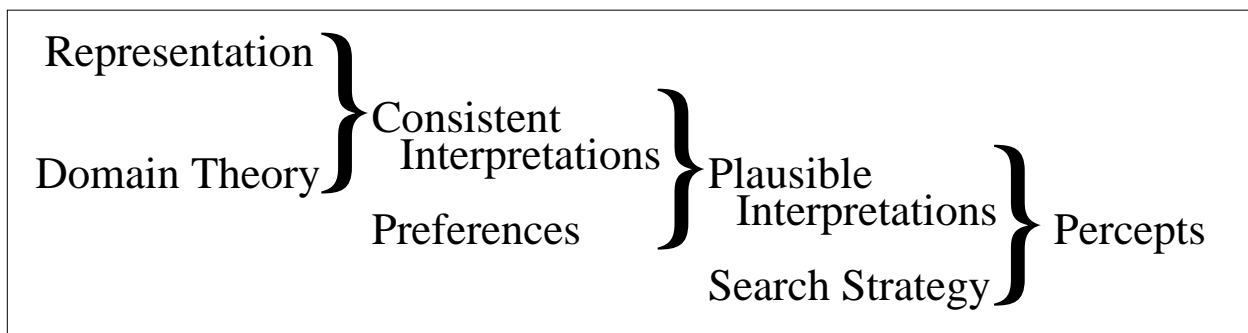
Perception: The process of finding plausible interpretations for data.

The Computational Theory of Perception:



We need to specify

- **Representation** – how interpretations are represented.
- **Domain Theory** – what constitutes a consistent interpretation.
- **Preferences** – when one interpretation should be preferred over another.
- **Search Strategy** – how to search for preferred interpretations.



Examples

Assignment 1: Separating diffuse and specular reflection.

- **Models:** Lighting and reflectance model.
- **Preferences:** Simple noise model.
- **Search:** By hand.

Assignment 2: Counting cells and buds in microscope images.

- **Models:** Smooth background, circular boundaries.
- **Preferences:** Simple background model, parsimonious circle fitting.
- **Search:** Robust estimation, RANSAC, greedy model comparison.

Assignment 3: Human eyes.

- **Models:** Eigen eyes, skin colour.
- **Preferences:** Mainly consistency. A pair of eyes in a skin-coloured blob with a known eye-to-eye separation.
- **Search:** Robust estimation, eigen-detection.

Assignment 4: Tracking pigeon parts.

- **Models:** Distribution of poses of pigeon parts.
- **Preferences:** Posterior probability (filtering distribution).
- **Search:** Particle filter.

Blocks-World Study: Motivation

The simple blocks-world domain serves to illustrate the basic components of a perceiver:

- Scene representation.
- Domain theory.
- Plausibility theory.
- Search for plausible models.

The domain also illustrates two basic problems in image interpretation:

- Providing a definition of an appropriate plausibility ordering on possible interpretations.
- Controlling the search for interpretations to relatively few plausible choices.

Image Line Data

Khoros line-finder results on blocks-world images:

