EXPLICIT ROLES IN GENERATIVE PLANNING: RETRAINING, SELECTION, AND ROBUSTNESS

Anonymous authors
Paper under double-blind review

ABSTRACT

The contribution of explicit role information to generative planning is difficult to isolate when geometry, demonstrations, and candidate selection can convey related information. We present a controlled intervention framework for vessel planning that separates training-time role availability, inference-time generator inputs, and role-dependent candidate selection. The framework combines architecture-matched retraining, generator–selector factorial interventions, and paired robustness tests to distinguish learning with explicit roles from reacting to their removal or corruption at inference. Across MLP diffusion, temporal U-Net diffusion, and flow matching, we pair 30 newly trained explicit-role-blind policies with 30 preserved conditioned counterparts. Our prospective evaluation comprises 1,608,000 episode executions in controlled synthetic environments, using ten training seeds per architecture, shared scene banks, and crossed resampling. Among six primary retraining contrasts, only MLP diffusion with correct selection exhibits a positive arrival difference supported by multiplicity-adjusted confidence intervals, with a gain of 1.76 percentage points. In contrast, descriptive factorial experiments show that cyclically incorrect generator codes reduce arrival by 7.03–19.07 percentage points relative to correct codes across the three architectures, with eight candidates and a correct selector held fixed. Paired noise–delay and current–lag experiments further reveal non-additive changes in closed-loop outcomes, highlighting the importance of evaluating coupled disturbances. Together, these findings distinguish the benefit of explicit-role training from sensitivity to incorrect role inputs and establish a reproducible evaluation protocol for examining conditioning, candidate selection, and robustness in autonomous vessel planning.

1 INTRODUCTION

A conditioned planner can succeed without using the explicit code that appears to explain its behavior. Geometry can reveal the same role, demonstrations can encode role-dependent actions, and selection can reintroduce role information after candidate generation. Test-time input ablation therefore answers a different question from retraining without that input.

Our vessel context contains five synthetic role coordinates alongside geometry and motion. *Explicit-role-blind* models zero only these coordinates; diagnostics examine information that remains.

The previous phase found that U-Net improved on a matched MLP without establishing an advantage over GMM; flow remained below GMM. They motivate separately specified new test namespaces and contrasts.

We ask what changes after architecture-matched retraining without the explicit code under correct or neutral selection; how separate generator and selector interventions interact; and how checkpoint choice and joint observation/execution errors qualify these comparisons. We separate these effects using established generative models, retaining failures and variability.

Anonymous code and videos: `https://anonymous.4open.science/r/lclr2027a9f3/`.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2 RELATED WORK

Conditioning and interpretation. Imitation learning can confuse predictive information with the causal structure relevant to control (de Haan et al., 2019); shortcut learning describes rules that fit an evaluation setting but transfer poorly (Geirhos et al., 2020). Underspecification provides a further reason to inspect behavior beyond a single heldout score (D'Amour et al., 2022). They motivate the evaluation questions, without making our explicit-channel intervention complete causal identification.

Generative planning and alternatives. Diffusion objectives and sampling methods provide the generative foundations (Ho et al., 2020; Nichol & Dhariwal, 2021; Song et al., 2021a). Diffuser and Diffusion Policy established planning and receding-horizon action generation with learned distributions (Janner et al., 2022; Chi et al., 2023); FiLM supplies feature-wise conditioning (Perez et al., 2018). Flow matching and rectified flow motivate interpolation-based velocity regression (Lipman et al., 2023; Liu et al., 2023). Mixture policies and sampling-based MPC remain essential alternatives (Bishop, 1994; Mandlekar et al., 2022; Pinneri et al., 2021; Williams et al., 2017a).

Scope of evidence. Constrained diffusion, belief-conditioned planning and safety-filter methods address distinct problems (Römer et al., 2025; Puthumanaillam et al., 2025; Xiao et al., 2025; Wabersich & Zeilinger, 2021). Finite candidate ranking inherits none of their safety guarantees. Marine models and numerical analysis inform the stipulated dynamics (Fossen, 2021; Butcher, 2016); seed variability and separate-factor resampling motivate explicit uncertainty accounting (Agarwal et al., 2021; Henderson et al., 2018; Owen, 2007). Appendix C records the wider source map and citation boundaries.

3 TASK, INFORMATION AND INTERVENTIONS

Task and interface. The planar state $s = (x, y, \psi, u, v, r)$ contains position, heading and body velocities. Two bounded actions control thrust and yaw moment. Context $c = (z, q_g) \in \mathbb{R}^{25}$ combines twenty motion/goal/mask features with five role coordinates; q is the fixed oracle initial role. Policies propose $H = 24$ actions over 12 s and execute four before replanning. Targets are nonreciprocal.

Arrival requires reaching the 5 m goal radius within 60 s before a 4 m center-distance collision; collision wins simultaneous events. Arrival, collision and timeout partition outcomes. Synchronized-chord event checks are numerical approximations. Synthetic initial roles are not evolving legal obligations.

Learned proposals. MLP and temporal U-Net DDPM predict Gaussian noise under a 32-step cosine schedule. Flow uses a dimension-matched MLP and linear Gaussian-to-action interpolation:

$$A_\tau = \sqrt{\bar{\alpha}_\tau} A_0 + \sqrt{1 - \bar{\alpha}_\tau}\epsilon, \qquad \mathcal{L}_D = \frac{1}{2H}\mathbb{E}\|\epsilon - \epsilon_\theta(A_\tau, \tau, c)\|_2^2, \tag{1}$$

$$X_t = (1 - t)Z + tA_0, \qquad \mathcal{L}_F = \frac{1}{2H}\mathbb{E}\|v_\theta(X_t, t, c) - (A_0 - Z)\|_2^2. \tag{2}$$

U-Net uses temporal convolutions, GroupNorm and FiLM. Flow uses 16 Heun intervals (32 network evaluations) and clips final actions. Table 1 reports parameters and budgets; Appendix A gives implementation details.

Independent retraining. New models train and validate with cloned contexts whose last five coordinates are zero. Both arms retain the objective and 0.15 role-dropout call, preserving intended within-seed draw order. Conditioned training already exposes the zero code, so inference nulling differs from retraining. Geometry and teacher actions remain available.

Generator and selector roles. On fixed conditioned weights, correct, null and cyclic-wrong generator inputs are e_q, 0_5 and $e_{(q+1) \bmod 5}$; blind-trained weights always receive 0_5. The selector uses

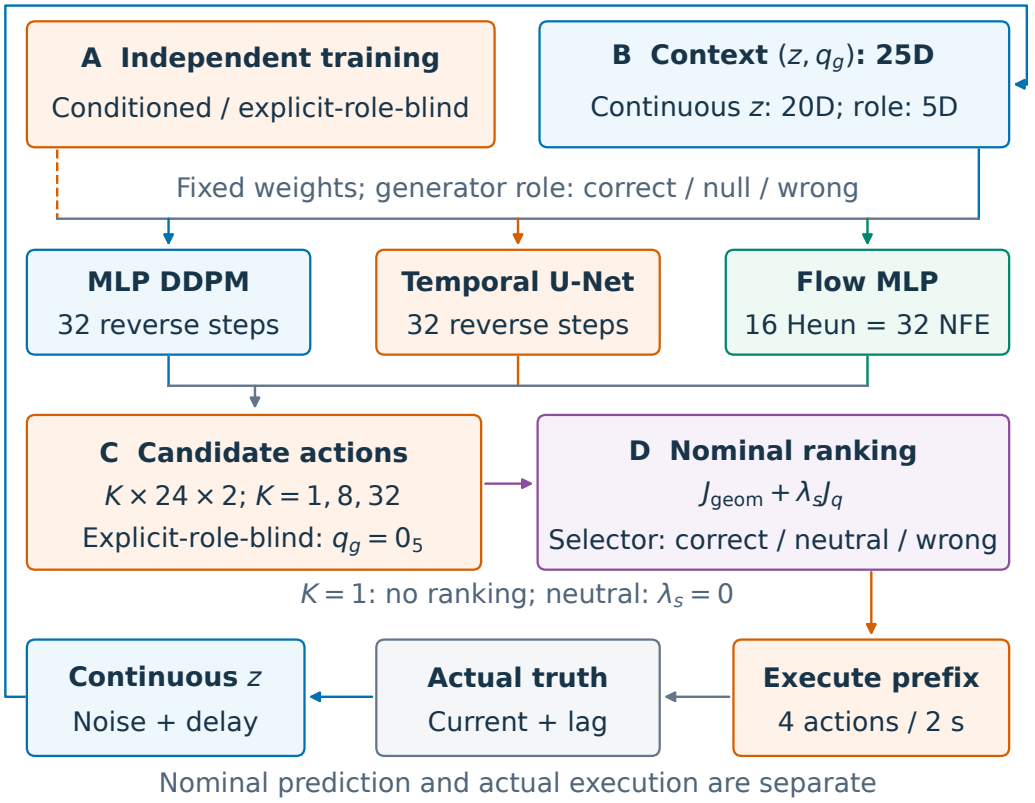


Figure 1: Training, generator input and selector cost are separate intervention sites. Blind training zeros five role channels; neutral selection sets $\lambda_s = 0$; wrong selection changes its own role copy. Feedback perturbs continuous observations while oracle initial q stays clean. $J_{\text{geom}} = J_g + 15J_c + 0.02J_a$ denotes the non-role score.

Table 1: Model families and registered main inference. Each of the three generative architectures has ten preserved conditioned and ten new explicit-role-blind training runs. All use 30,000 updates. Equal updates or close parameter counts do not equalize training FLOPs.

Family	Parameters	Main proposal budget
MLP DDPM	171,056	$K = 8$, 32 reverse steps
Temporal U-Net	179,906	$K = 8$, 32 reverse steps
Flow MLP	171,056	$K = 8$, 16 Heun steps
BC	150,576	One deterministic sequence
GMM-BC	337,672	Eight sampled sequences
CEM	—	640 / 4,096 proposals

its own q_s and weight λ_s:

$$k^\star = \arg\min_{1 \le k \le K} \{J_g(A^k) + 15J_c(A^k) + 0.02J_a(A^k) + \lambda_s J_q(A^k; q_s)\},$$
$$A^\star = A^{k^\star}. \tag{3}$$

Correct selection uses $(q_s, \lambda_s) = (q, 10)$; neutral uses $(q, 0)$; wrong uses $((q+1) \bmod 5, 10)$. Only a selector-side copy changes. Environment state and generator input stay unchanged. At $K = 1$, ranking is bypassed. For $K > 1$, the four costs measure goal distance, soft proximity, effort and role proxies; none is a feasibility certificate.

Table 2: Registered v05 extension. Realized counts are audited in Table 6. Main and last share one test bank. Other blocks use separate namespaces; all settings within each block share scenes. Executions are repeated episodes, not independent scene counts.

Block	Scenes	Settings	Jobs	Executions
Main	2,000	18	180	360,000
Fixed last	2,000	12	120	240,000
Role factorial	300	84	840	252,000
Noise × delay	300	108	1,080	324,000
Current × lag	300	144	1,440	432,000
Total	—	366	3,660	1,608,000

Reference systems. BC, eight-component GMM-BC and CEM share the nominal interface. CEM uses $128 \times 5 = 640$ or $512 \times 8 = 4096$ proposals at $H = 24$; Appendix B retains historical controls.

4 Prospective extension and statistical design

Stage boundary. The earlier phase contains 1,181,000 executions. Table 2 specifies 1,608,000 new executions in 3,660 jobs, designed before new test outcomes were available. The combined 2,789,000 counts executions, not a pooled confirmatory sample. Table 6 verifies completion.

Data and training pairs. Demonstrations contain 4,096 training and 512 validation scenes (45,985/5,877 decision records), including teacher failures. Uniform record sampling is not role-balanced. A fixed, unfitted rule recovers the initial role for all 4,096 training and 512 validation scenes. Across dependent teacher decision windows, frequency-weighted accuracy is 83.21% and 83.46%; head-on recall is only 44.63% and 49.12%. Initial recoverability does not establish redundancy throughout rollout. This particular rule also provides no bound on optimal recoverability. Appendix Figure 8 retains every role. Thirty new runs pair with thirty preserved CUDA-conditioned runs across three architectures and seeds 100–109. Within-architecture initialization and intended draw order are checked by short-trace tests; there is no cross-architecture pairing. Both arms use the same 30,000-update recipe. The first full-validation minimum defines the primary checkpoint; update 30,000 is the fixed-last sensitivity.

Main and factorial blocks. On 2,000 shared test scenes, six primary contrasts compare conditioned minus blind arrival for three architectures under correct/neutral selection. BC uses $K = 1$; GMM/CEM-640 use both selectors; CEM-4096 uses correct selection. Fixed-last reevaluates the twelve generative settings on this same bank.

A separate 300-scene validation bank crosses four generator conditions (correct, null, wrong, independently blind), three selectors and $K \in \{1, 8, 32\}$. At $K = 1$, ranking bypass makes selectors identical, so each generator is executed once, leaving 84 settings. Different K values do not form nested closed-loop candidate banks.

Joint stresses. The perception grid crosses position-noise standard deviations $\{0, 0.5, 1.0\}$ m with delays $\{0, 0.5, 1, 2\}$ s; heading and velocity noise scale with the frozen noise suite. Only continuous state/target observations are perturbed or delayed; the fixed oracle role, goal and masks stay clean. Perception robustness is therefore conditional on clean role side information. The execution grid crosses hidden world-frame y current $\{0, 0.15, 0.3, 0.45\}$ m/s with actuator lag $\{0, 0.25, 0.5, 1\}$ s. The latter uses 0.1 s RK4 substeps while planners retain nominal 0.5 s Euler prediction. Each grid compares nine methods on its own 300-scene test bank. Baselines and two-factor differences always use that block's matched zero-stress condition, never another bank's mean.

Table 3: Eighteen main settings: ten seeds, 2,000 shared scenes. Outcome percentages with descriptive 95% intervals. C/B: conditioned/explicit-role-blind; c/n: correct/neutral selection. Primary differences use 99.1667% intervals.

System	Arrival [CI]	Collision [CI]	Timeout [CI]
MLP C/c	56.83[54.98, 58.70]	41.71[39.75, 43.60]	1.47[0.97, 2.06]
MLP C/n	58.82[56.42, 61.19]	38.60[36.38, 40.75]	2.59[1.71, 3.54]
MLP B/c	55.07[52.89, 57.29]	44.11[41.81, 46.35]	0.83[0.57, 1.13]
MLP B/n	58.65[56.25, 61.03]	40.77[38.36, 43.12]	0.58[0.28, 0.98]
U-Net C/c	66.13[64.62, 67.69]	33.81[32.26, 35.33]	0.06[0.01, 0.14]
U-Net C/n	71.00[69.37, 72.68]	29.00[27.33, 30.63]	0.00[0.00, 0.00]
U-Net B/c	66.26[64.74, 67.81]	33.71[32.16, 35.22]	0.04[0.00, 0.10]
U-Net B/n	71.28[69.70, 72.91]	28.71[27.08, 30.30]	0.02[0.00, 0.05]
flow C/c	62.20[60.70, 63.73]	36.32[34.79, 37.79]	1.48[1.12, 1.88]
flow C/n	67.13[65.59, 68.70]	31.52[29.93, 33.06]	1.36[0.85, 1.92]
flow B/c	61.27[59.66, 62.89]	37.56[35.94, 39.15]	1.18[0.89, 1.51]
flow B/n	67.50[65.82, 69.18]	32.05[30.37, 33.72]	0.45[0.26, 0.67]
BC/c	64.09[62.54, 65.65]	35.80[34.24, 37.35]	0.12[0.04, 0.22]
GMM/c	66.55[65.06, 68.07]	33.14[31.62, 34.64]	0.31[0.18, 0.47]
GMM/n	69.39[67.84, 70.95]	30.26[28.69, 31.81]	0.36[0.20, 0.54]
CEM640/c	65.32[63.80, 66.84]	34.63[33.11, 36.13]	0.06[0.01, 0.15]
CEM640/n	73.46[71.95, 74.95]	26.55[25.05, 28.06]	0.00[0.00, 0.00]
CEM4096/c	65.74[64.20, 67.30]	34.26[32.70, 35.80]	0.01[0.00, 0.03]

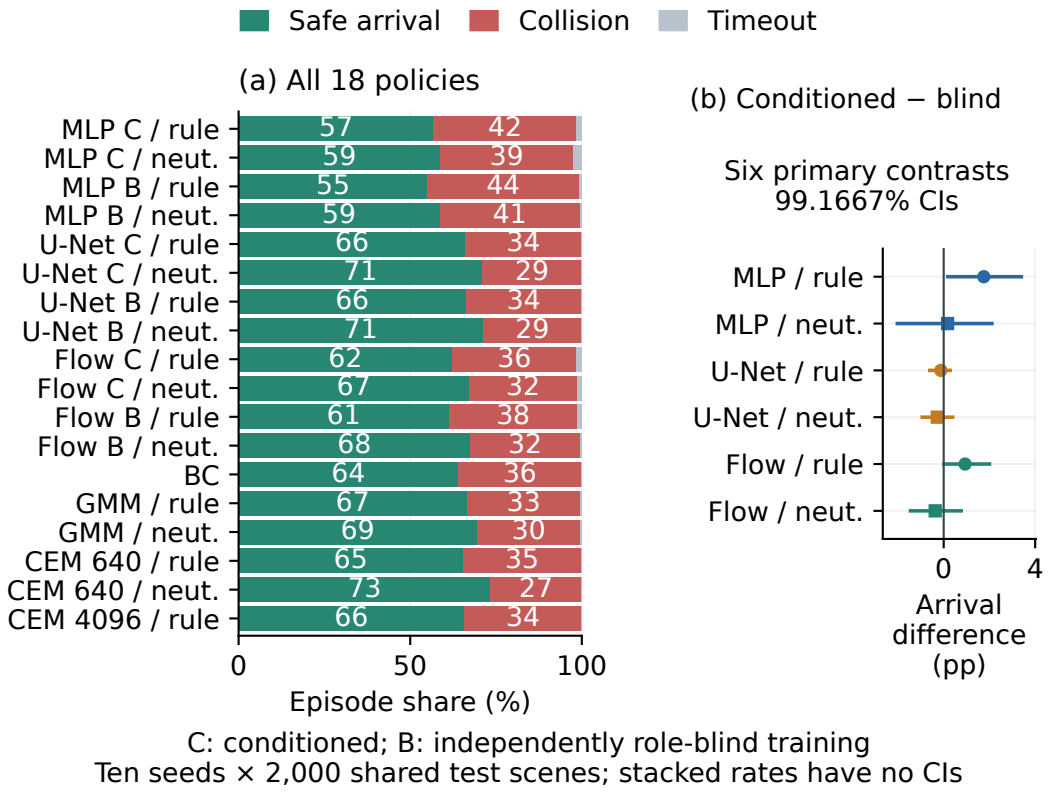


Figure 2: Eighteen main outcome decompositions and six conditioned-minus-blind arrival contrasts. Forest intervals are 99.1667%; stacked bars show point estimates. Seeds are variability blocks.

