Catching heuristics are optimal control policies

Boris Belousov, Gerhard Neumann, Constantin A. Rothkopf, Jan Peters
email: {belousov, neumann, peters}@ias.tu-darmstadt.de, rothkopf@psychologie.tu-darmstadt.de

Motivation

There are two seemingly contradictory theories of ball interception:

- humans predict the ball trajectory to optimally plan future actions,
- humans employ heuristics to reactively choose actions based on visual feedback.

We show that interception strategies appearing to be heuristics can be understood as computational solutions to the optimal control problem faced by a ball-catching agent acting under uncertainty.

Catching heuristics

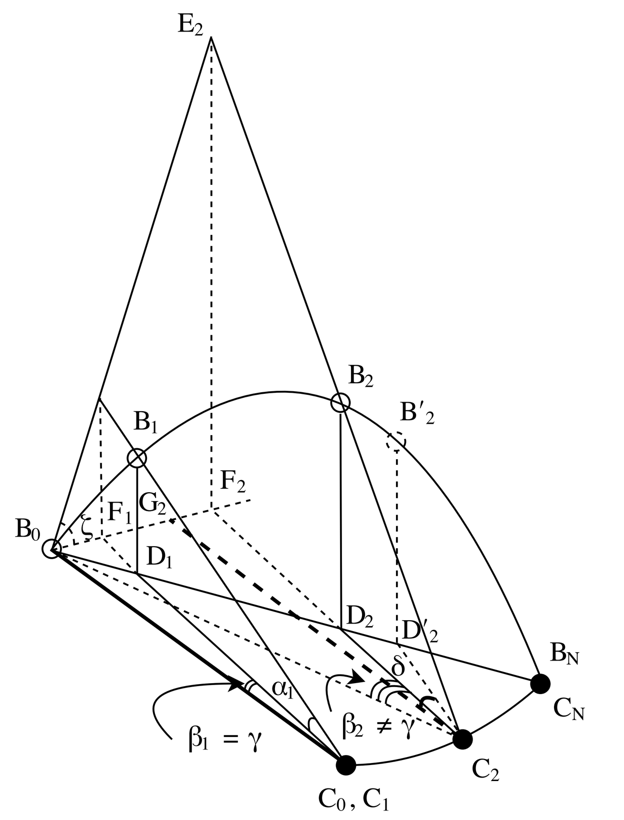

Figure 1: Prominent catching heuristics in one figure.

A number of heuristics have been proposed to explain how humans catch balls. Figure 1 shows four well-supported by experiments heuristics: **optic acceleration cancellation** (OAC), **constant bearing angle** (CBA), **generalized optic acceleration cancellation** (GOAC), and **linear optical trajectory** (LOT). OAC and CBA together form Chapman's theory.

In Figure 1, the ball B follows a parabolic trajectory $B_{0:N}$ while the agent C follows $C_{0:N}$ to intercept it. Angle α is the **elevation angle**; angle γ is the **bearing angle** with respect to direction C_0B_0 (or C_2G_2, which is parallel). Due to delayed reaction, the agent starts running when the ball is already in the air.

The heuristics can be formulated as follows:

$d\tan\alpha/dt = \text{const}$	OAC
$\gamma = \text{const}$	CBA
$\delta \approx \gamma$	GOAC
$\tan\alpha/\tan\beta = \text{const}$	LOT

Ball catching as optimal control under uncertainty

System dynamics:

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{\epsilon}_{k+1}, \qquad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}),$$
$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{\delta}_k, \qquad \boldsymbol{\delta}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{x}_k)).$$

Belief state is approximated by the **normal distribution**, $\boldsymbol{b}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Future observations are assumed to coincide with their **most likely** values, $\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{\mu}_k)$, for the purpose of planning. Under these assumptions, the **extended Kalman filter** (EKF) equations result in **determenistic belief dynamics**

$$\boldsymbol{\mu}_k = \boldsymbol{f}(\boldsymbol{\mu}_{k-1}, \boldsymbol{u}_{k-1}),$$
$$\boldsymbol{\Sigma}_k = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\bar{\boldsymbol{\Sigma}}_k.$$

At every time step, the agent solves a **constrained nonlinear optimization** problem

$$\min_{\boldsymbol{u}_{0:N-1}} \; J(\boldsymbol{\mu}_{0:N}, \boldsymbol{\Sigma}_{0:N}; \boldsymbol{u}_{0:N-1})$$
$$\text{s.t.} \quad \boldsymbol{u}_k \in \mathcal{U}_{\text{feasible}}, \quad k = 0 \ldots N-1,$$
$$\boldsymbol{\mu}_k \in \mathcal{X}_{\text{feasible}}, \quad k = 0 \ldots N,$$

to obtain an **optimal sequence of controls** $\boldsymbol{u}_{0:N-1}$ minimizing the objective function J.