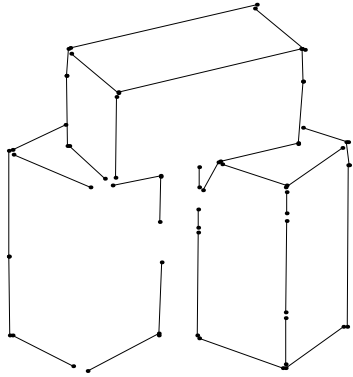


Image "9": 34 Segments

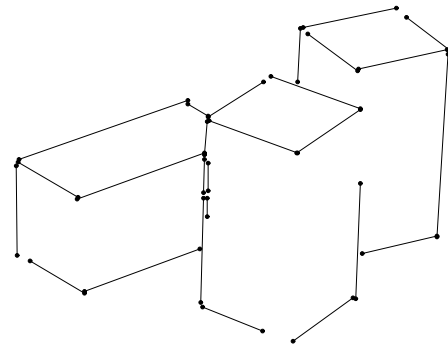


Image "5": 26 Segments

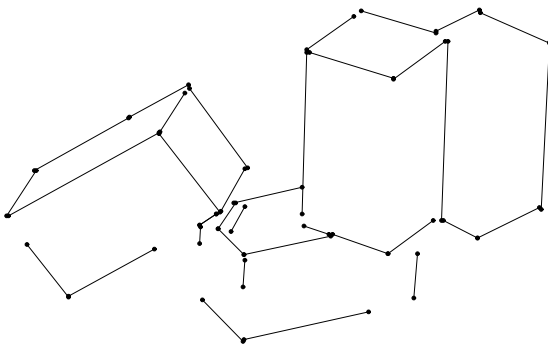


Image "E": 34 Segments

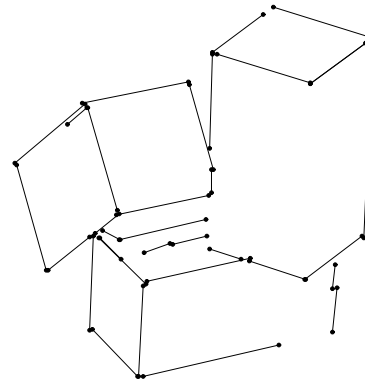


Image "D": 34 Segments

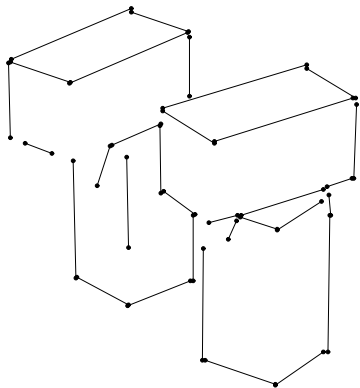


Image "7": 32 Segments

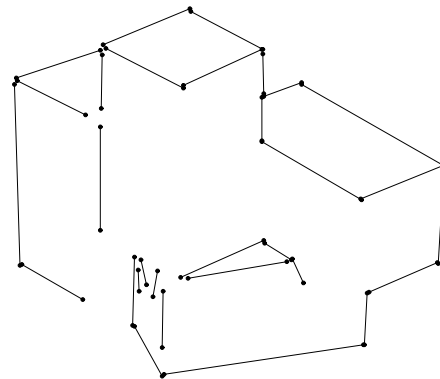
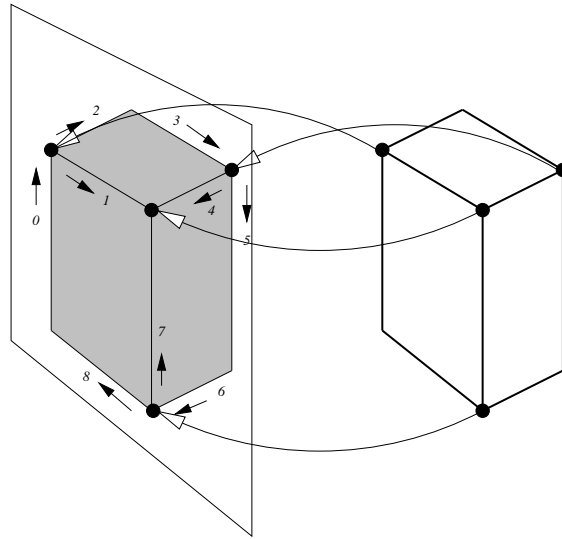


Image "4": 30 Segments

Representation: Card-Blocks World



Card-Block Model

Scene

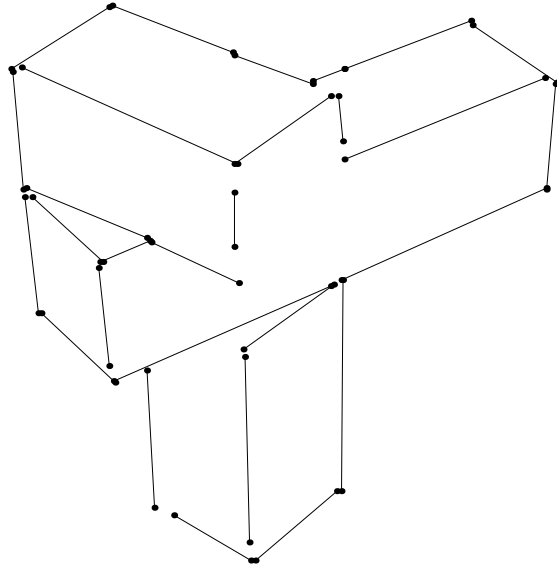
We consider scene models consisting of:

- Card-blocks arranged in depth layers.
- Sticks (isolated line segments).