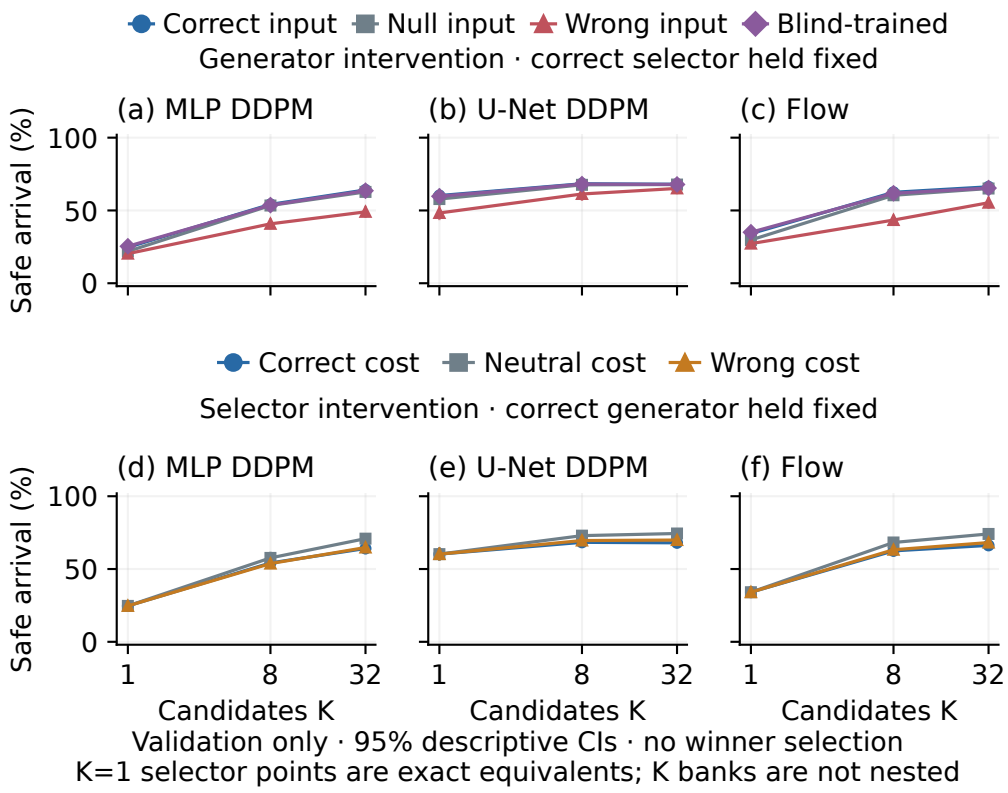


Figure 3: The 84-setting validation factorial. Separate blind training, conditioned-model nulling and cyclic wrong codes are distinct interventions. At $K = 1$, selector-equivalent points reuse the single executed setting.

5 Results and Registered Interpretations

5.1 Explicit Conditioning After Independent Retraining

Table 3 reports every main outcome and Figure 2 shows all six primary differences. Conditioned-minus-explicit-role-blind arrival differences, with correct then neutral selection, are MLP: $+1.76[0.10, 3.48]$, $+0.17[-2.11, 2.19]$; U-Net: $-0.13[-0.69, 0.36]$, $-0.28[-1.02, 0.47]$; flow: $+0.94[-0.07, 2.08]$, $-0.37[-1.53, 0.85]$ percentage points (99.1667% intervals). Adjusted intervals exclude zero for MLP with correct selection (above zero). The 5 zero-containing contrasts remain uncertain; they do not establish equivalence. Correct-selector arrival is BC $64.09[62.54, 65.65]$, GMM $66.55[65.06, 68.07]$, CEM640 $65.32[63.80, 66.84]$, CEM4096 $65.74[64.20, 67.30]$ percent. With neutral selection, GMM and CEM640 attain $69.39[67.84, 70.95]$ and $73.46[71.95, 74.95]$ percent (all reference intervals descriptive 95%). These controls preclude a claim of uniform generative superiority. Retraining holds architecture and recipe fixed while changing explicit-channel availability. Correct selection can supplement proposals with role costs; neutral selection removes that contribution. Geometry and demonstrations can still convey role information.

5.2 Independent Generator and Selector Interventions

At the prespecified $K = 8$ and correct selector, correct-generator minus null, wrong, and independently blind arrival differences are, in that order: MLP $+0.77[-1.13, 2.70]$, $+13.37[10.83, 15.97]$, $+0.50[-1.83, 2.83]$; U-Net $+0.73[-0.20, 1.63]$, $+7.03[4.10, 10.43]$, $+0.23[-0.73, 1.17]$; flow $+2.03[0.63, 3.40]$, $+19.07[15.87, 22.43]$, $+0.67[-0.90, 2.20]$ percentage points (descriptive 95% intervals). Figure 3 separates generator and selector changes; the complete 84 settings, including the single executed selector at $K = 1$, are retained in the appendix and data archive. These descrip-

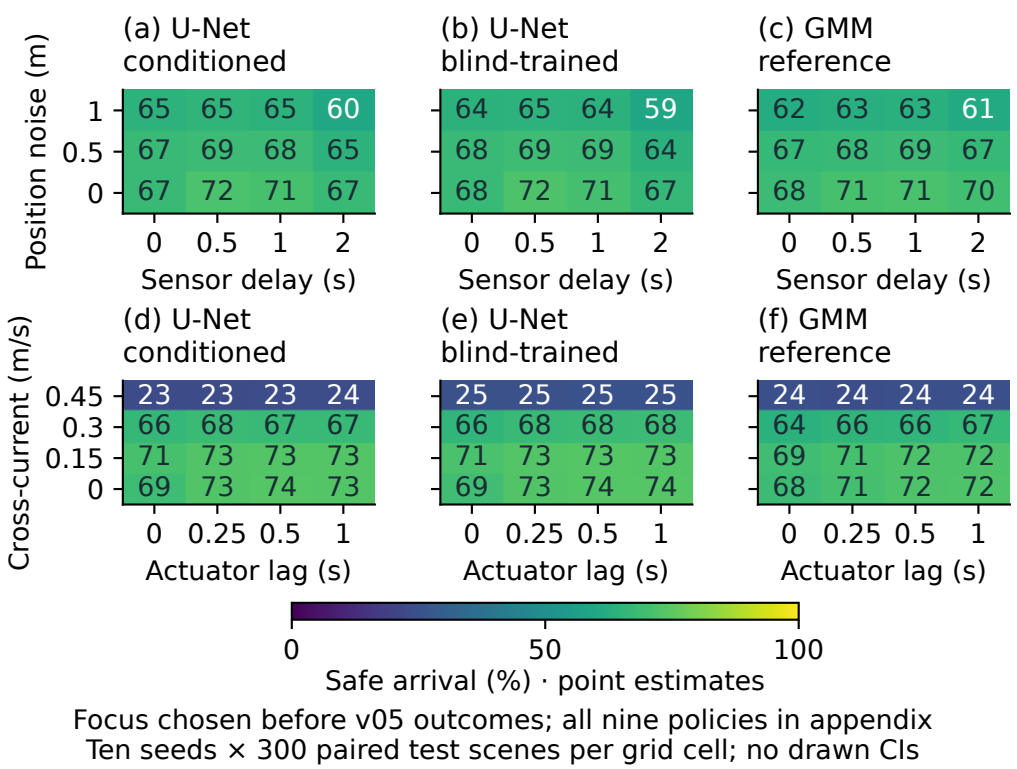


Figure 4: The prespecified noise–delay and current–lag focus panels, with complete nine-method grids in the appendix. Every shift and interaction uses its own matched zero-stress bank. Point-estimate heatmaps do not display uncertainty.

tive validation interventions do not replace the main family. Changing K changes both sampling opportunities and subsequent random allocation, precluding nested-candidate causal interpretation.

5.3 Joint perception and execution errors

At the fixed joint endpoint (1.0 m position noise and 2 s delay), arrival/collision/timeout percentages are conditioned U-Net 59.87/36.07/4.07; blind U-Net 59.17/36.37/4.47; GMM 61.13/36.73/2.13. In the same order, arrival changes from this block's matched zero-stress cells are $-7.60[-11.13, -4.27]$, $-8.50[-11.77, -5.40]$, $-7.27[-10.03, -4.53]$ percentage points; departure-from-additivity values are $-4.77[-7.87, -1.83]$, $-4.00[-7.10, -0.97]$, $-2.33[-5.03, 0.33]$ percentage points. All these intervals are descriptive 95%; the endpoint and three displayed policies were fixed independently of new outcomes. Noise and delay alter continuous observations while retaining the clean oracle role and nominal planner model. For outcome rate f, departure from additivity is

$$I(x, y) = f(x, y) - f(x, 0) - f(0, y) + f(0, 0). \tag{4}$$

Reported as $100I$ percentage points, this probability-scale interaction is positive when the rate exceeds its additive expectation and negative when below it, separately for each outcome. It does not identify a physical mechanism or transferable robustness. At the fixed joint endpoint (0.45 m/s current and 1 s actuator lag), arrival/collision/timeout percentages are conditioned U-Net 23.77/28.10/48.13; blind U-Net 25.27/28.23/46.50; GMM 24.40/28.73/46.87. In the same order, arrival changes from this block's matched zero-stress cells are $-45.17[-50.20, -39.97]$, $-43.97[-49.17, -38.57]$, $-44.03[-48.87, -39.17]$ percentage points; departure-from-additivity values are $-4.17[-7.07, -1.30]$, $-4.00[-7.20, -0.87]$, $-3.87[-6.53, -1.20]$ percentage points. All these intervals are descriptive 95%; the endpoint and three displayed policies were fixed independently of new outcomes. Lag can smooth as well as delay control; improved arrival alone

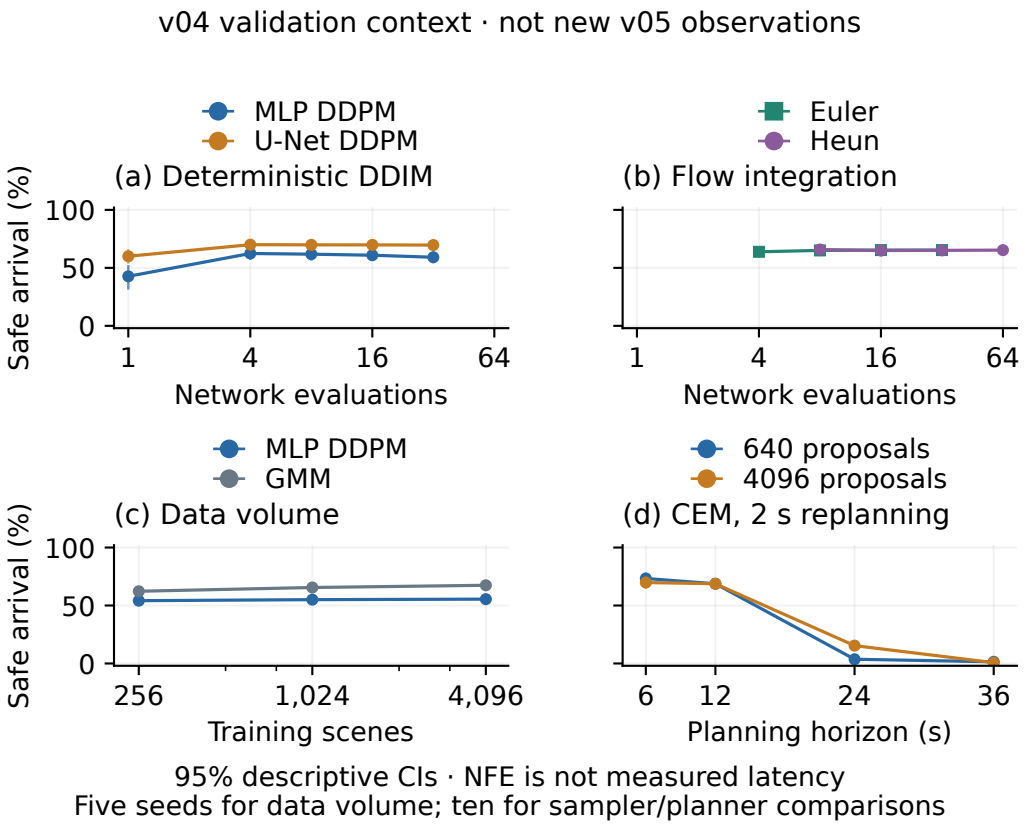


Figure 5: Earlier-phase v04 validation diagnostics, retained as historical context rather than new v05 outcomes. Data volume, sampler NFE and horizon use their original banks and repetitions, outside the new primary family.

does not establish a learned actuator mechanism. Collision and timeout decompositions distinguish different failure tradeoffs.

5.4 Budget context and checkpoint sensitivity

Longer-horizon CEM also increases search dimension at fixed effort; its poor archived result is not a verdict on externally tuned CEM (Appendix B). For correct selectors, fixed-last minus validation-best arrival differences for conditioned/blind training are MLP: $-0.05[-0.30, 0.11]$, $+0.17[-0.48, 0.96]$; U-Net: $+0.09[-0.31, 0.51]$, $-0.05[-0.37, 0.27]$; flow: $-0.13[-0.53, 0.10]$, $-0.19[-0.70, 0.16]$ percentage points (descriptive 95% intervals). Table 8 retains all twelve settings, including neutral selection. Both checkpoints share the main bank. Different validation-loss scales across DDPM and flow do not justify selecting checkpoints from test outcomes.

5.5 Matched behavior and failure interpretation

The prespecified seed-100 head-on (scene 0) and crossing give-way (scene 12) cases end in collision for all three architectures under both training regimes (twelve runs). Head-on termination occurs at 1.885–1.886 s, before the scheduled second replan at 2 s. This does not establish unavoidable collision or an optimal-control bound. The fixed cases do not estimate population differences. These fixed failures illustrate local behavior; aggregate comparisons determine population-level conclusions.

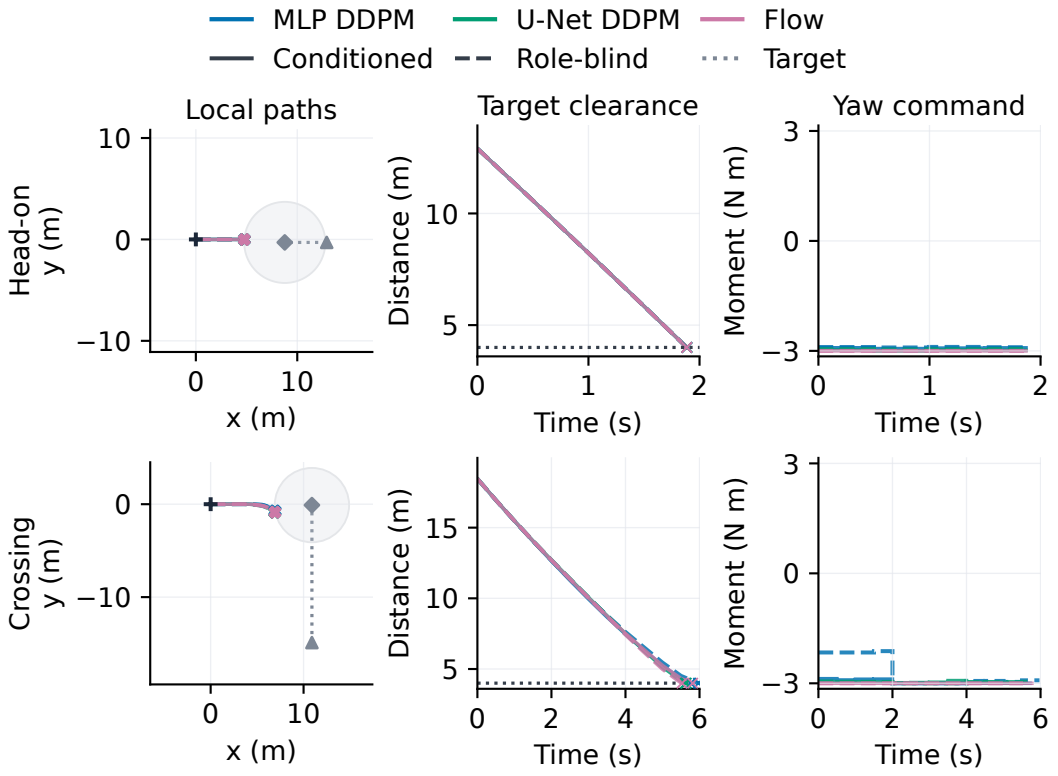


Figure 6: Fixed seed-100 head-on/crossing cases under both training regimes. Local views can omit goals/targets; distance uses all active targets with a 4 m threshold. The last column shows executed yaw moment $3a^M$ (N m), held between samples. No outcome-based case selection or interpolation is used.

6 Conclusion

We presented a controlled intervention framework for evaluating explicit role information in generative vessel planning. Its central contribution is to separate training-time role availability, inference-time generator inputs, and role-dependent candidate selection, distinguishing independent retraining from test-time input ablation and downstream ranking. Across MLP diffusion, temporal U-Net diffusion, and flow matching, we paired thirty newly trained explicit-role-blind policies with thirty preserved conditioned counterparts. The prospective evaluation comprises 1,608,000 episode executions, integrating architecture-matched comparisons, generator–selector factorial interventions, checkpoint sensitivity analysis, and paired robustness tests within a shared statistical protocol. Our findings distinguish the benefit of explicit-role training from sensitivity to incorrect role inputs. Among the six primary retraining contrasts, MLP diffusion with correct selection is the only setting whose multiplicity-adjusted interval supports a positive arrival difference, with an estimated gain of 1.76 percentage points. Meanwhile, the descriptive factorial analysis shows that replacing correct generator codes with cyclically incorrect codes reduces arrival by 7.03–19.07 percentage points across the three architectures at eight candidates and a fixed correct selector. Together, these results demonstrate why training-time availability and inference-time corruption require separate evaluation. This work provides a reproducible basis for assessing the contribution of explicit conditioning and examining how candidate generation, selection, and disturbances shape autonomous planning performance. Future work will investigate learned, temporally evolving role estimates and uncertainty-aware candidate selection, and extend the evaluation to reciprocal multi-vessel encounters and real-world vessel trials.

REFERENCES

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron Courville, and Marc G. Bellemare. Deep Reinforcement Learning at the Edge of the Statistical Precipice. In *Advances in Neural Information Processing Systems*, 2021. doi: 10.48550/arXiv.2108.13264. URL https://arxiv.org/abs/2108.13264. DOI identifies the arXiv preprint.

Aaron D. Ames, Xiangru Xu, Jessy W. Grizzle, and Paulo Tabuada. Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control*, 2017. doi: 10.1109/TAC.2016.2638961. URL https://doi.org/10.1109/TAC.2016.2638961.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis M. Kochmann. Physics-Informed Diffusion Models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/096347b4efc264ae7f07742fea34af1f-Abstract-Conference.html.

Christopher M. Bishop. Mixture Density Networks. Technical Report NCRG/94/004, Aston University, 1994. URL https://publications.aston.ac.uk/id/eprint/373/1/NCRG_94_004.pdf.

J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 3 edition, 2016. doi: 10.1002/9781119121534. URL https://onlinelibrary.wiley.com/doi/book/10.1002/9781119121534.

João Carvalho, An Thai Le, Mark Baierl, Dorothea Koert, and Jan Peters. Motion planning diffusion: Learning and planning of robot motions with diffusion models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2023. URL https://arxiv.org/abs/2308.01557.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. In *Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026. URL https://www.roboticsproceedings.org/rss19/p026.pdf.

Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models. In *Advances in Neural Information Processing Systems*, 2018. doi: 10.48550/arXiv.1805.12114. URL https://arxiv.org/abs/1805.12114. DOI identifies the arXiv preprint.

Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D. Hoffman, Farhad Hormozdiari, Neil Houlsby, Shaobo Hou, Ghassen Jerfel, Alan Karthikesalingam, Mario Lucic, Yian Ma, Cory McLean, Diana Mincu, Akinori Mitani, Andrea Montanari, Zachary Nado, Vivek Natarajan, Christopher Nielson, Thomas F. Osborne, Rajiv Raman, Kim Ramasamy, Rory Sayres, Jessica Schrouff, Martin Seneviratne, Shannon Sequeira, Harini Suresh, Victor Veitch, Max Vladymyrov, Xuezhi Wang, Kellie Webster, Steve Yadlowsky, Taedong Yun, Xiaohua Zhai, and D. Sculley. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, 23(226):1–61, 2022. URL https://www.jmlr.org/papers/v23/20-1335.html.

Pim de Haan, Dinesh Jayaraman, and Sergey Levine. Causal confusion in imitation learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.nips.cc/paper_files/paper/2019/hash/947018640bf36a2bb609d3557a285329-Abstract.html.

Pete Florence, Corey Lynch, Andy Zeng, Oscar A. Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit Behavioral Cloning. In *Proceedings of the 5th Conference on Robot Learning*, 2022. URL https://proceedings.mlr.press/v164/florence22a.html.

Thor I. Fossen. *Handbook of Marine Craft Hydrodynamics and Motion Control*. Wiley, 2 edition, 2021. doi: 10.1002/9781119575016. URL https://doi.org/10.1002/9781119575016.

Thor I. Fossen. Fossen's Marine Craft Model. Official author technical reference, n.d. URL https://www.fossen.biz/html/marineCraftModel.html. Official reference; accessed September 22, 2026.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020. doi: 10.1038/s42256-020-00257-z. URL https://doi.org/10.1038/s42256-020-00257-z.

Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep Reinforcement Learning That Matters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. doi: 10.1609/aaai.v32i1.11694. URL https://ojs.aaai.org/index.php/AAAI/article/view/11694.

Jonathan Ho and Tim Salimans. Classifier-Free Diffusion Guidance, 2022. URL https://arxiv.org/abs/2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020. URL https://arxiv.org/abs/2006.11239.

Sture Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 1979. doi: 10.2307/4615733. URL https://www.jstor.org/stable/4615733.

Michael Janner, Yilun Du, Joshua Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9902–9915, 2022. URL https://proceedings.mlr.press/v162/janner22a.html.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2210.02747. URL https://arxiv.org/abs/2210.02747. DOI identifies the arXiv preprint.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 1678–1690, 2022. URL https://proceedings.mlr.press/v164/mandlekar22a.html.

D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 2000. doi: 10.1016/S0005-1098(99)00214-9. URL https://www.sciencedirect.com/science/article/abs/pii/S0005109899002149.

D. Q. Mayne, M. M. Seron, and S. V. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 2005. doi: 10.1016/j.automatica.2004.08.019. URL https://www.sciencedirect.com/science/article/abs/pii/S0005109804002870.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Art B. Owen. The pigeonhole bootstrap. *The Annals of Applied Statistics*, 2007. doi: 10.1214/07-AOAS122. URL https://artowen.su.domains/reports/AOAS122.pdf.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual Reasoning with a General Conditioning Layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. doi: 10.1609/aaai.v32i1.11671. URL https://ojs.aaai.org/index.php/AAAI/article/view/11671.

Cristina Pinneri, Shambhuraj Sawant, Sebastian Blaes, Jan Achterhold, Joerg Stueckler, Michal Rolinek, and Georg Martius. Sample-efficient Cross-Entropy Method for Real-time Planning. In *Proceedings of the 2020 Conference on Robot Learning*, 2021. URL https://proceedings.mlr.press/v155/pinneri21a.html.

Gokul Puthumanaillam, Aditya Penumarti, Manav Vora, Paulo Padrao, Jose Fuentes, Leonardo Bobadilla, Jane Shin, and Melkior Ornik. Belief-conditioned one-step diffusion: Real-time trajectory planning with just-enough sensing. In *Proceedings of the 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 68–92, 2025. URL https://proceedings.mlr.press/v305/puthumanaillam25a.html.

robomimic core team. Models: Implemented policy networks, n.d. URL https://robomimic.github.io/docs/modules/models.html. Official robomimic documentation; accessed September 22, 2026.

Ralf Römer, Alexander von Rohr, and Angela P. Schoellig. Diffusion predictive control with constraints. In *Proceedings of the 7th Annual Learning for Dynamics and Control Conference*, volume 283 of *Proceedings of Machine Learning Research*, pp. 791–803, 2025. URL https://proceedings.mlr.press/v283/romer25a.html.

R. Rubinstein. The Cross-Entropy Method for Combinatorial and Continuous Optimization. *Methodology and Computing in Applied Probability*, 1999. doi: 10.1023/A:1010091220143. URL https://link.springer.com/article/10.1023/A:1010091220143.

Nur Muhammad Shafiullah, Zichen Jeff Cui, Ariuntuya Altanzaya, and Lerrel Pinto. Behavior Transformers: Cloning k modes with one stone. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2206.11251.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*, 2021a. doi: 10.48550/arXiv.2010.02502. URL https://arxiv.org/abs/2010.02502. DOI identifies the arXiv preprint.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021b. URL https://arxiv.org/abs/2011.13456.

United States Coast Guard. USCG Amalgamated Navigation Rules International & U.S. Inland. Official Navigation Center reference, 2024. URL https://www.navcen.uscg.gov/navigation-rules-amalgamated. Official reference; accessed September 22, 2026.

Kim Peter Wabersich and Melanie N. Zeilinger. A predictive safety filter for learning-based control of constrained nonlinear dynamical systems. *Automatica*, 2021. doi: 10.1016/j.automatica.2021.109597. URL https://www.sciencedirect.com/science/article/abs/pii/S0005109821001175.

Grady Williams, Andrew Aldrich, and Evangelos A. Theodorou. Model Predictive Path Integral Control: From Theory to Parallel Computation. *Journal of Guidance, Control, and Dynamics*, 2017a. doi: 10.2514/1.G001921. URL https://arc.aiaa.org/doi/10.2514/1.G001921.

Grady Williams, Paul Drews, Brian Goldfain, James M. Rehg, and Evangelos A. Theodorou. Information theoretic model predictive control: Theory and applications to autonomous driving, 2017b. URL https://arxiv.org/abs/1707.02342.

Wei Xiao, Tsun-Hsuan Wang, Chuang Gan, Ramin Hasani, Mathias Lechner, and Daniela Rus. SafeDiffuser: Safe Planning with Diffusion Probabilistic Models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/f95606d8e870020085990d9650b4f2a1-Paper-Conference.pdf.

Shanliang Yao, Runwei Guan, Zhaodong Wu, Yi Ni, Zile Huang, Ryan Wen Liu, Yong Yue, Weiping Ding, Eng Gee Lim, Hyungjoon Seo, Ka Lok Man, Jieming Ma, Xiaohui Zhu, and Yutao Yue. WaterScenes: A Multi-Task 4D Radar-Camera Fusion Dataset and Benchmarks for Autonomous Driving on Water Surfaces. *IEEE Transactions on Intelligent Transportation Systems*, 2024. doi: 10.1109/TITS.2024.3415772. URL https://arxiv.org/abs/2307.06505.

STATEMENT ON THE USE OF AI

Generative artificial intelligence tools were used to polish part of the wording, improving clarity and grammatical accuracy.

ANONYMOUS ARTIFACTS

Code and videos repository

https://anonymous.4open.science/r/lclr2027a9f3/

A IMPLEMENTATION AND COMPLETE PROTOCOL

A.1 TRAINING INTERVENTION AND SAMPLING MEASURE

The new trainer preserves the three architecture kinds (conditional MLP DDPM, temporal U-Net DDPM and MLP flow matching). It clones each $N \times 25$ condition tensor and zeros coordinates 20–24, using zero-based indexing, for explicit-role-blind training. The original objective then still performs its 0.15 role-dropout draw. This retains the call order for dropout, noise/time and mini-batch sampling rather than switching to the older unconditional-model objective, which omitted the dropout draw. Validation applies the same explicit-channel mask and its fixed RNG stream. Because conditioned training includes role dropout, inference nulling does not introduce a wholly unseen code. The main retraining contrast changes explicit-feature availability during training, validation selection and inference; the registered fixed-last sensitivity removes the validation-best selection step.

Paired seeds are 100–109 within an architecture. The design preserves the initialization and intended stochastic stream relative to the archived CUDA trainer, with short-trace implementation tests. Optimization trajectories and final weights need not match because the conditioning information differs. Seeds across architectures do not imply shared initial parameters. Best checkpoints use the first strictly lowest full-validation objective, evaluated at update 1 and every 1,000 updates through 30,000. The fixed-last analysis always uses exactly update 30,000.

The training-record role counts are $(10{,}163, 5{,}517, 6{,}516, 9{,}723, 14{,}066)$ in clear/head-on/give-way/stand-on/overtaking order; validation counts are $(1{,}239, 623, 962, 1{,}297, 1{,}756)$. Initial scene counts are nearly balanced, but uniform record sampling gives these unequal frequencies. An unfitted rule first identifies clear scenes by the primary mask; otherwise it reconstructs the primary target's body velocity from relative velocity and own velocity, separates crossings by the sign of lateral motion, and separates head-on/overtaking by longitudinal direction. It recovers every audited initial label on 4,096 training and 512 validation scenes. The same unmodified rule recovers 38,266/45,985 training windows (83.21%) and 4,905/5,877 validation windows (83.46%). Excluding initial states gives 81.57% and 81.88%. Head-on recall is 2,462/5,517 (44.63%) in training and 306/623 (49.12%) in validation; later-only recall is 1,642/4,697 (34.96%) and 203/520 (39.04%). Aggregate accuracy weights recorded teacher windows by their frequency, not equally by role, scene or time. Windows within a scene are dependent; these descriptive proportions carry no window-independent confidence interval. The target is the stored initial role, and geometry changes during encounters. Initial recoverability therefore does not establish redundancy throughout rollout. Conversely, errors of this particular unfitted rule establish neither optimal unrecoverability nor an information-theoretic bound. Neither this diagnostic nor a null role code proves that a trained network uses or ignores the recoverable information.

Training uses float32, batch 512, AdamW learning rate 3×10^{-4}, weight decay 10^{-4} and gradient clipping at 1. Records are sampled uniformly. Role-stratified scene construction does not imply equal numbers of trajectory windows or equal per-role loss weight. The source contains all teacher trajectories, including failures; no test-dependent demonstration filtering is allowed.

A.2 CONTINUOUS CONTEXT AND NOMINAL DYNAMICS

Actions $a \in [-1, 1]^2$ map to thrust $F = 15(a^F + 1)$ N and yaw moment $M = 3a^M$ N m. The state components are world position (x, y), heading ψ, body surge/sway (u, v) and yaw rate r. The temporal U-Net uses lengths $24/12/6/12/24$ with channels $24/48/96/48/24$.

The twenty non-role coordinates contain three body-velocity features, two body-frame goal coordinates, and three target slots with five coordinates each (relative position, relative velocity and an activity mask). Inactive slots are zero. Position, speed and yaw-rate scales are 60 m, 3 m/s and 0.3 rad/s. The nominal equations are

$$\begin{aligned} \dot{x} &= u\cos\psi - v\sin\psi + c_x, & \dot{y} &= u\sin\psi + v\cos\psi + c_y, & \dot{\psi} &= r, \\ m\dot{u} &= F - d_u u + mvr, & m\dot{v} &= -d_v v - mur, & I\dot{r} &= M - d_r r. \end{aligned} \tag{5}$$

Here $m = 50$ kg, $I = 20$ kg m^2, and $(d_u, d_v, d_r) = (10, 20, 8)$ in the associated units. The nominal planner sets $c = 0$ and uses 0.5 s Euler transitions. Constant world-frame current is imposed kinematically, not as a full relative-flow force model.

BC predicts a bounded sequence with squared imitation loss. GMM-BC has eight diagonal Gaussian components in inverse-tanh action coordinates, with learned mixture weights and variances and the tanh Jacobian in density evaluation. Saturated demonstrations are clipped to $[-0.99999, 0.99999]$ for that continuous density. DDPM uses 32 sinusoidal time features and three 256-wide SiLU layers in the MLP case. The U-Net uses fixed linear interpolation for deterministic CUDA backward compatibility. Heun flow evaluates predictor and corrector at each interval, for 32 calls at 16 intervals.

A.3 Nominal objective and exact selector meanings

For predicted active-target distances d_{hj} and masks b_j,

$$J_c = H^{-1} \sum_{h,j} b_j \ \mathrm{softplus}((4 - d_{hj})/1)^2, \qquad J_a = H^{-1} \sum_{h} [((a_h^F + 1)/2)^2 + (a_h^M)^2].$$

J_g is normalized terminal squared goal distance. J_q uses the initial role and risk to penalize starboard, crossing-ahead, stand-on and separation proxies as configured. Neutral selection changes only the role penalty weight from ten to zero. Wrong selection cyclically changes only the selector's integer role copy. The true environment label and the generator tensor remain intact unless their own registered intervention is applied.

At $K = 1$, scoring is bypassed for all selector modes. The factorial therefore executes one selector setting per architecture and generator condition, not three duplicate runs. Per architecture it has $4 + 4 \times 3 \times 2 = 28$ settings, or 84 across the three architectures. At $K = 8$ and 32 the full selector crossing remains. Distinct closed-loop K values do not form nested proposal banks because batch shape and episode termination change random allocation.

The wrong-selector intervention is the specific cyclic map $q \mapsto (q + 1) \bmod 5$, not symmetric semantic corruption. Overtaking $4 \to 0$ disables the role penalty, whereas clear $0 \to 1$ has zero masked primary-target risk; the role costs are not normalized to equal strength. Interpret each setting under this implemented map.

A.4 Banks, reference methods and complete accounting

Main and fixed-last use seed namespace 81001 with 2,000 scenes. The factorial uses validation namespace 82001 with 300 scenes. Perception and execution use separate test namespaces 83001 and 84001 with 300 scenes each. New seeds reuse the frozen scenario generator; they do not constitute a new scene-generation implementation. Changing seed or scene count is not a prefix operation.

The main has twelve generative settings (three architectures, two training regimes, two selectors), BC at its unique one-candidate selector setting, GMM at correct/neutral selectors, CEM-640 at correct/neutral selectors, and CEM-4096 at correct selection: eighteen settings. The last-checkpoint block contains only the twelve generative settings. The stress grids use six generative methods plus BC, GMM and CEM-640, all with the correct selector. GMM and BC retain their archived checkpoints; CEM repetitions are planner RNG rather than independent training runs. Backend and checkpoint provenance are retained per run.

The planned five-block matrix has 366 settings, 3,660 jobs and 1,608,000 episode executions. Only completed, hash-verified reports may establish final realized counts. Thirty new explicit-role-blind runs extend the existing training archive; preserved conditioned checkpoints must not be counted as newly trained runs.

Policy proposal sampling acts only on the alive batch. The policy RNG is seeded once for each evaluation replicate and used across sequential chunks; changing termination times or alive-batch shapes can change which Gaussian draws reach a particular scene, even at equal K. Shared policy seeds are replicate blocks, not per-scene pathwise candidate-noise coupling. Observation innovations have a different contract: each macrostep resets a separate CPU generator from the fixed chunk offset and step and perturbs the full stored state/target batch before alive selection, coupling these innovations across matched evaluations. This observation coupling is distinct from the intended within-architecture training draw order and its short-trace tests.

A.5 PAIRED TWO-FACTOR STRESSES

Perception position-noise standard deviations are 0, 0.5 and 1.0 m. At the 0.5 m level, associated heading, linear-velocity and yaw-rate standard deviations are 2°, 0.1 m/s and 0.01 rad/s; they scale proportionally for other levels. Delays are 0, 0.5, 1 and 2 s. Noise and history buffering affect continuous state/target observations while the nominal controller model remains unchanged. The initial oracle role q, goal and target masks are not perturbed or reclassified. Conditioned generators and correct selectors therefore retain clean symbolic side information even at the largest observation stress; this is not end-to-end role-perception robustness.

Execution uses current $(0, c_y)$ with $c_y \in \{0, 0.15, 0.3, 0.45\}$ m/s and actuator time constant $\tau \in \{0, 0.25, 0.5, 1\}$ s. For $\tau > 0$, actual wrench follows

$$\dot{w} = (w_{\mathrm{cmd}} - w)/\tau, \qquad w(h) = w_{\mathrm{cmd}} + (w_0 - w_{\mathrm{cmd}})e^{-h/\tau}.$$

For zero lag, the actuator is instantaneous. Truth uses 0.1 s RK4 substeps and synchronized float64 chord-event calculations; retained state arrays may use float32. Wrench integration, event truncation and raw metric reconstruction follow the audited physical evaluator. No integration-refinement superiority is inferred merely from using a smaller numerical step.

Outcome-rate interactions follow Equation 4 with common within-block seed/scene weights. Different banks are never pooled for an interaction. The execution zero-current/zero-lag reference still uses RK4 truth, not the nominal main bank's Euler truth. We report $100I$ in percentage points: positive means the specified outcome rate is above its additive expectation, negative below it. This is probability-scale nonadditivity, not automatic physical synergy. The descriptive interaction scale is distinct from an odds ratio, a hazard model or a causal claim about real vessels.

Saved-trajectory auditing distinguishes exact saved-chord event-label agreement from rounding-boundary compatibility. The latter uses coordinate quantization bounds plus an explicit float32 arithmetic budget: an engineering allowance, not a theorem covering every GPU operation. Boundary-compatible cases are retained and separately counted, without relabeling outcomes or enlarging the physical collision/goal radii. Thus passing the raw audit is not a claim that every event was reconstructed without tolerance, nor an independent full policy or continuous-physics rerun.

A.6 RESAMPLING AND REPORTING CONTRACT

The bootstrap uses 10,000 draws and seed 95001, independently resampling the training/planner replicate axis and role-stratified scene axis. Paired settings share the same weights. The six primary main contrasts are conditioned minus explicit-role-blind for MLP, U-Net and flow, each under correct and neutral selection. Bonferroni individual confidence is $1 - 0.05/6 = 0.991666\ldots$. The earlier phase's four-comparison 98.75% family is separate.

All other intervals are 95% descriptive intervals, including fixed-last, per-role, factorial and grid interactions. More rows do not confer multiplicity-adjusted discovery status. We make no equivalence or noninferiority claim. All outcome denominators, finite continuous-metric counts, seed rows, hashes and failures must remain available. Any later latency or qualitative display is separately labeled unless explicitly present in this registered matrix.

B EARLIER V04 EVIDENCE RETAINED SEPARATELY

This section reports completed historical evidence that was available before v05 design. The v04 phase has 115 registered training runs, 3,580 evaluation jobs and 1,181,000 executions. Its primary family is distinct: U-Net minus matched MLP, flow minus matched MLP, U-Net minus GMM and flow minus GMM, using 5,000 crossed-bootstrap draws and 98.75% individual intervals.