Detailed model of the catching agent for belief-space optimal control

Several model components are essential to faithfully describe catching behavior:

- damped dynamics $\ddot{\boldsymbol{r}}_c = \boldsymbol{F} - \lambda \dot{\boldsymbol{r}}_c$,
- direction-dependent magnitude of the maximal applicable force $|\boldsymbol{F}_{\max}(\theta)| = F_1 + F_2 \cos\theta$,
- state-dependent observation uncertainty $\sigma_o^2 = s(\sigma_{\max}^2(1-\cos\Omega) + \sigma_{\min}^2)$.

The catching agent trades-off success with effort

$J = w_0 \lVert \boldsymbol{\mu}_b - \boldsymbol{\mu}_c \rVert_2^2$	[final position]
$+\; w_1 \operatorname{tr} \boldsymbol{\Sigma}_N$	[final uncertainty]
$+\; \tau w_2 \sum_{k=0}^{N-1} \operatorname{tr} \boldsymbol{\Sigma}_k$	[running uncertainty]
$+\; \tau \sum_{k=0}^{N-1} \boldsymbol{u}_k^T \boldsymbol{M} \boldsymbol{u}_k$	[total energy].

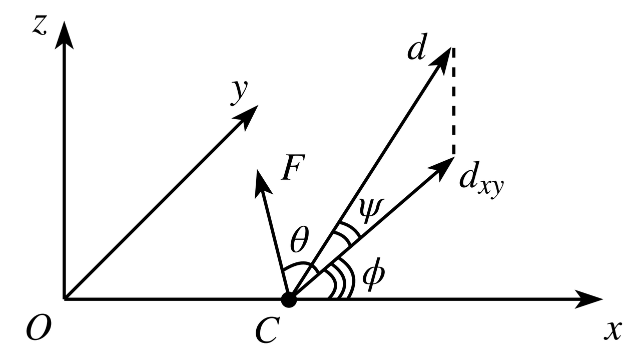

Figure 2: Catcher C applies force $\boldsymbol{F}$ to move in xy-plane. Unit vector $\boldsymbol{d}$, parameterized by angles ϕ and ψ, specifies the gaze direction. The catcher controls the **module of the force** F along with the **direction** of its application θ, and **angular velocities** ω_ϕ and ω_ψ.

Simulated experiments and results

Continuous tracking of an outfielder—heuristics hold

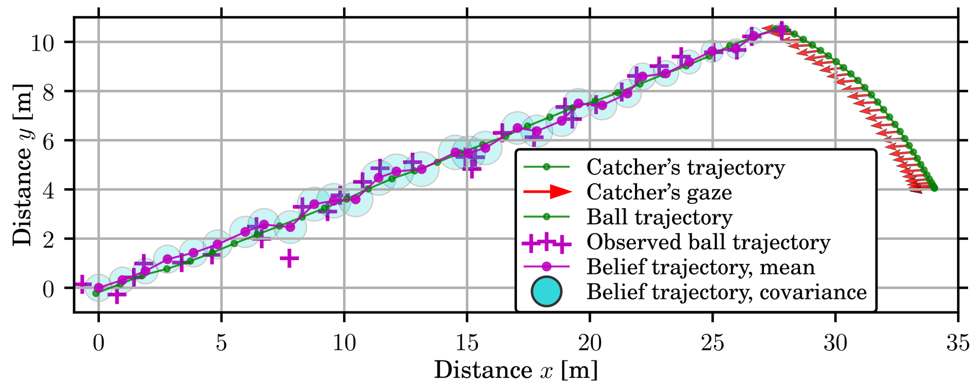

Figure 3: A typical trajectory of a successful catch.