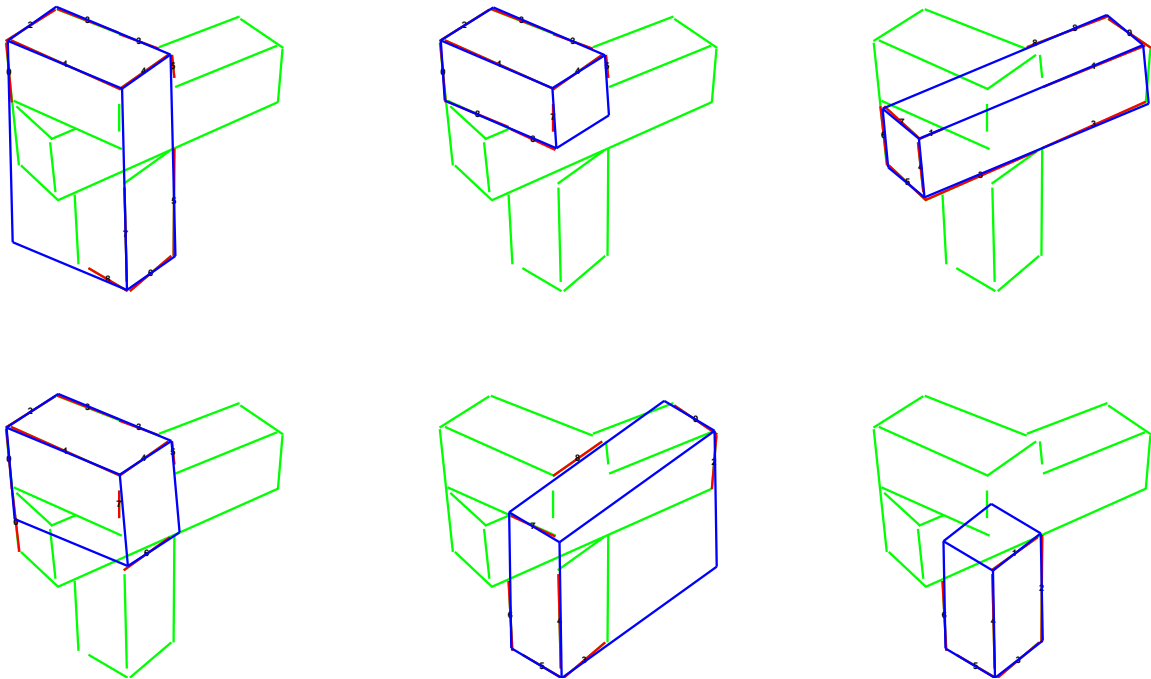
The interior of a card-block is taken to be opaque, and therefore it can occlude other objects.

Image interpretations must explain all image edges in terms of card-blocks or sticks.