Those differences are $+9.40\,[8.01, 10.79]$, $+5.51\,[4.05, 6.96]$, $-0.64\,[-1.49, 0.26]$ and $-4.53\,[-5.52, -3.57]$ percentage points. The U-Net–GMM interval includes zero without demonstrating equivalence; flow is below GMM in this registered family. At 4,096 CEM proposals, horizon 24 versus 48 gives 64.01% versus 15.02% arrival and 0.01% versus 53.33% timeout. Longer horizon also enlarges search dimensionality at fixed optimization effort, so this is not a verdict on fully tuned external CEM.

Table 4: Completed earlier-phase v04 main test on 2,000 shared scenes and ten repetitions; all outcomes are percentages. Arrival has descriptive 95% crossed-bootstrap intervals; collision and timeout are mean percentages. The four primary differences use 98.75% intervals in the text. Latency is median batch-one CUDA planner time (ms), measured separately with ordinary, non-strict-deterministic inference.

System	Arrival [95% CI]	Coll.	Time.	ms
BC	62.23 [60.66, 63.82]	37.53	0.24	1.18
GMM-BC	65.28 [63.75, 66.77]	34.38	0.34	11.27
MLP DDPM (legacy)	55.22 [53.14, 57.25]	43.51	1.27	34.86
Role-blind DDPM	53.30 [51.01, 55.50]	45.98	0.72	34.83
Matched MLP DDPM	55.23 [53.38, 57.03]	43.36	1.41	35.47
Temporal U-Net DDPM	64.63 [63.03, 66.20]	35.32	0.05	109.57
Flow matching	60.75 [59.18, 62.29]	37.69	1.57	28.14
CEM 640/H24	63.76 [62.27, 65.27]	36.21	0.03	47.73
CEM 4096/H24	64.01 [62.50, 65.55]	35.98	0.01	76.47
CEM 4096/H48	15.02 [14.11, 15.90]	31.65	53.33	138.10
MPPI-style	45.06 [43.60, 46.56]	35.93	19.02	56.82
Action-box optimizer	64.33 [62.84, 65.84]	35.64	0.03	732.10
Random shooting	62.80 [61.26, 64.31]	36.07	1.13	10.14

These numbers are not pooled with new v05 outcomes. The earlier archive preserves all data-size, sampler, planning, OOD, physics, geometry, candidate and training-backend diagnostics with their original banks and intervals.

C Literature Map and Claim Boundaries

Table 5 links methodological choices to their sources and records the limits of that connection. This is a provenance map, not a list of reproduced baselines. The separate audit records the reading scope for each entry, including papers consulted only through abstracts or selected sections and books checked through publisher metadata.

Table 5: Sources supporting the design and the claims they do not establish for this study.

Topic and sources	Direct support in this study	Claims that require additional evidence or assumptions
Diffusion foundations (Ho et al., 2020; Nichol & Dhariwal, 2021; Song et al., 2021b)	Noise prediction, the cosine schedule, and the relation to continuous-time generative modeling.	A cosine schedule does not reproduce all improved-DDPM choices. Discrete action diffusion is not the complete score-SDE framework.
Diffusion planning (Janner et al., 2022; Carvalho et al., 2023)	Planning with learned trajectory or motion distributions; context for proposal generation and subsequent planning.	Different state/action spaces and optimization procedures preclude calling our action generator a reproduction or claiming planning novelty from diffusion alone.
Temporal conditioning (Chi et al., 2023; Perez et al., 2018)	Conditional action sequences, receding-horizon execution, temporal convolutions, and feature-wise affine modulation.	A low-dimensional temporal U-Net is not the full visuomotor Diffusion Policy; FiLM and time-structured denoising are established components.

Continued on the next page

Table 5 continued

Topic and sources	Direct support in this study	Claims that require additional evidence or assumptions
Sampling and guidance (Song et al., 2021a; Ho & Salimans, 2022)	Reduced-step DDIM sampling and the distinction between score guidance and input ablations.	Nulling one input is not classifier-free score mixing. Reduced NFE alone does not establish comparable quality or exact sampling.
Flow objectives (Lipman et al., 2023; Liu et al., 2023)	Conditional interpolation paths and velocity regression for an architecture-matched generative comparison.	Finite integration and endpoint clipping do not inherit exact transport results. No reflow, distillation, or optimal coupling is implied.
Mixture policies (Bishop, 1994; Mandlekar et al., 2022; robomimic core team, n.d.)	Conditional mixture densities, policy-network practice, and demonstration-quality sensitivity.	Mixture components, candidate counts, covariance parameterization, and temporal factorization still require explicit reporting and fair control.
Other multimodal policies (Florence et al., 2022; Shafiullah et al., 2022)	Energy-based and transformer-based alternatives establish that multimodal imitation is broader than diffusion.	These are related-work gaps unless implemented and evaluated. Our GMM is neither IBC nor a Behavior Transformer.
CEM variants (Rubinstein, 1999; Pinneri et al., 2021)	Sampling-distribution updates, memory, and temporally structured sampling motivate strong online-search controls.	Finite-budget search is not globally optimal. Warm start alone does not reproduce the full iCEM algorithm.
Path-integral MPC (Williams et al., 2017b;a)	Information-theoretic and importance-sampling perspectives motivate the MPPI-style comparator.	The two 2017 publications differ. Theory does not automatically transfer to changed noise, control costs, clipping, or sampling corrections.
Learned dynamics (Chua et al., 2018)	Probabilistic ensembles and uncertainty propagation illustrate an additional model-based baseline class.	Known-dynamics CEM is not PETS. Ensemble learning and its data-collection and calibration protocol are not supplied by the optimizer alone.
MPC guarantees (Mayne et al., 2000; 2005)	Stability, terminal conditions, feasibility, and bounded-disturbance assumptions frame the limits of receding-horizon control.	Replanning alone does not prove stability or recursive feasibility. Linear bounded-disturbance results do not cover arbitrary nonlinear mismatch or Gaussian disturbances.
Closest specialized planners (Römer et al., 2025; Puthumanaillam et al., 2025)	Constrained diffusion predictive control and belief-conditioned fast planning delimit the neighboring problem space.	Finite candidate ranking is not complete DPCC. A fixed synthetic role label is not a belief estimate or one-step distillation.
Safety and physics constraints (Xiao et al., 2025; Bastek et al., 2025; Ames et al., 2017; Wabersich & Zeilinger, 2021)	Diffusion-time invariance, physics residuals, barrier conditions, and predictive safe-backup mechanisms distinguish several constraint strategies.	Nominal checking or a penalty is not certified safety. Continuous-time execution, model uncertainty, feasibility, and terminal backup conditions need separate treatment.
Marine dynamics and integration (Fossen, 2021; n.d.; Butcher, 2016)	Marine model structure and standard numerical integration motivate the stipulated dynamics and integration-accuracy study.	No identified vessel model is supplied. World-frame translation is not full relative-flow hydrodynamics; RK4 accuracy is not physical validity or collision safety.

Continued on the next page

Table 5 continued

Topic and sources	Direct support in this study	Claims that require additional evidence or assumptions
Rules and external data (United States Coast Guard, 2024; Yao et al., 2024)	Navigation rules provide contextual obligations; WaterScenes supplies a water-surface perception benchmark.	Descriptive role proxies are not legal compliance. A perception dataset does not supply this study's closed-loop control benchmark.
Statistical reliability (Agarwal et al., 2021; Henderson et al., 2018; Owen, 2007; Holm, 1979)	Training variability, separate-factor resampling, and predeclared multiplicity control motivate transparent uncertainty reporting.	Many scenarios do not replace training replicates. Ten main seeds still limit precision; the archived data-scale study uses five. Holm's sequential procedure was not applied; the v04 four-comparison and v05 six-comparison families use separate Bonferroni-adjusted intervals.
Conditioning and interpretation (de Haan et al., 2019; D'Amour et al., 2022; Geirhos et al., 2020)	General motivation for interventions, testing beyond one heldout score, and considering shortcut behavior.	Only abstracts and primary metadata were consulted for these additions. They do not validate our explicit-role intervention as complete causal identification or establish role-free learning.

Our role-stratified, paired seed–scenario bootstrap is a study-specific adaptation of crossed-factor resampling. Reported descriptive intervals should not be read as simultaneous confidence statements. Similarly, an information-privileged finite-budget controller is a diagnostic comparator, not a certified performance upper bound. Earlier negative results remain part of the interpretation of additional architectures and experimental rounds.

D Expanded mechanism detail

Figure 7 separates the training masks, temporal architecture and inference interfaces. Generator-code and selector-objective interventions act at distinct points; the execution and observation grids change continuous quantities while the clean oracle initial role remains fixed. The expanded diagram is structural and does not report measured performance.

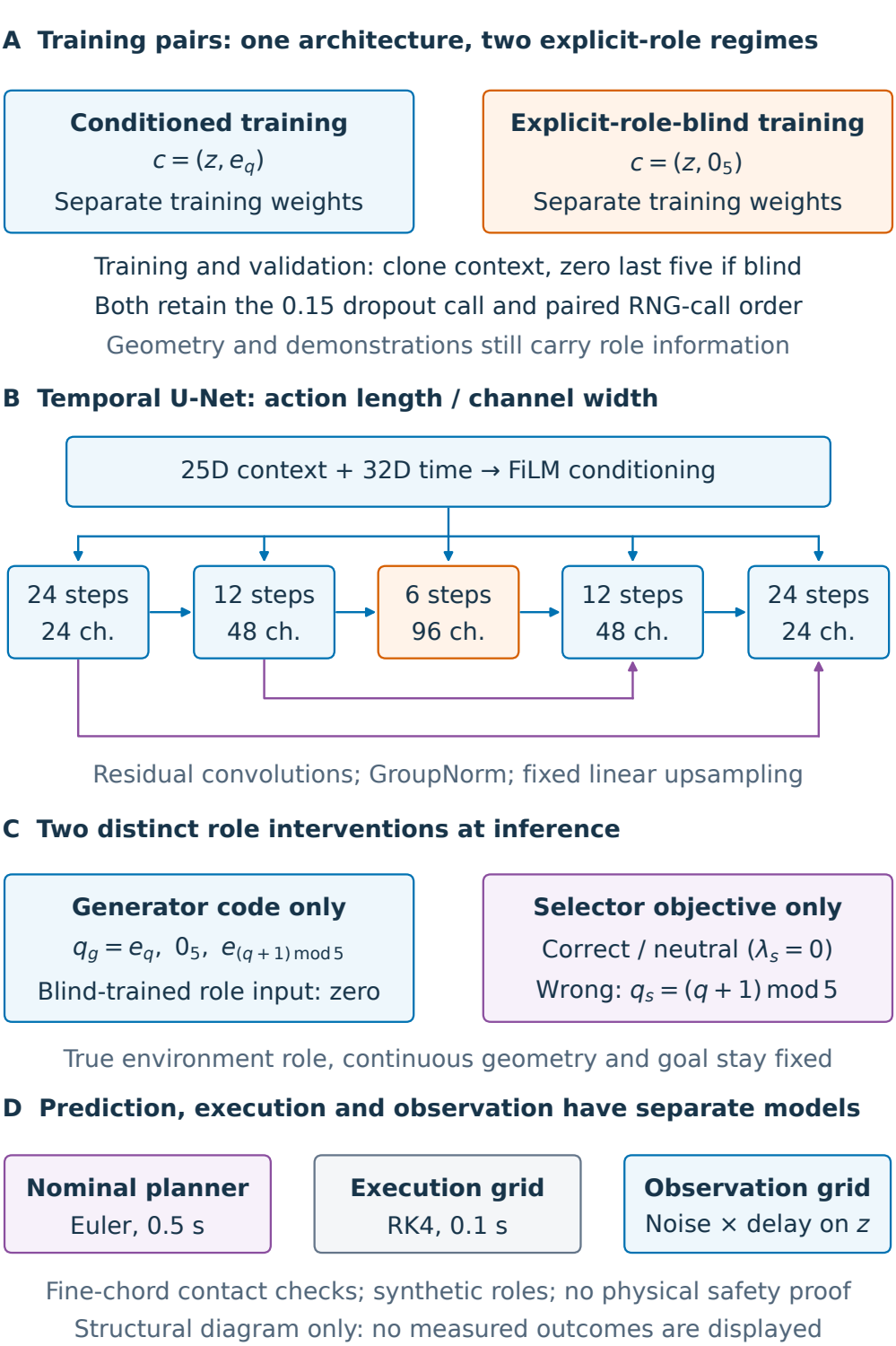


Figure 7: Separate training masks, temporal U-Net structure, independent generator/selector role interventions, and the distinction between nominal prediction, execution and observation. Only continuous state/target observations are perturbed; the oracle initial role remains fixed and clean. Native width is 5.5 inches. This is structural evidence, not a measured outcome plot.

Table 6: Completed v05 accounting after full source/checkpoint, saved-bank and raw-episode audits. Thirty new blind trainings and thirty preserved conditioned trainings are distinct; smoke runs are excluded.

Block	Jobs	Settings	Executions
main	180	18	360,000
checkpoint last	120	12	240,000
factorial	840	84	252,000
perception grid	1,080	108	324,000
execution grid	1,440	144	432,000

Table 7: Six prespecified conditioned-minus-explicit-role-blind arrival differences (percentage points). Intervals have 99.1667% confidence; zero inclusion is not equivalence.

Architecture	Selector	Difference [99.1667% CI]
MLP	correct	$+1.76[0.10, 3.48]$
MLP	neutral	$+0.17[-2.11, 2.19]$
U-Net	correct	$-0.13[-0.69, 0.36]$
U-Net	neutral	$-0.28[-1.02, 0.47]$
flow	correct	$+0.94[-0.07, 2.08]$
flow	neutral	$-0.37[-1.53, 0.85]$

Table 8: All twelve fixed-last minus validation-best arrival differences (percentage points, descriptive 95% intervals). The last checkpoint is fixed at update 30,000. Labels follow Table 3.

Setting	Last minus best [95% CI]
MLP C/c	$-0.05[-0.30, 0.11]$
MLP C/n	$+0.19[0.00, 0.65]$
MLP B/c	$+0.17[-0.48, 0.96]$
MLP B/n	$+0.05[-0.66, 0.81]$
U-Net C/c	$+0.09[-0.31, 0.51]$
U-Net C/n	$-0.23[-1.12, 0.60]$
U-Net B/c	$-0.05[-0.37, 0.27]$
U-Net B/n	$-0.45[-1.25, 0.26]$
flow C/c	$-0.13[-0.53, 0.10]$
flow C/n	$-0.13[-0.63, 0.22]$
flow B/c	$-0.19[-0.70, 0.16]$
flow B/n	$-0.16[-0.53, 0.06]$

E COMPLETE NEW NUMERICAL EVIDENCE

The machine-readable archive preserves all 366 settings, 3,660 seed runs, five role strata and three terminal outcomes. The tables above and complete figure gallery do not convert descriptive intervals into a larger confirmatory family. The same 10,000 crossed-bootstrap draws bind marginal outcomes, paired contrasts and stress interactions; the only adjusted family is the six prespecified main arrival contrasts.

Saved-event audit scope. Of 1,176,000 nominal-evaluator episodes, 595,215 have verified saved-chord event-label agreement, 580,785 are compatible only within the separately recorded float32 rounding envelope, and 0 remain unresolved. The physics audit instead checks 432,000 episodes and 121,328,690 executed fine intervals using its frozen coordinate/control tolerances; the nominal terminal-category counter does not apply there. Neither audit changes outcomes or replays the complete policy/ODE. Passing is saved-evidence consistency under the stated numerical allowances, not tolerance-free event reconstruction or a continuous-physics certificate.

Table 9: Main arrival percentages by training/planner seed. Each cell evaluates 2,000 shared scenes; seeds are repetition blocks. One decimal is used here; full-precision rows remain in the data archive.

Setting	100	101	102	103	104	105	106	107	108	109
MLP C/c	57.3	55.6	58.2	55.5	58.2	57.6	57.2	53.1	56.4	59.4
MLP C/n	59.6	55.1	63.2	57.6	61.2	59.2	58.8	52.3	59.6	61.9
MLP B/c	53.7	55.9	58.1	53.5	58.8	57.3	54.4	49.9	54.8	54.6
MLP B/n	57.0	60.1	62.8	56.4	63.2	59.6	58.1	52.0	58.8	58.7
U-Net C/c	65.8	65.8	66.5	66.2	65.3	67.3	66.1	65.5	66.1	66.9
U-Net C/n	71.7	69.6	72.1	71.5	69.2	72.4	70.8	69.6	71.2	72.2
U-Net B/c	66.0	65.7	66.4	66.4	66.4	66.9	65.7	65.7	66.0	67.4
U-Net B/n	71.9	70.4	72.4	71.1	70.8	71.8	71.0	70.7	70.2	72.7
flow C/c	63.1	62.0	62.1	62.6	61.8	61.9	62.6	62.0	61.9	62.3
flow C/n	68.1	65.8	68.4	66.8	67.6	67.6	66.8	67.6	66.0	66.8
flow B/c	61.6	60.8	61.8	62.2	61.0	60.6	59.3	62.1	61.5	62.0
flow B/n	67.2	65.8	69.1	68.0	67.6	67.8	64.9	67.9	68.2	68.6
BC/c	64.7	64.2	63.9	64.4	63.9	63.8	63.3	65.1	63.9	63.8
GMM/c	66.5	65.9	66.4	66.7	66.6	66.9	66.7	66.7	66.6	66.8
GMM/n	69.7	68.3	69.5	69.6	69.7	69.5	68.8	69.9	69.6	69.5
CEM640/c	64.9	65.6	65.3	65.4	65.2	65.4	65.9	65.4	65.6	64.7
CEM640/n	73.8	73.5	73.6	73.4	73.7	73.3	73.4	73.5	73.5	73.1
CEM4096/c	65.9	65.9	65.6	65.8	65.8	65.7	65.7	65.7	65.9	65.7