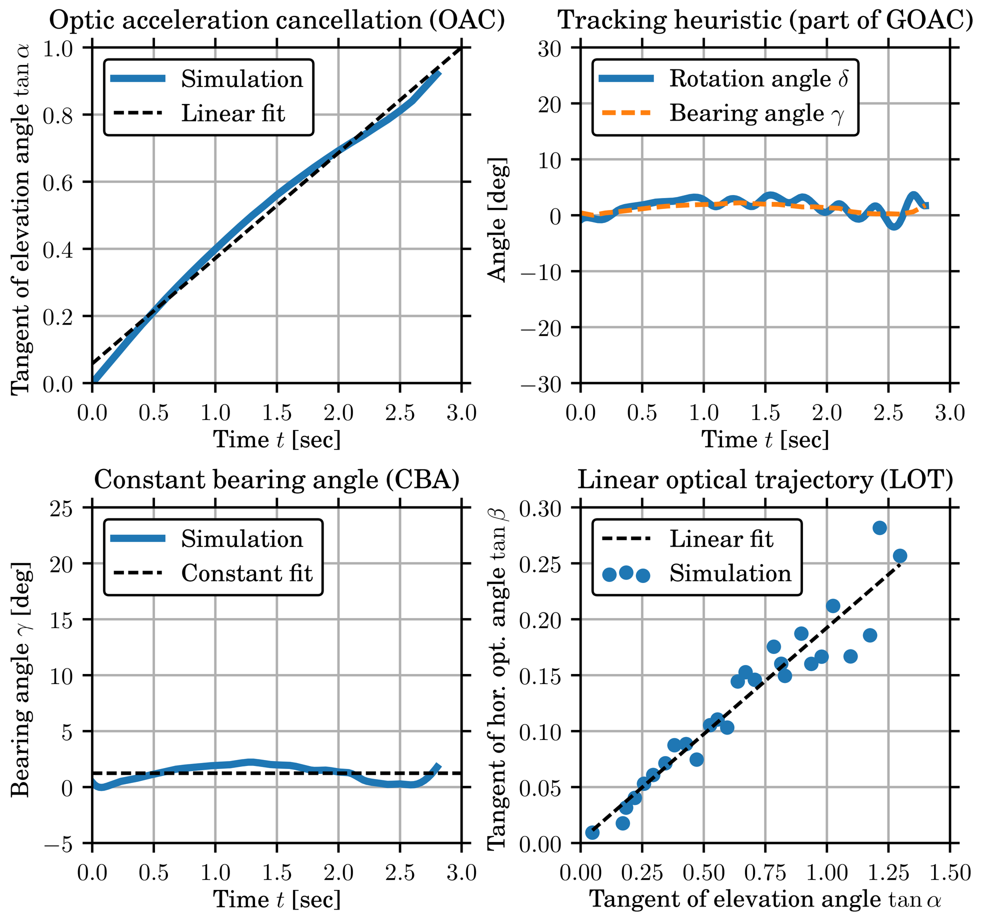

Figure 4: Heuristics for the successful catch.

In Figure 3, the agent starts sufficiently close to the interception point to continuously visually track the ball, therefore he is able to efficiently reduce uncertainty and intercept the ball while keeping it in sight. Note that the agent does not follow a straight trajectory but a curved one, in agreement with human experiments.

Figure 4 shows that resulting from our optimal control formulation policies always fulfill the heuristics (OAC, CBA, GOAC, and LOT) with approximately the same precision as in the original human experiments:

- $\tan\alpha$ grows linearly (OAC),
- γ remains constant (CBA),
- δ oscillates around γ (GOAC),
- $\tan\beta \propto \tan\alpha$ (LOT).

Thus, in this well-studied case, the model produces an optimal policy that exhibits behavior which is fully in accordance with the heuristics.

Interrupted tracking during long passes—heuristics break but prediction is required

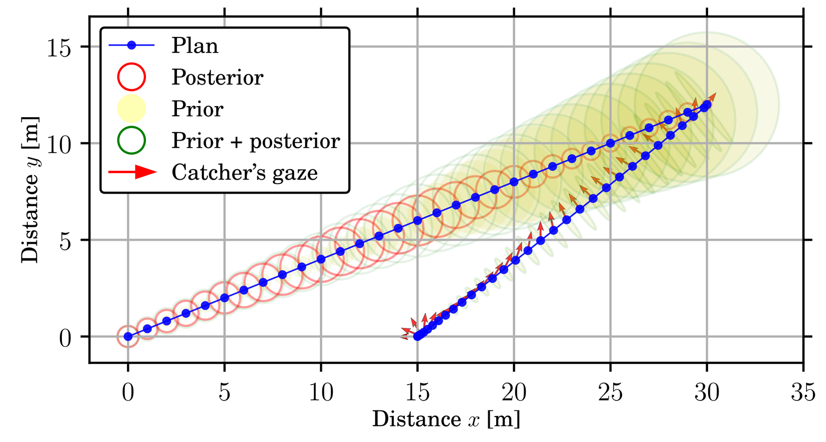

Figure 5: Catch that violates heuristics.