Example: Multiple Candidates

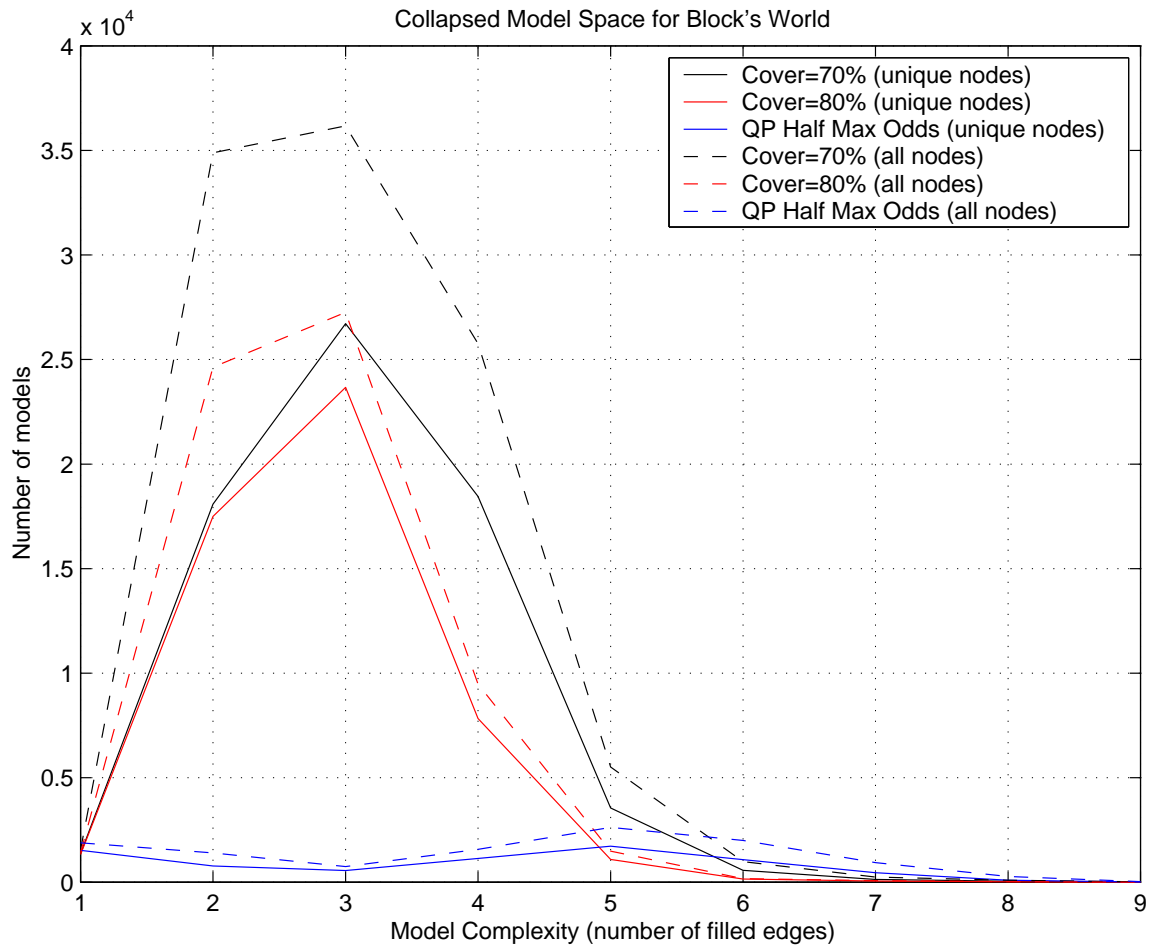


Some plausible individual blocks, amongst thousands.

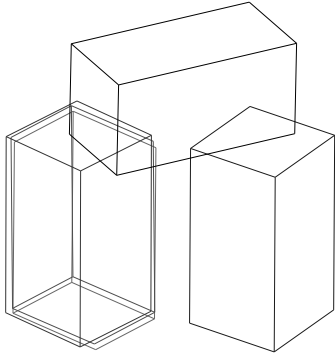


Number of Individual Blocks

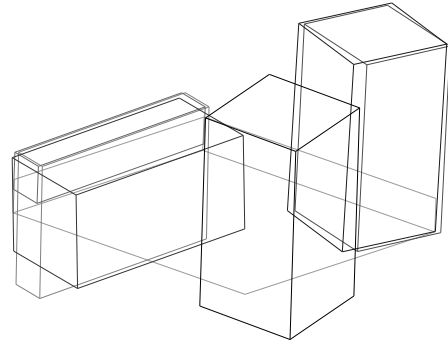
Just how big is that search space?