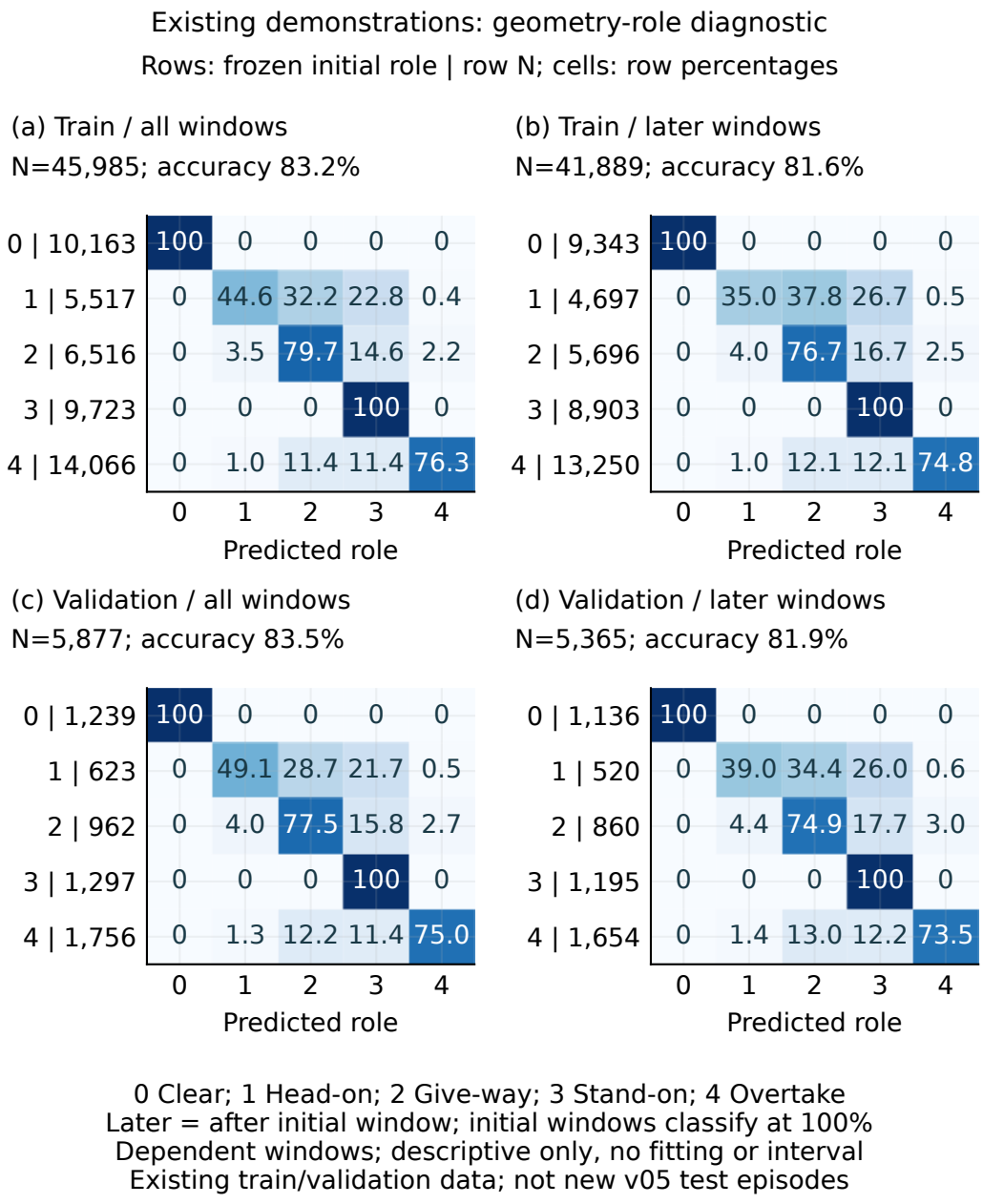


Figure 8: Existing-demonstration diagnostic, separate from new v05 evaluation: the same fixed, unfitted geometry rule predicts the stored initial role for training and validation, using all or only later teacher decision windows. Cells are row-normalized confusion proportions; row labels retain role-specific denominators. Aggregate accuracy weights dependent windows by their observed frequency, not equally by role or scene. Initial-state recovery does not establish redundancy throughout rollout. This diagnostic rule is not an optimal classifier or an information-theoretic bound; its errors do not establish that role is unrecoverable from geometry. No window-independent confidence intervals are implied.

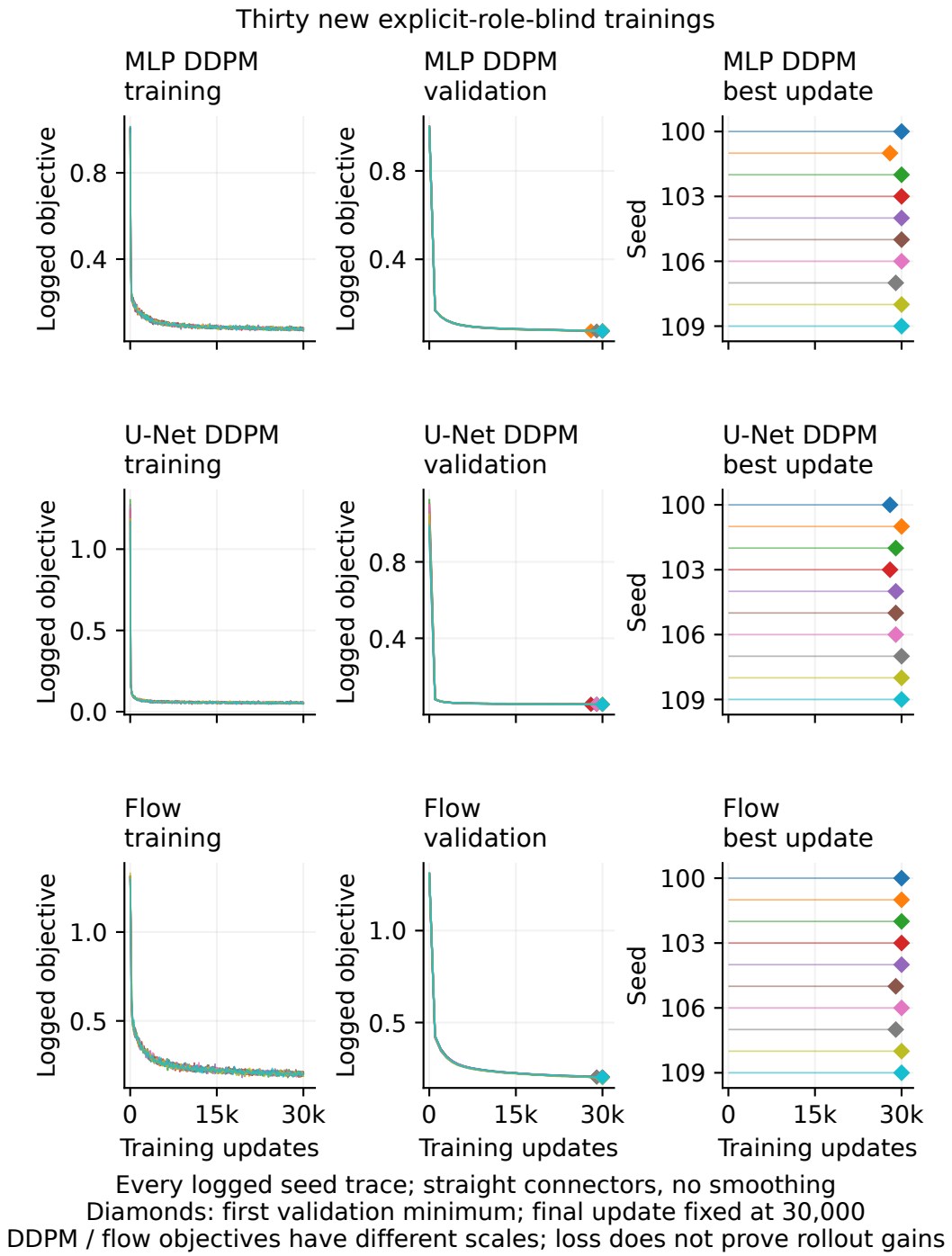


Figure 9: All thirty new explicit-role-blind training runs: unsmoothed recorded training and validation losses and validation-selected checkpoint updates. Loss scales differ across architectures and conditioning regimes; these curves are optimization diagnostics, not cross-model performance or training-compute comparisons.

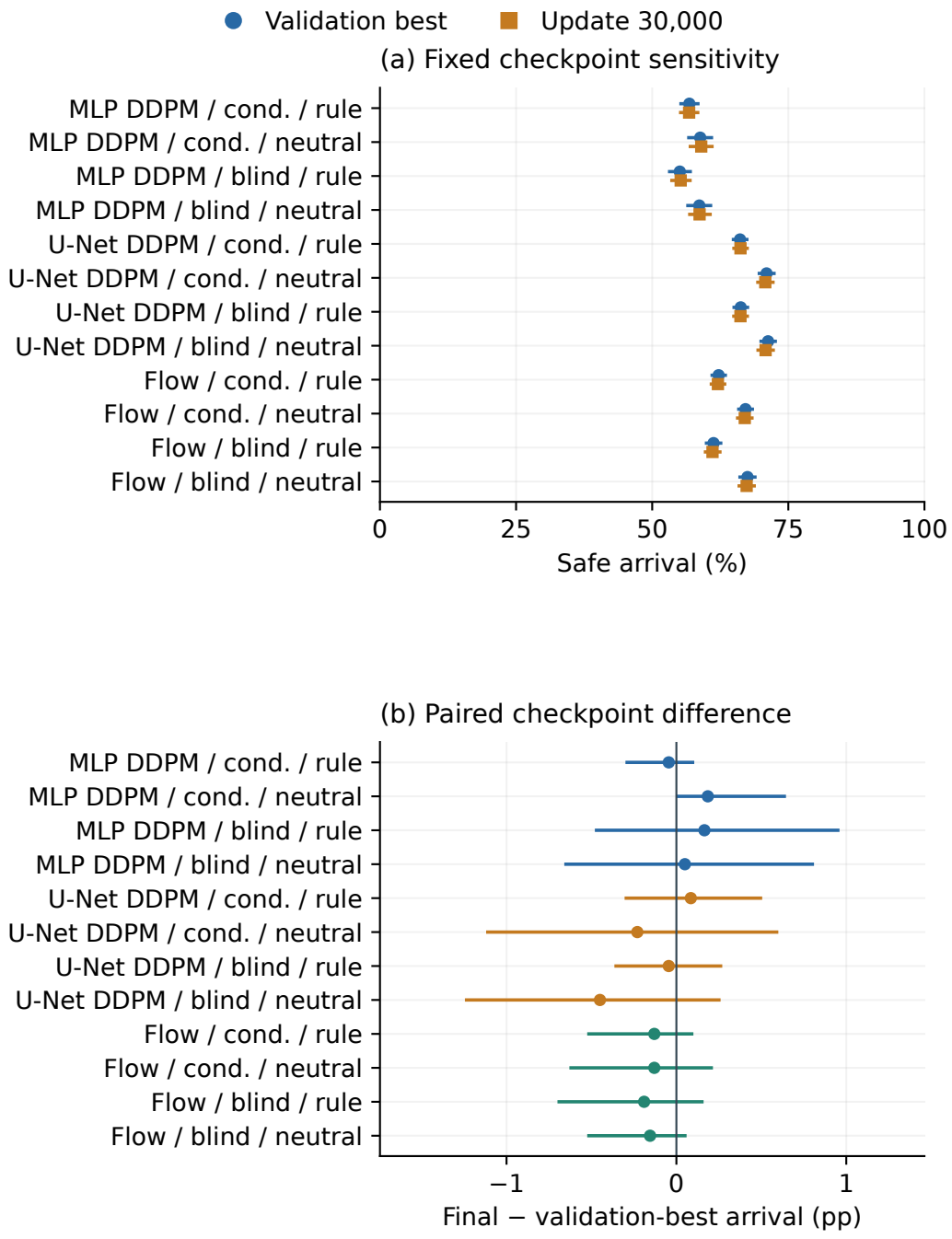


Figure 10: Validation-best and fixed-last checkpoints for all twelve generative settings on the same main bank. Marginal and paired-difference intervals are descriptive 95% crossed-bootstrap intervals. The fixed-last checkpoint uses update 30,000 and is never selected using test outcomes.

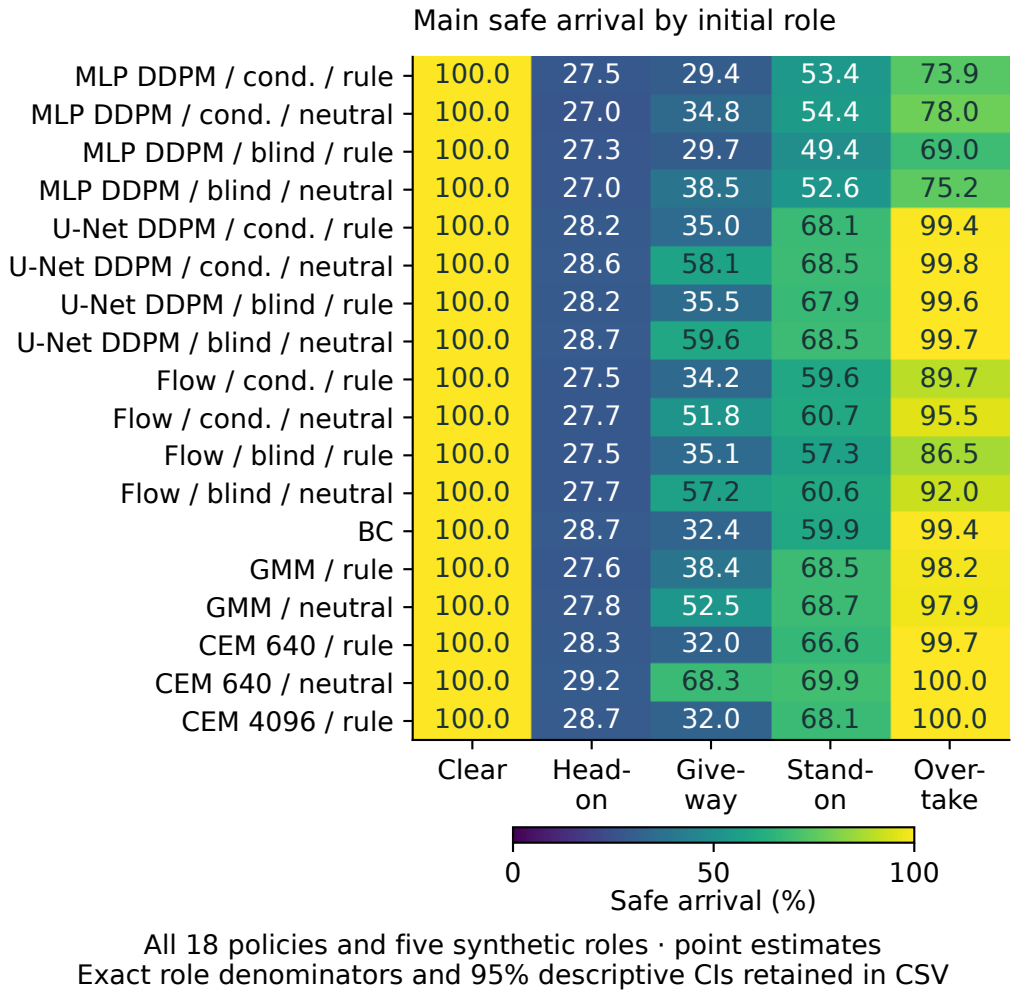


Figure 11: Arrival point estimates for all eighteen main settings and five initial synthetic roles. Exact role-specific denominators and descriptive 95% intervals are preserved in the source CSV. Cell colors and numbers show point estimates, without an uncertainty layer.

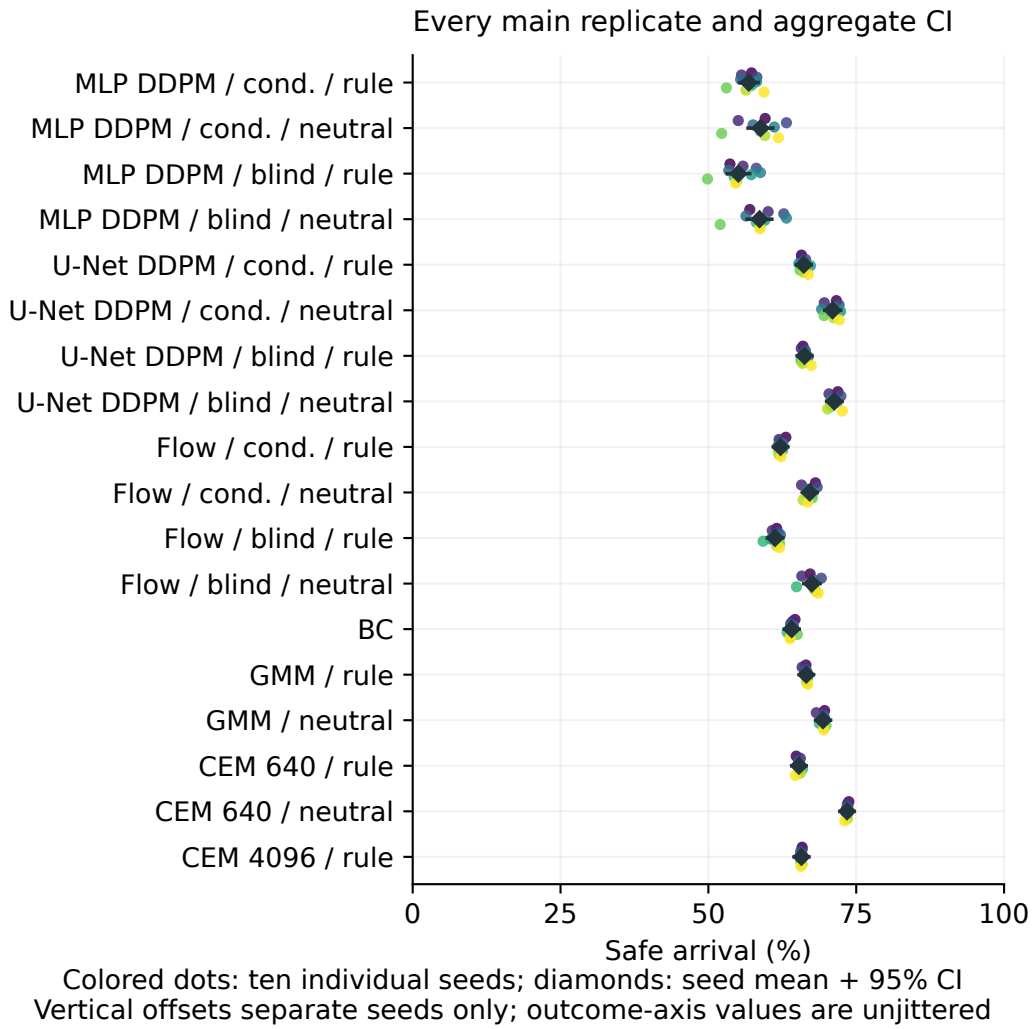


Figure 12: All ten training/planner seed means for each of the eighteen main settings. Diamonds and intervals are aggregate means with descriptive 95% crossed-bootstrap intervals. Vertical offsets separate seeds visually; arrival-axis values are not jittered.