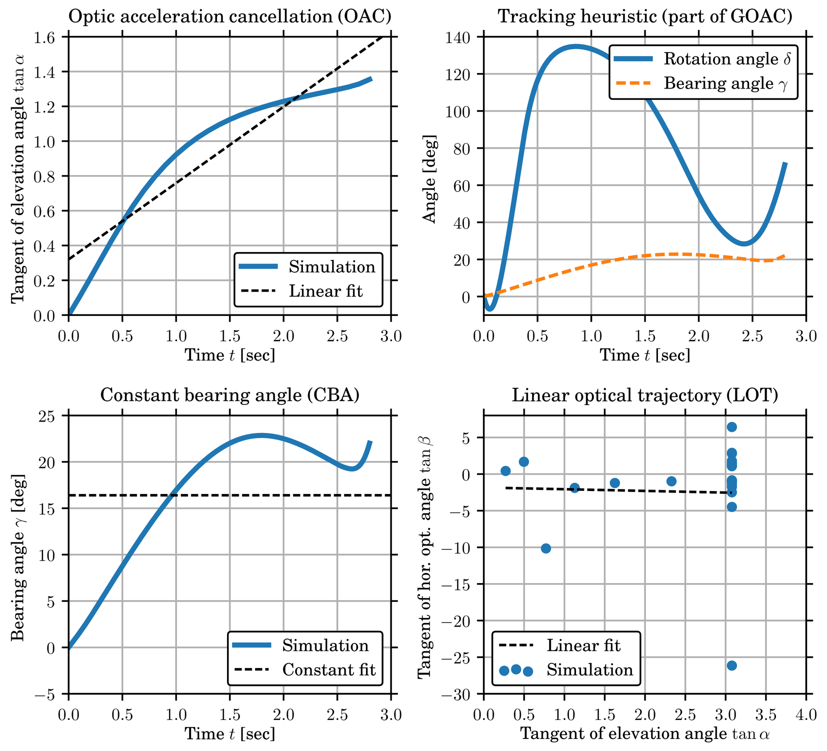

Figure 6: Heuristics do not hold.

An interception plan that leads to successful catch despite violating heuristics is shown in Figure 5. The agent would not be able to reach the interception point in time while running backwards and, thus, has to turn forward to run faster.

As seen from Figure 6, the heuristics fail to explain this catch—even during the final stage of the catch when the agent is continuously tracking the ball. OAC deviates from linearity, CBA is not constant, the tracking heuristic wildly deviates from the prediction, and LOT is highly non-linear.

Switching behaviors when uncertainty and reaction time are varied

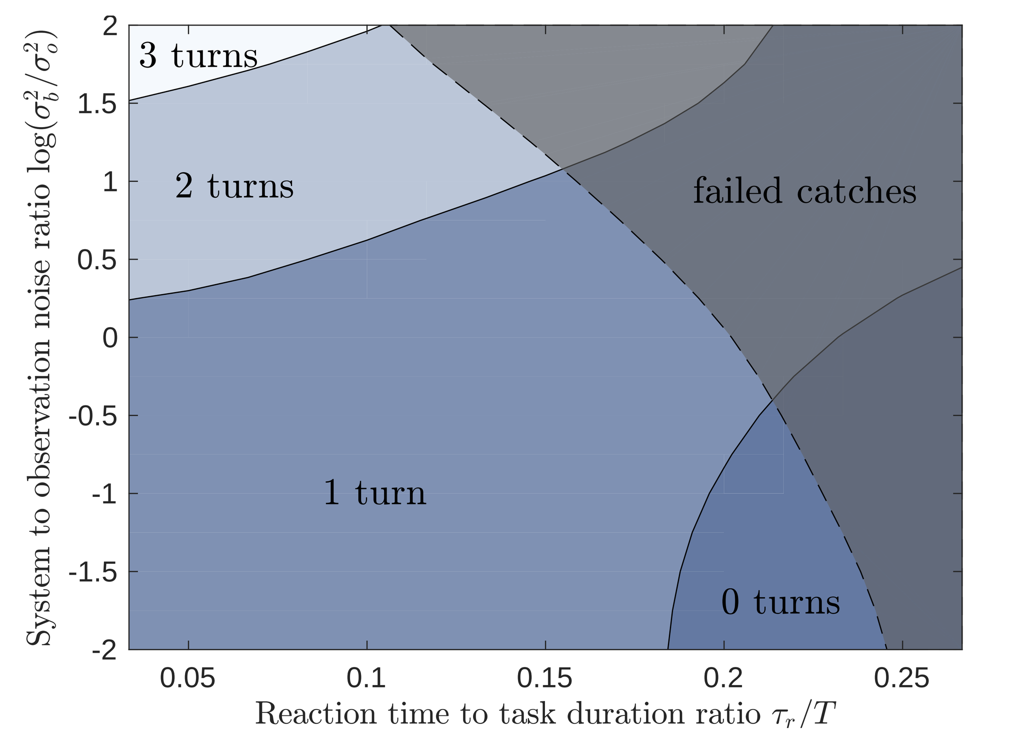

Figure 7: Switches between reactive and feedforward policies are determined by uncertainities and latency.

When reaction delays are long and predictions are reliable, the agent turns towards the interception points and runs as fast as he can (purely predictive strategies; lower right corner in Figure 7). When predictions are not sufficiently trustworthy, the agent has to switch multiple times between a reactive policy to gather information and a predictive feedforward strategy to successfully fulfill the task (upper left corner). When reaction time and system noise become sufficiently large, the agent fails to intercept the ball (upper right grayed out area). Thus, seemingly substantially different behaviors can be explained by means of a single model.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640554.