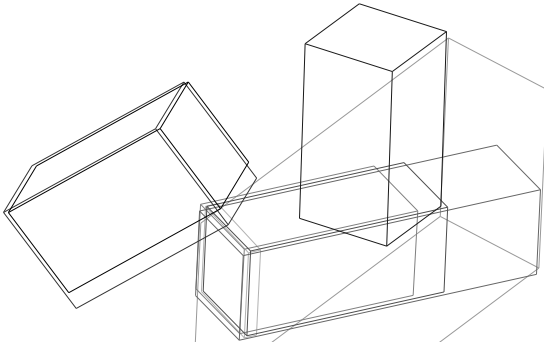
Plausible Individual Blocks



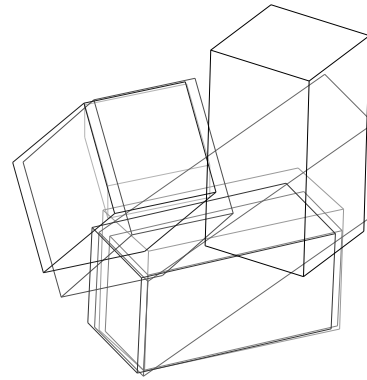
Im9: 5 blocks



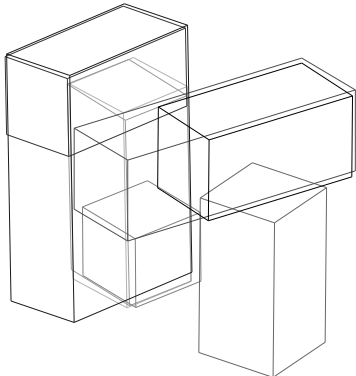
Im5: 7 blocks



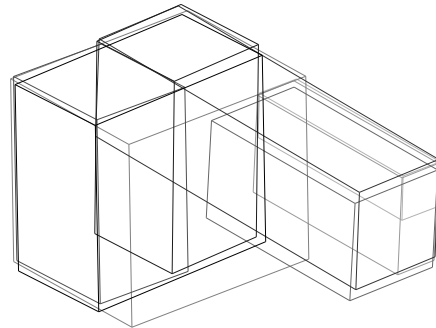
ImE: Best 8 of 12 blocks



ImD: Best 8 of 11 blocks

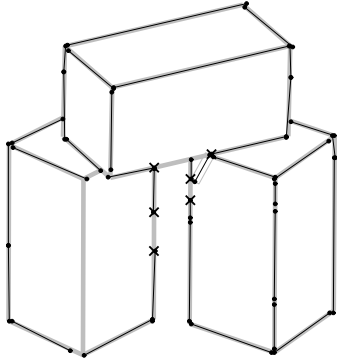


Im7: Best 8 of 10 blocks

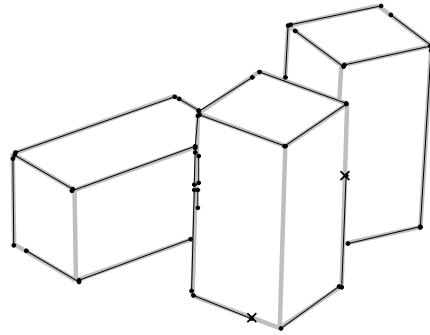


Im4: Best 8 of 18 blocks

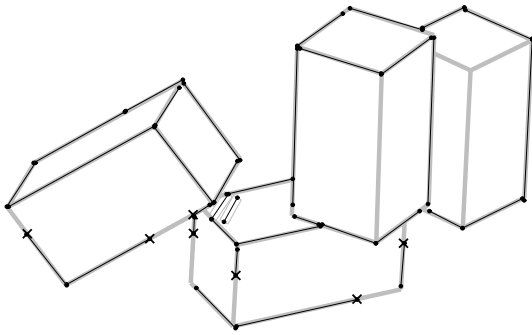
Plausible Interpretations



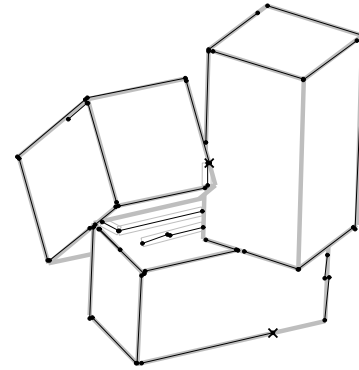
Im9: Score: $\epsilon^{34} \delta^5$
(next best $\epsilon^{42} \delta^7$)