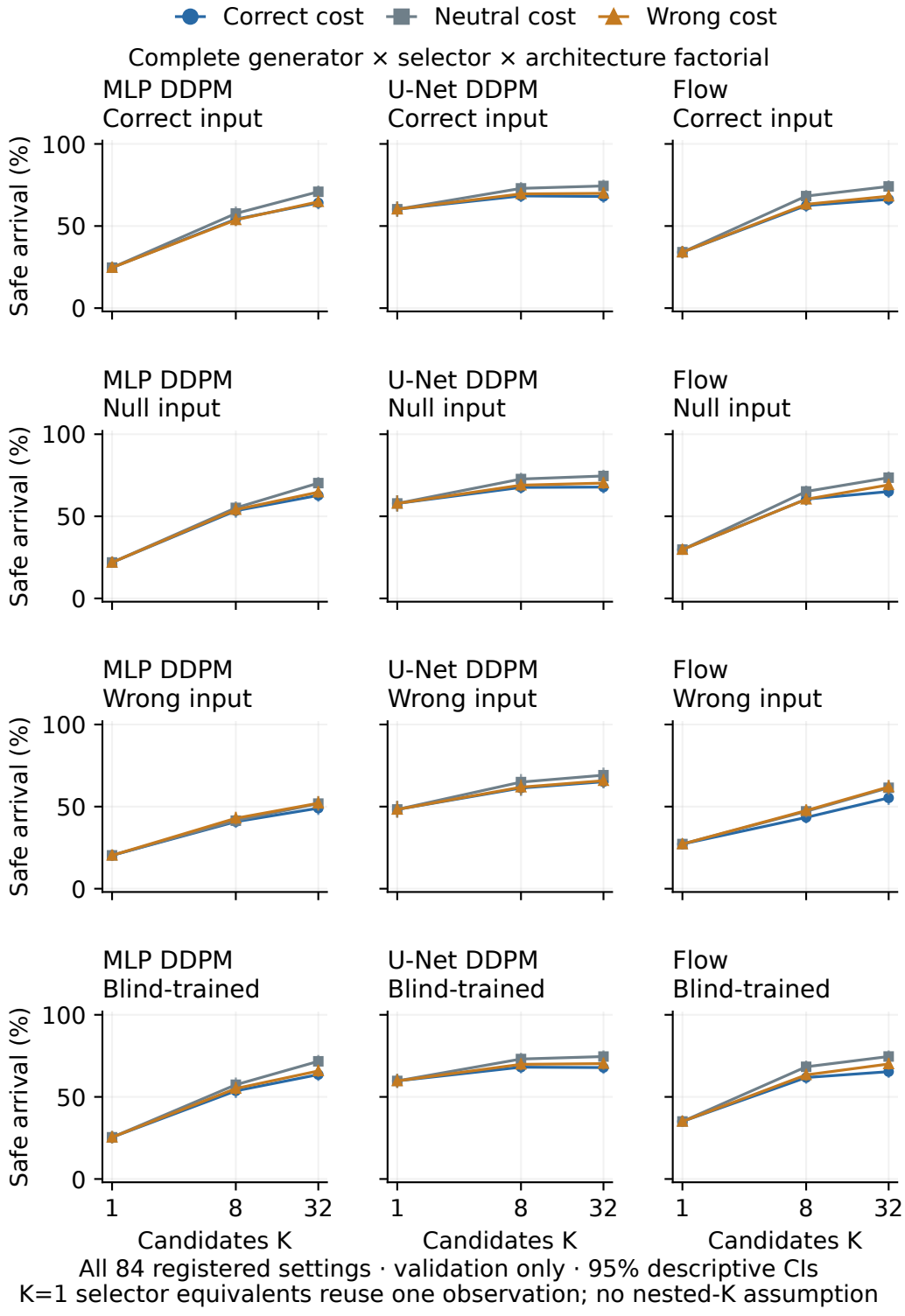


Figure 13: Safe arrival across the complete 84-setting validation factorial: three architectures, four generator conditions, three selectors and $K = 1, 8, 32$. Intervals are descriptive 95%. At $K = 1$ the sole executed setting is displayed for equivalent selector paths; repeated display points add no observations.

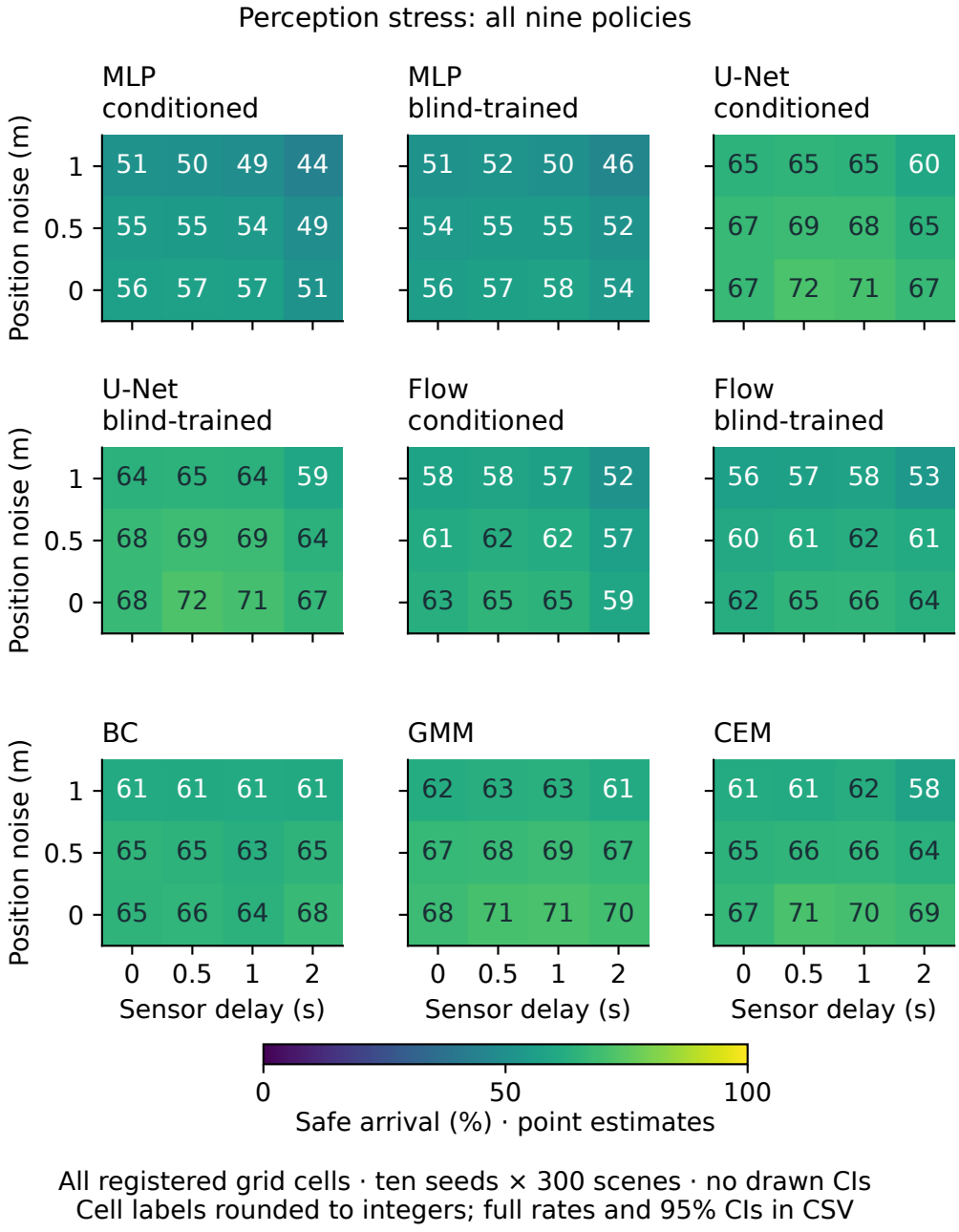


Figure 14: Safe arrival percentages for all nine policies over the full position-noise and sensor-delay grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

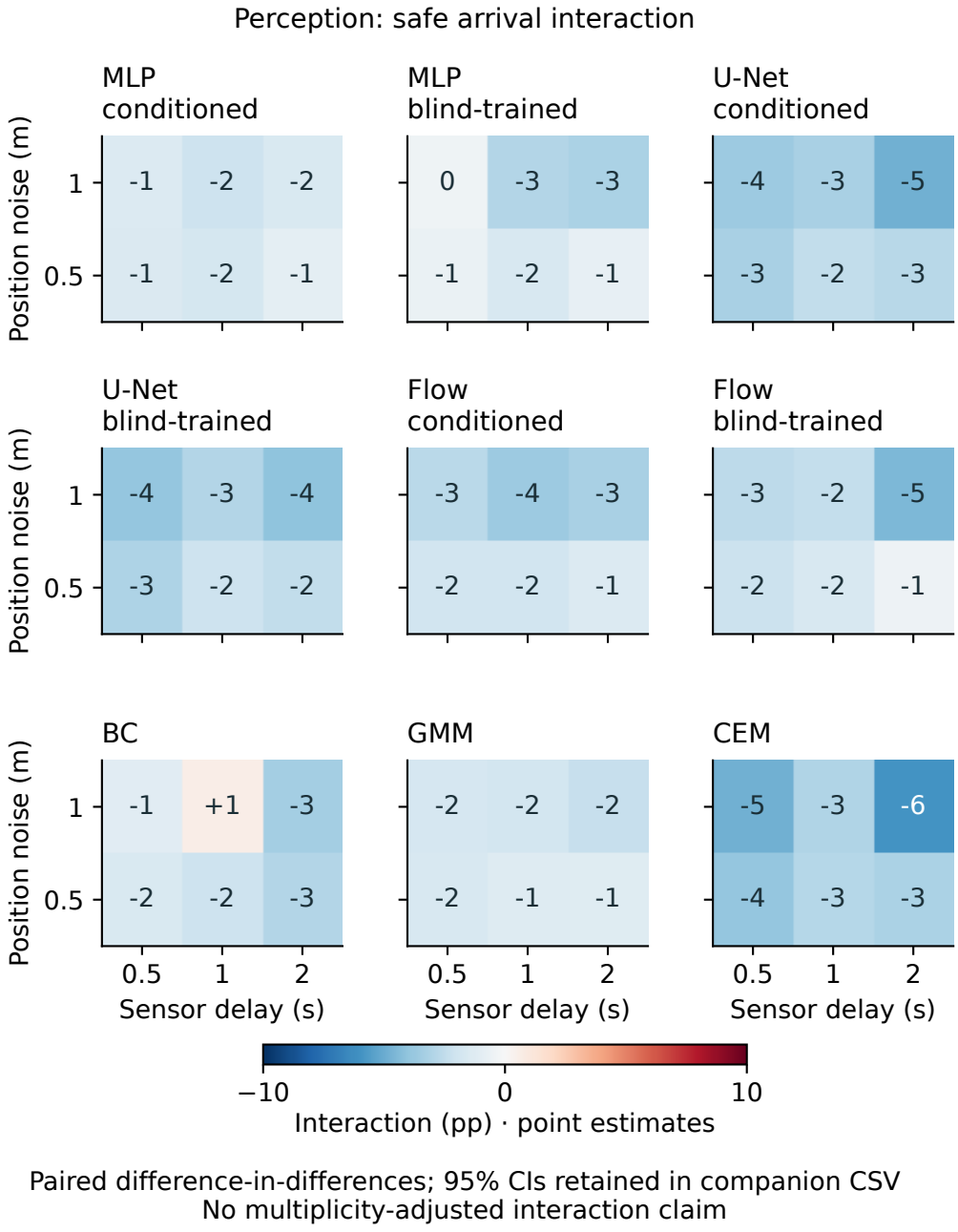


Figure 15: Safe arrival departure from additivity, in percentage points, for all nine policies over the position-noise and sensor-delay grid: $f(x, y) - f(x, 0) - f(0, y) + f(0, 0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.

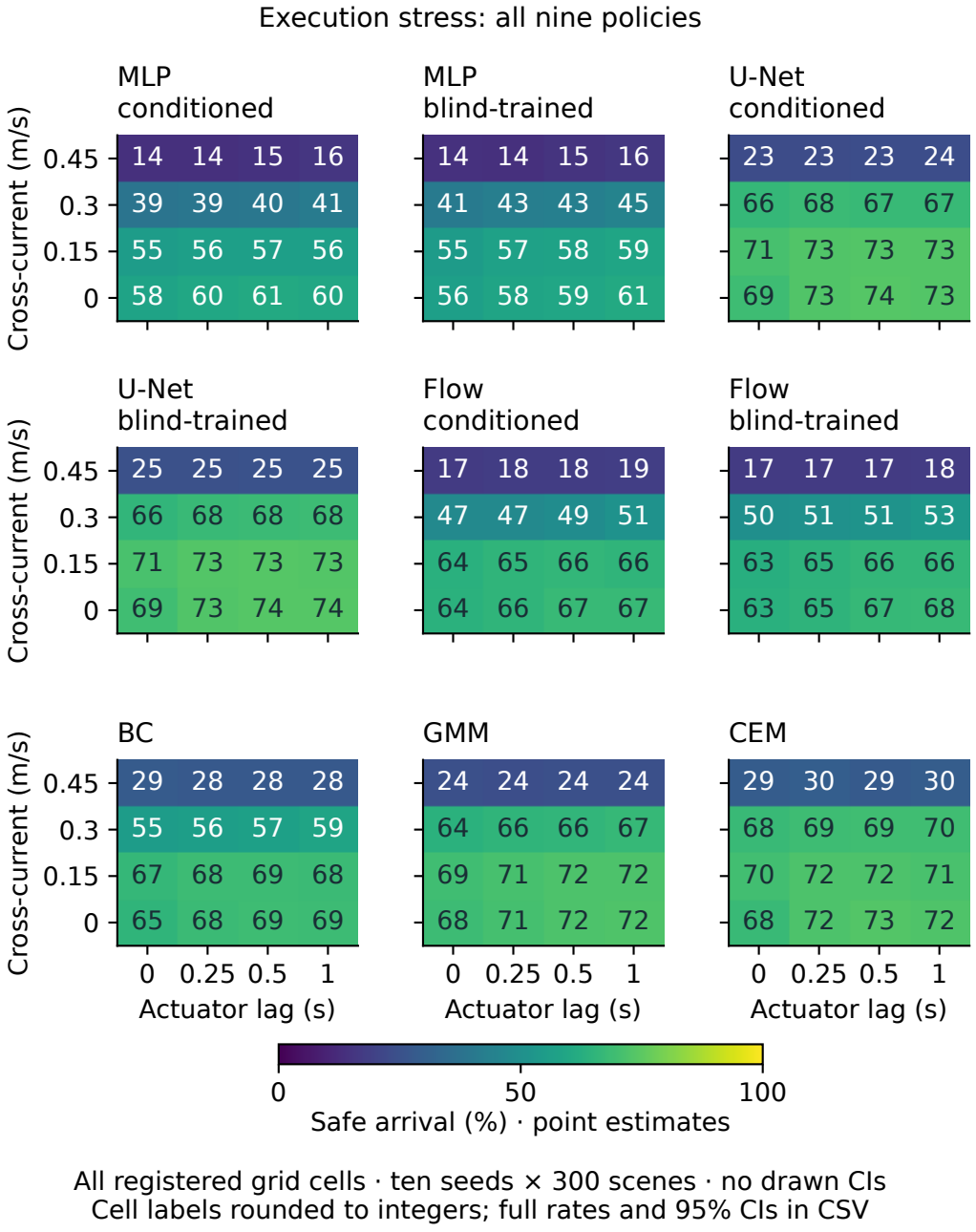


Figure 16: Safe arrival percentages for all nine policies over the full current and actuator-lag grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

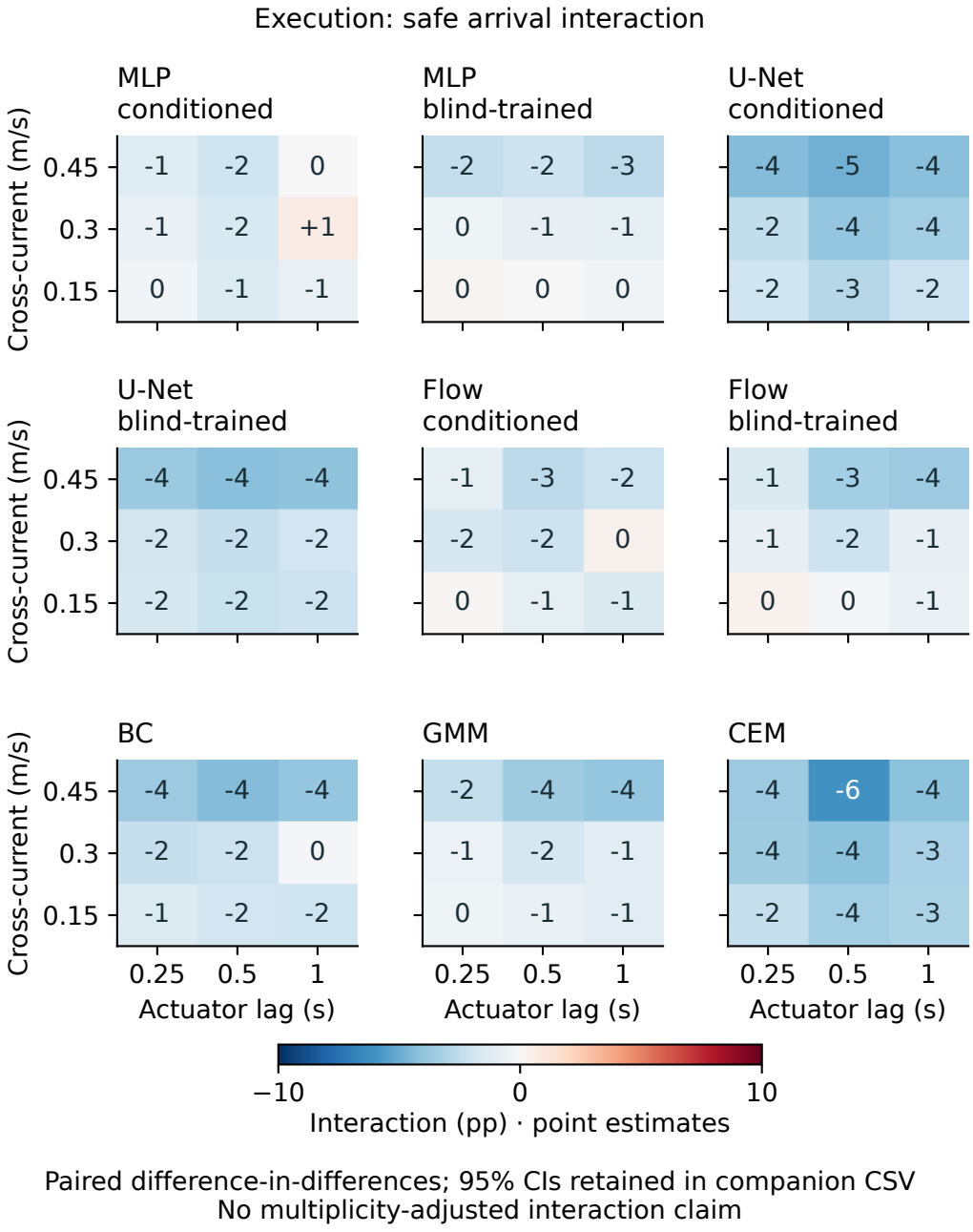


Figure 17: Safe arrival departure from additivity, in percentage points, for all nine policies over the current and actuator-lag grid: $f(x, y) - f(x, 0) - f(0, y) + f(0, 0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.