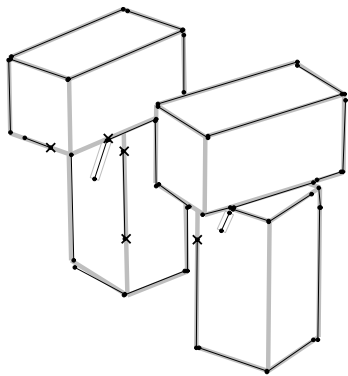
Im5: Score $\epsilon^{26} \delta^5$
(next best $\epsilon^{35} \delta^7$)



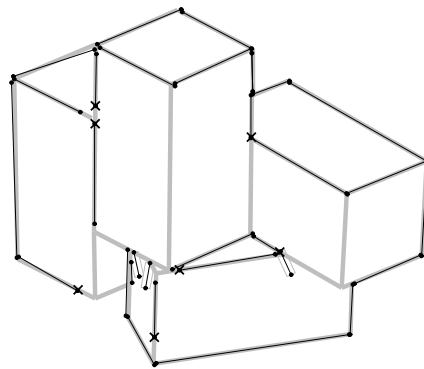
ImE: Score $\epsilon^{47} \delta^{10}$
(1 other within ϵ^2)



ImD: Score: $\epsilon^{42} \delta^7$
(1 other within ϵ^2)



Im7: Score: $\epsilon^{45} \delta^8$
(1 other within ϵ^2)



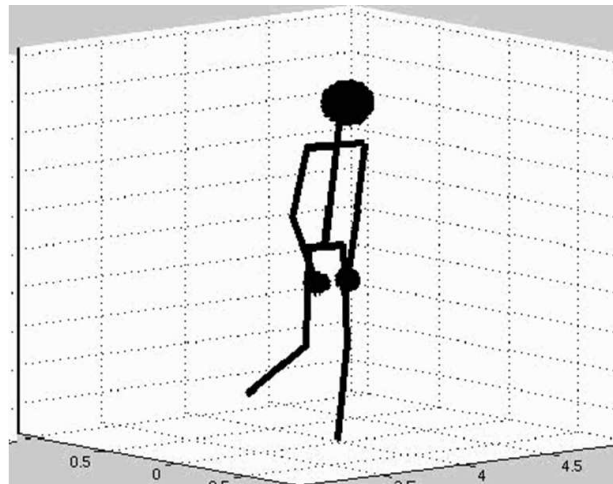
Im4: Score: $\epsilon^{51} \delta^9$
(7 others within ϵ^2)

Beyond Blocks

A much more complicated example is the perception of human gaits.



3D Model



But the big issues are the same. What sort of model? What's allowed in the model space? Given image data, which specific models are preferred? How can we find these preferred models efficiently?

Perceivers: Motion Understanding

Scene Dynamics slides go here

The Great Beyond

A big remaining challenge is to move beyond such simple models, to more powerful generic models, such as the ones you might use to describe a random postcard in some detail (say in a paragraph or two of text and a sketch from a different viewpoint).



Surprisingly challenging due to:

- the ambiguity of local image data,
- the desire for rich, general scene models,
- a tractable definition of plausibility is required,
- the size of the search space containing plausible models.

References:

1. E. H. Adelson and A.P. Pentland, The Perception of Shading and Reflectance MIT Media Lab Vismod TR#140 (1990) (see also their chapter in Perception as Bayesian Inference, Eds. W. Richards and D. Knill, Camb. Univ. Press (1996)).
2. A.D. Jepson and R. Mann, Qualitative probabilities for image interpretation, Proc. IEEE Int. Conf. on Computer Vision, Kerkyra, Greece, 1999, pp. 1123-1130.
3. A. Jepson and W. Richards, What is a percept?, Research in Biological and Computational Vision, Department of Computer Science, Univ. of Toronto, RBCV-TR-93-43, Apr. 1993 (see also MIT Cog. Sci. Memo 43, 1991).
4. R. Mann, A. Jepson, and J.M. Siskind, Computational perception of scene dynamics, Comp. Vision and Image Understanding, Vol. 65, No. 2 (Feb. 1997), pp.113-128.
5. P. Sinha and E. Adelson, Recovering reflectance and illumination in a world of painted polyhedra, MIT Media Lab TR#199, 1992 (see also ICCV '93).