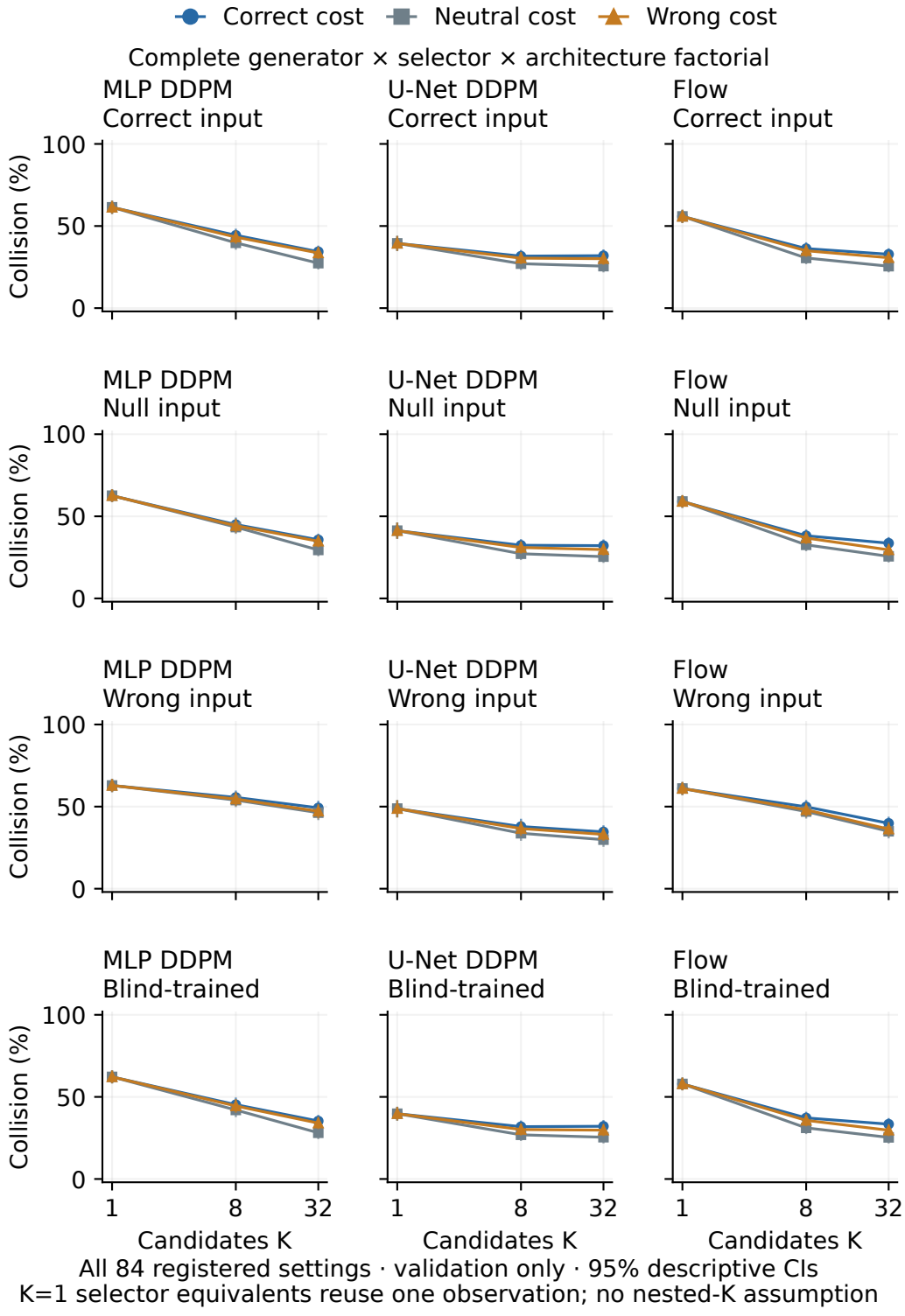


Figure 18: Collision across the complete 84-setting validation factorial: three architectures, four generator conditions, three selectors and $K = 1, 8, 32$. Intervals are descriptive 95%. At $K = 1$ the sole executed setting is displayed for equivalent selector paths; repeated display points add no observations.

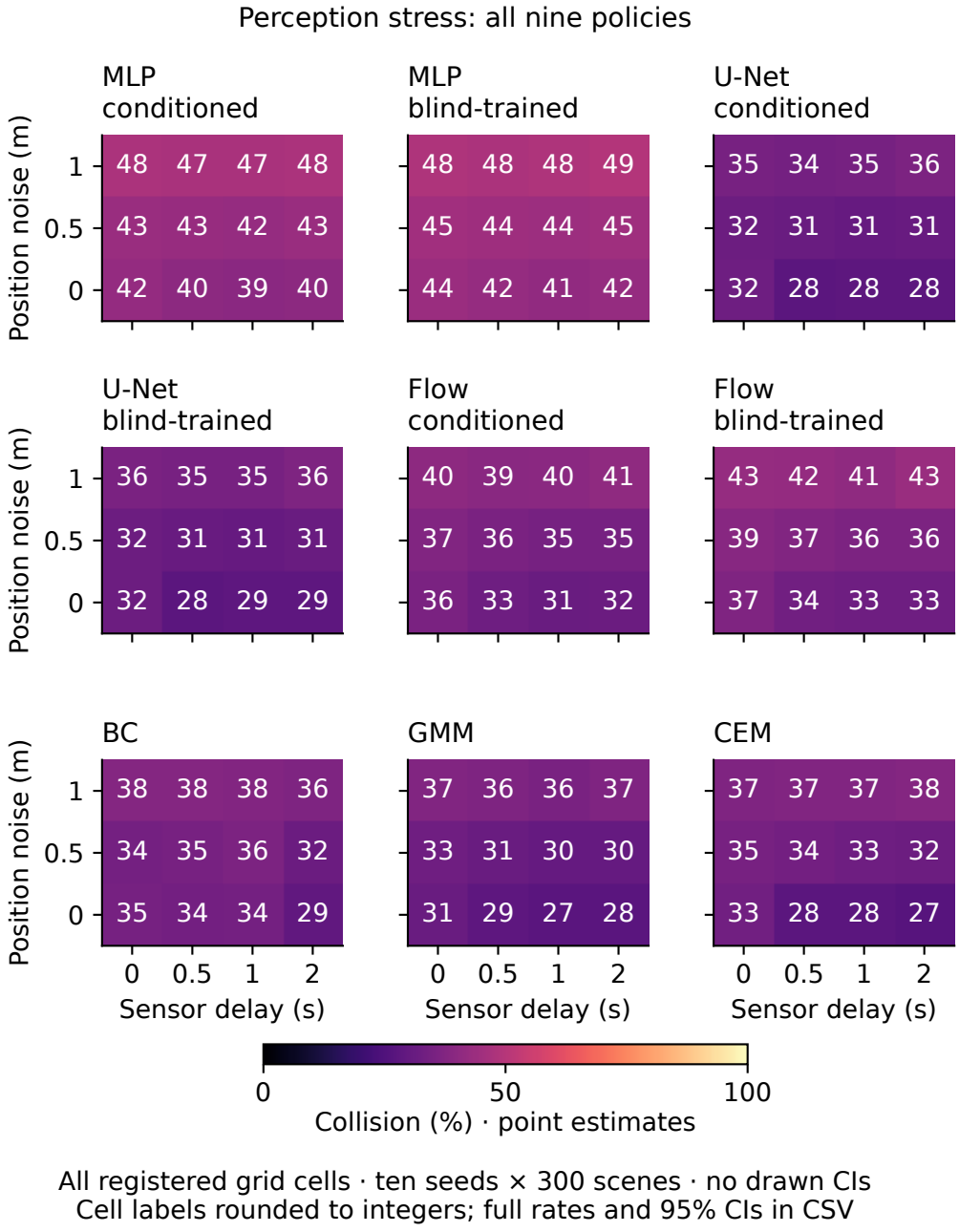


Figure 19: Collision percentages for all nine policies over the full position-noise and sensor-delay grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

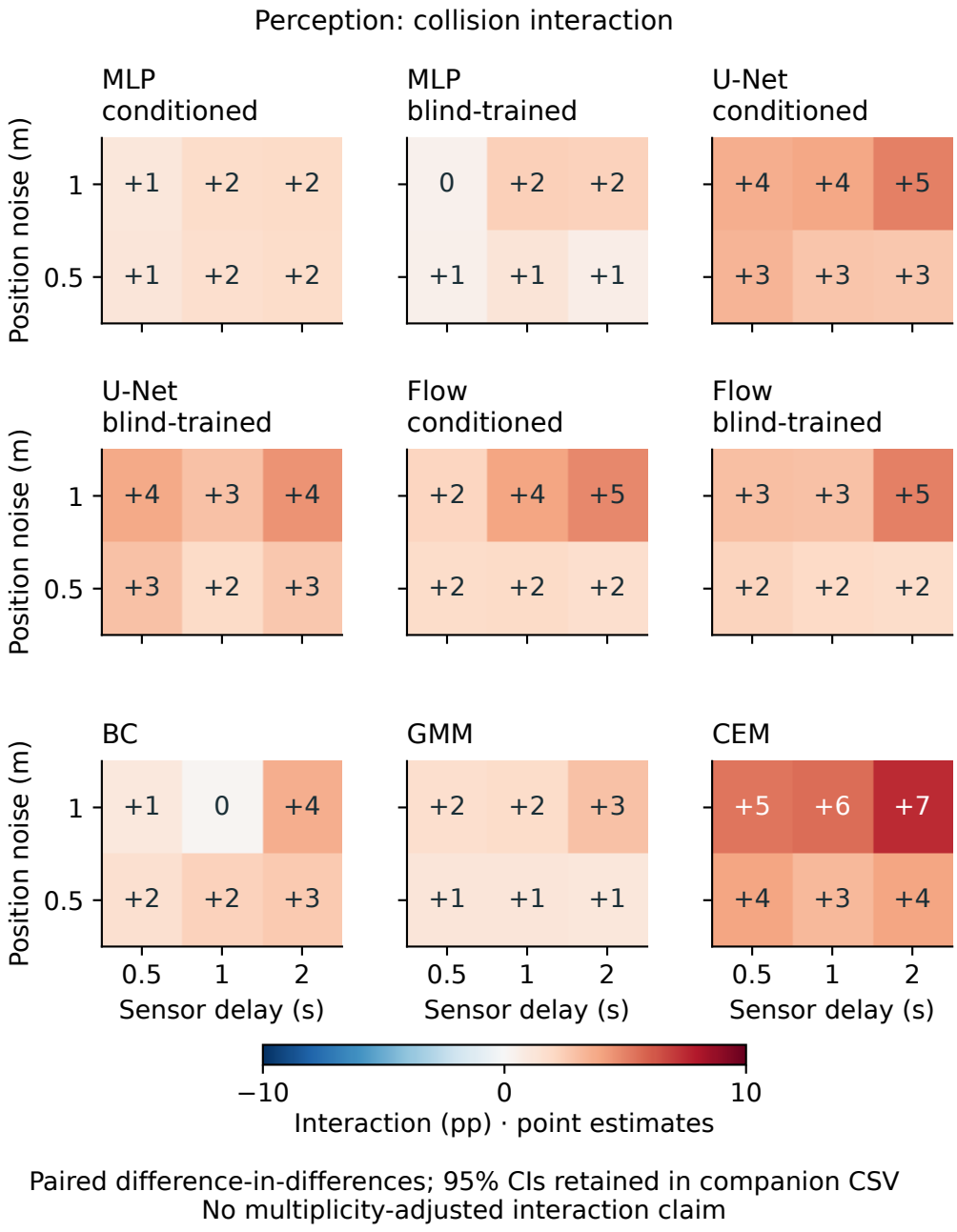


Figure 20: Collision departure from additivity, in percentage points, for all nine policies over the position-noise and sensor-delay grid: $f(x,y)-f(x,0)-f(0,y)+f(0,0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.

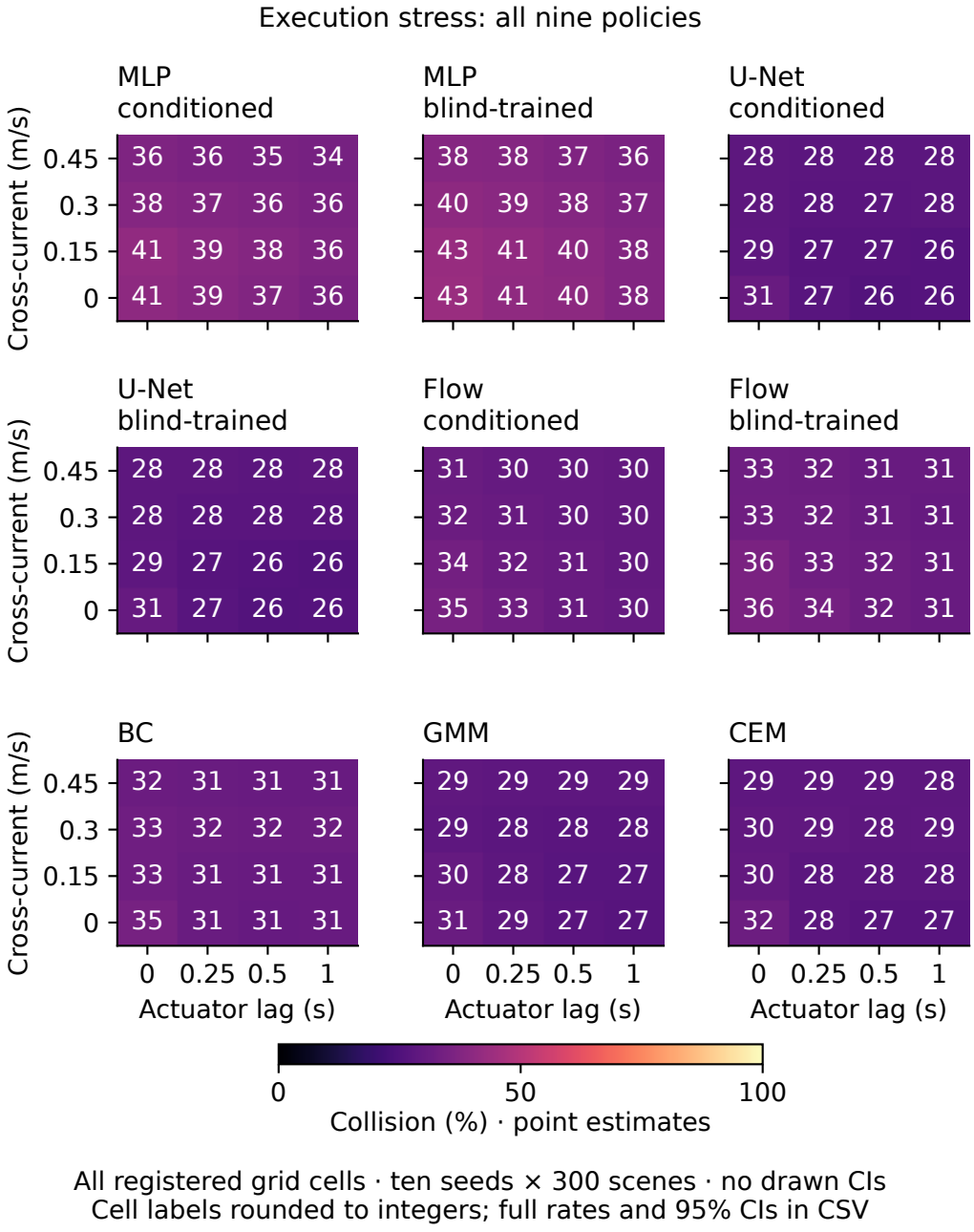


Figure 21: Collision percentages for all nine policies over the full current and actuator-lag grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

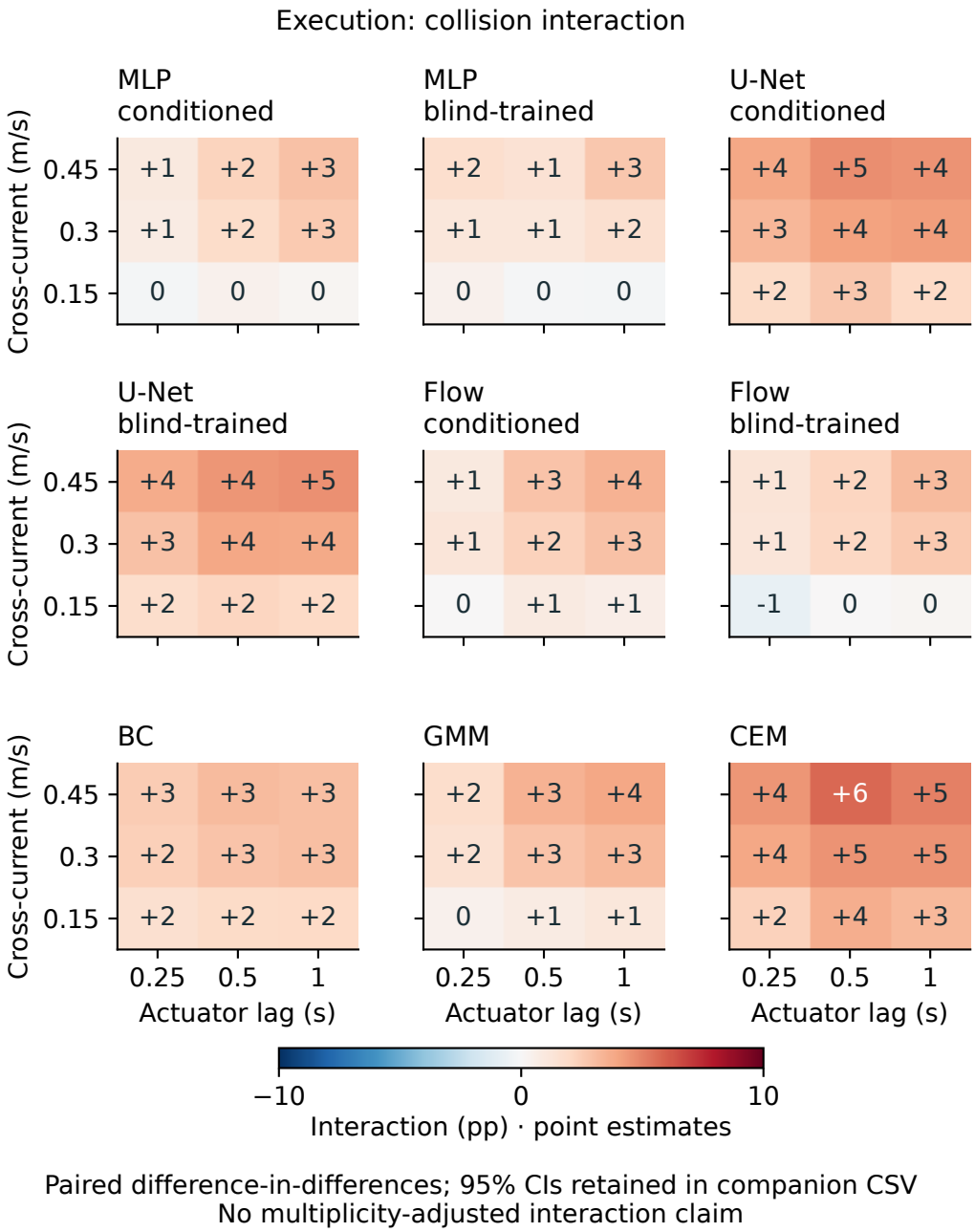


Figure 22: Collision departure from additivity, in percentage points, for all nine policies over the current and actuator-lag grid: $f(x, y) - f(x, 0) - f(0, y) + f(0, 0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.

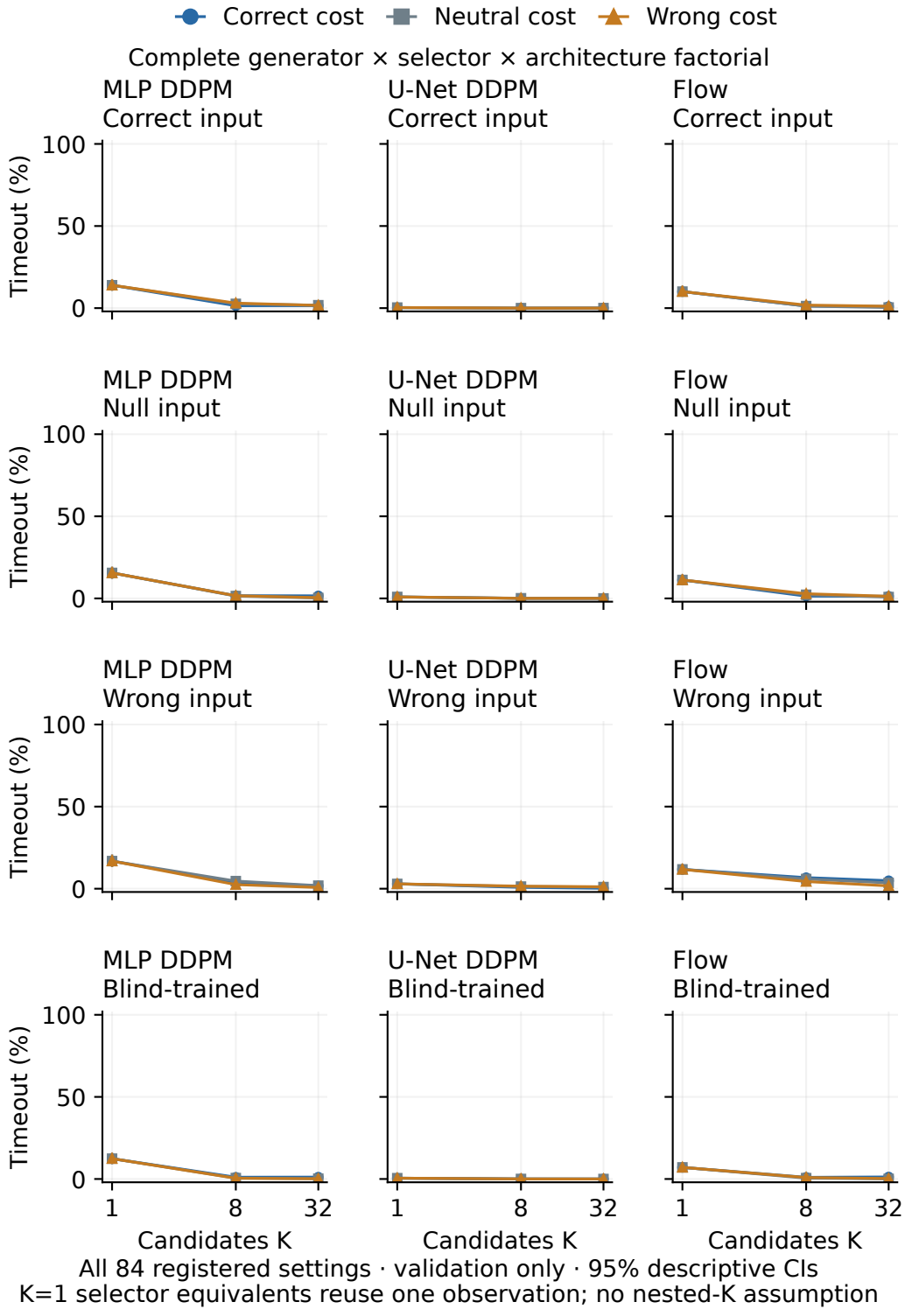


Figure 23: Timeout across the complete 84-setting validation factorial: three architectures, four generator conditions, three selectors and $K = 1, 8, 32$. Intervals are descriptive 95%. At $K = 1$ the sole executed setting is displayed for equivalent selector paths; repeated display points add no observations.

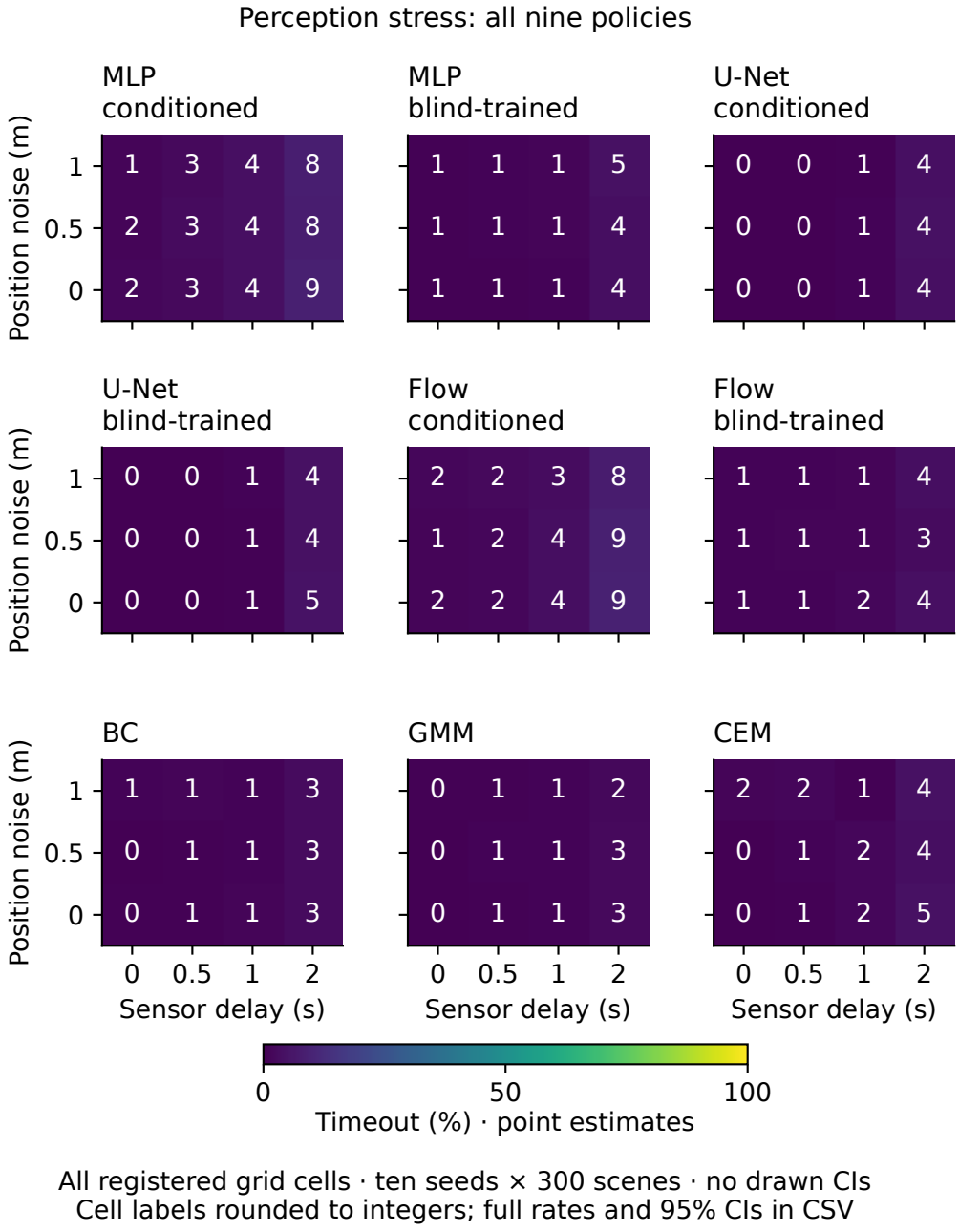


Figure 24: Timeout percentages for all nine policies over the full position-noise and sensor-delay grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

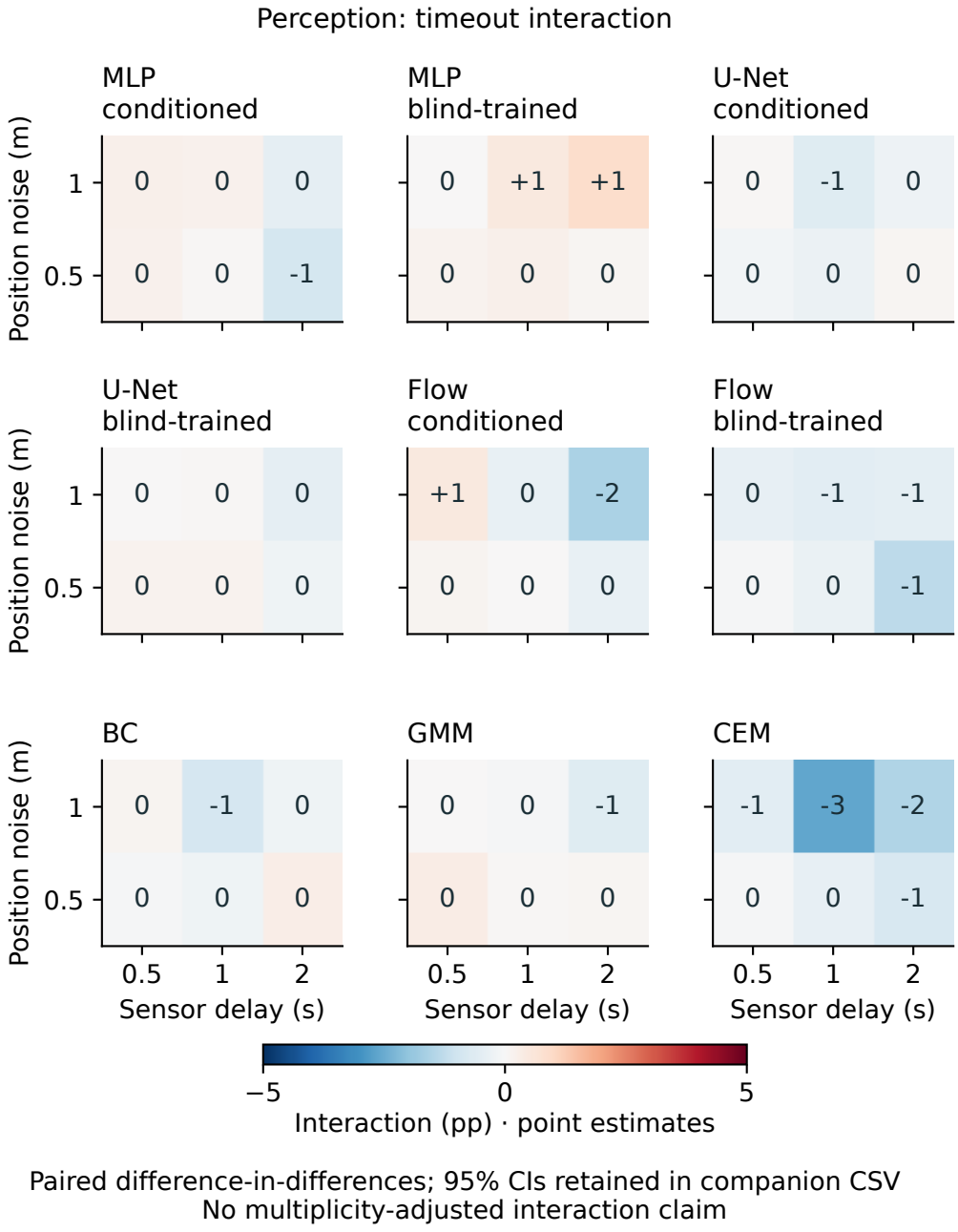


Figure 25: Timeout departure from additivity, in percentage points, for all nine policies over the position-noise and sensor-delay grid: $f(x,y)-f(x,0)-f(0,y)+f(0,0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.

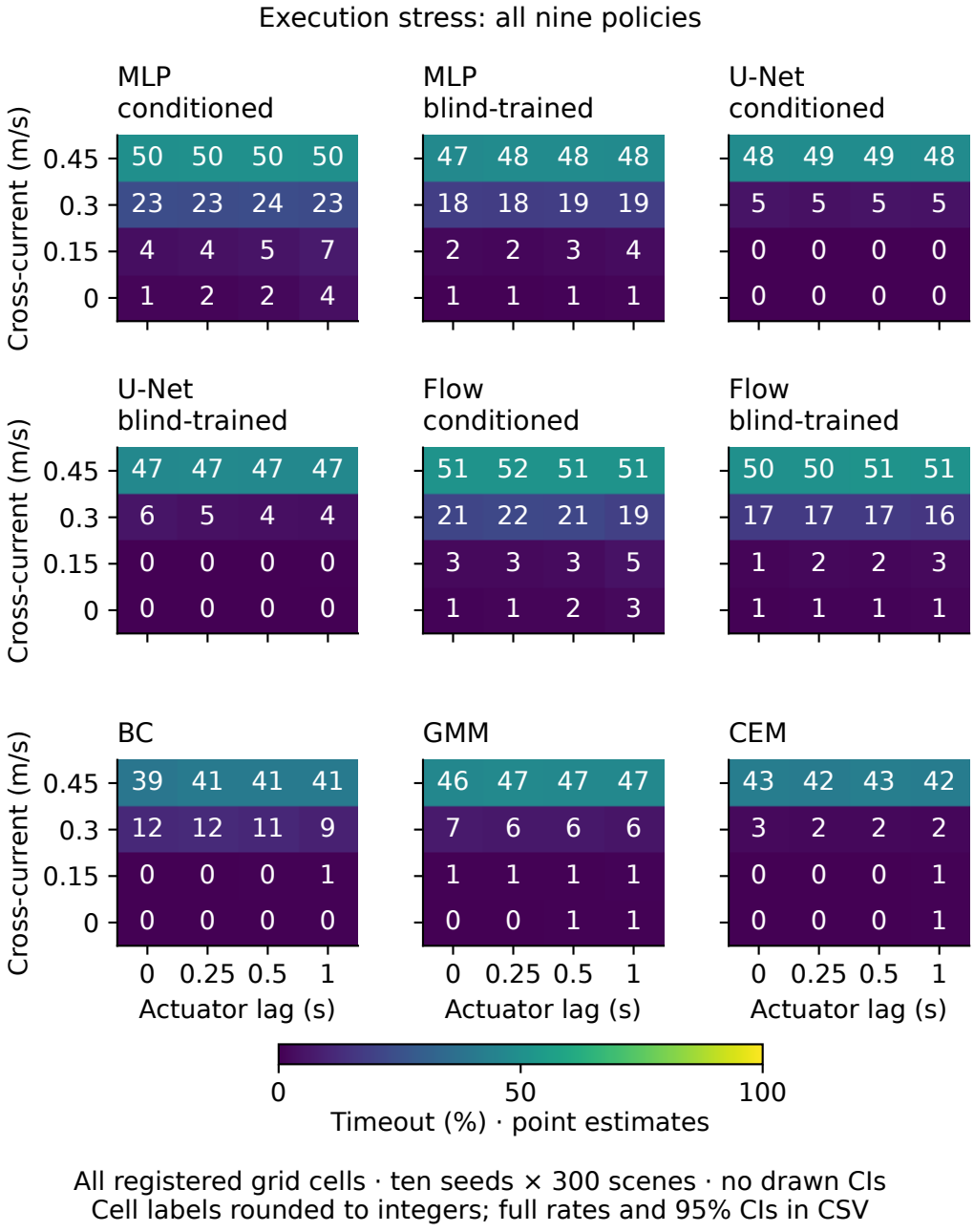


Figure 26: Timeout percentages for all nine policies over the full current and actuator-lag grid. Heat maps show point estimates; exact descriptive 95% intervals and all denominators remain in the source CSV. Each grid has its own matched scene bank.

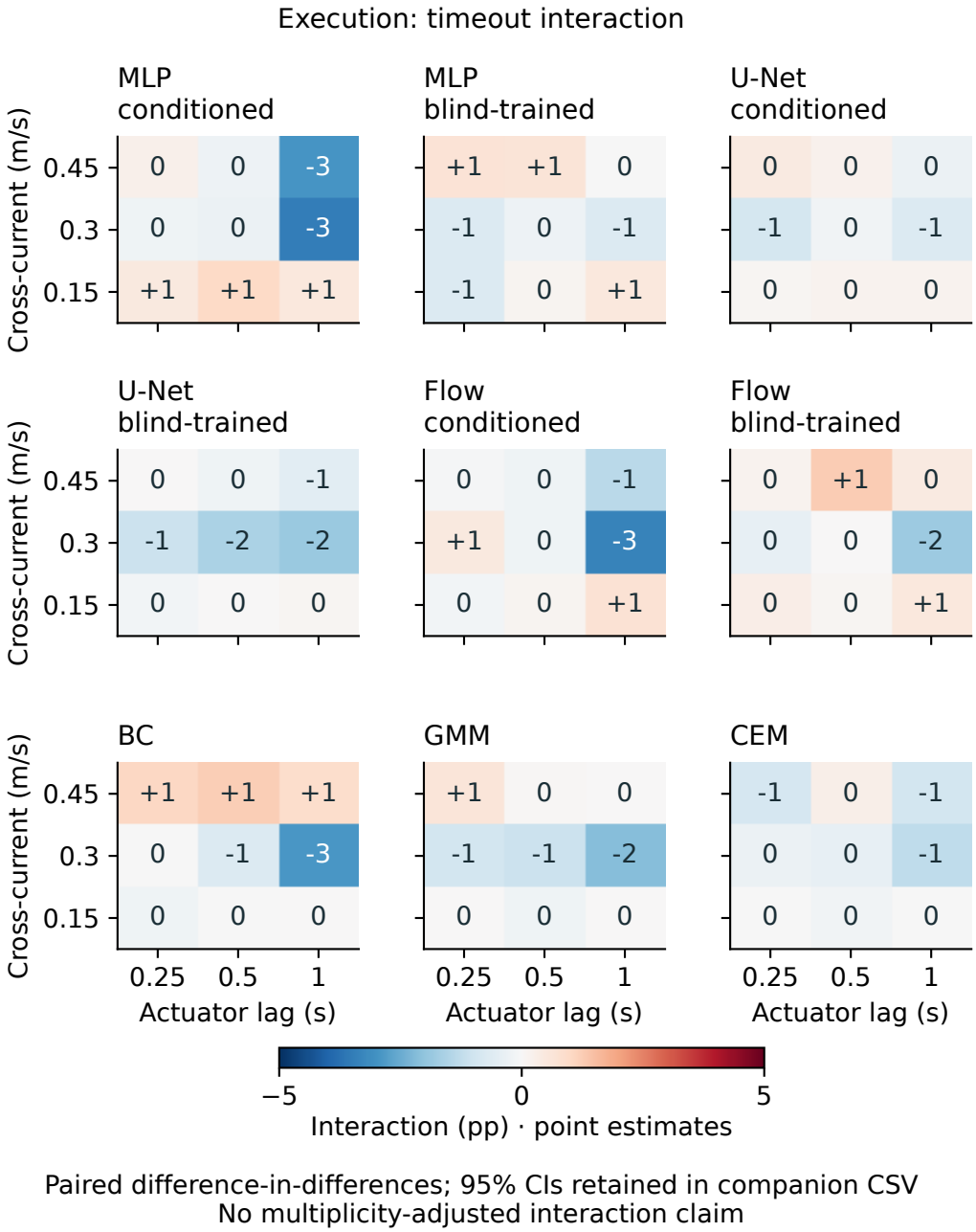


Figure 27: Timeout departure from additivity, in percentage points, for all nine policies over the current and actuator-lag grid: $f(x, y) - f(x, 0) - f(0, y) + f(0, 0)$. Point estimates are shown; descriptive 95% intervals are retained in the source CSV. The zero margins define the contrasts and are not additional interactions.