Stop Merging, Start Separating: Why Merging Learning and Modeling Won't Solve Manipulation but Separating the General From the Specific Will

Vito Mengers*,[1,2,3]
Adrian Sieler[1,2]
Alexander Koenig*,[1,2]
Aravind Battaje[1,2,3]
Xing Li*,[1,2,3]
Oliver Brock[1,2,3]

Abstract— **Recent progress in robot manipulation can be attributed to two developments: first, the application of novel learning methods, and second, the use of expertly crafted models. Consequently, merging these two developments seems a promising path for further progress. However, this only works if obtaining policies from learning and modeling possess synergistic properties. We argue that this is not necessarily the case. We discuss the reasons and suggest an alternative view of what can accelerate progress in manipulation. We then recall that this alternative view is already well-established in seminal works in robotics and show, based on our own work, that this view continues to produce advances in robotic manipulation.**

I. Introduction

Learning-based and model-based approaches both produce policies[1], but their information source differs. Learning-based approaches extract policies from data, whereas model-based approaches rely on human ingenuity. We conjecture that one of the main obstacles towards progress in robot manipulation is not a policy's information source but instead the inherent structure of the policy, irrespective of how it is obtained. If this is the case, merging learning and models alone is unlikely to lead to fundamental advances.

We propose that the inherent limitation of existing approaches to robot manipulation stems from the fact that they attempt to form the (almost) complete policy *prior* to task execution. Instead, we argue, it is better to build a preliminary policy template that captures the instance-independent, general information required for robustness and generality. Such a general policy template should represent the strongest possible inductive bias for the problem class while leaving those aspects of the policy unspecified that depend on the specific problem instance. Of course, the general policy template is insufficient; instance-specific information is essential to complete it. We argue this completion should be done *during execution*, based on the information obtained directly from the problem instance itself.

* Equal contributions
[1] Robotics and Biology Laboratory, Technische Universität Berlin
[2] Robotics Institute Germany
[3] Science of Intelligence, Research Cluster of Excellence, Berlin
ChatGPT suggested the first line of the title.
This work has been partially funded by the BMBF within the Robotics Institute Germany, grant No. 16ME1000, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project No. 390523135, and the German Priority Program – DFG-SPP 2100 "Soft Material Robotic Systems" – project No. 405033880.
[1]Or plan, program, controller—the terminology varies by community.

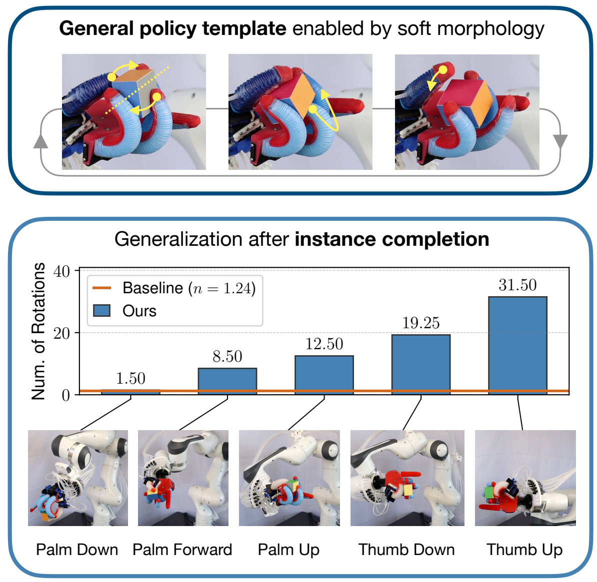

Fig. 1. The separation principle for in-hand manipulation. A general policy template dictates object rotation. Instance-specific completion occurs during execution: The soft hand's compliance adapts the motion to handle varying gravitational forces due to different wrist poses. Our two-stage approach achieves 14.65 complete object rotations on average—over an order of magnitude more than the current state-of-the-art learning-based technique [1]. We achieve more with less: Our hand is unsensorized, while the baseline uses custom tactile sensors. The videos are available at here.

Note that an extreme division into general policy template and instance-specific completion is always present when a feedback controller is running. However, in this extreme division, the policy is already fully instantiated, i.e., the policy represents an inductive bias that is overly specific. To make progress in manipulation, it is necessary to find the most appropriate, i.e., most specific inductive bias and a way to augment this inductive bias with instance-specific details.

We empirically support the two-stage policy construction approach by describing its application to three manipulation problems. The first application is dexterous in-hand manipulation, the second is long-horizon contact-rich manipulation, and the third is a novel and general action representation. We believe that the compelling and comprehensive evidence from real-world robotic manipulation solutions provide strong support for our hypothesis, namely, that the application of two-stage policy construction based on strong inductive biases and instance-specific completion enables advances in robotic manipulation.

II. Related Work

The idea of splitting task knowledge into general policy templates and instance-specific completion is not new. Early robotic systems embraced this separation out of necessity, constrained by limited computational resources. Choi and Latombe's navigation system [2] exemplifies this structure: "*a prior model of the environment is available, but this model is incomplete.*" Their system "*combines a planning component [...] to plan a lesser-committed motion plan and a reaction component that pilots the robot [...] through the unexpected obstacles.*" Robustness came not from full knowledge but from deferring specificity to execution.

In manipulation, Lozano-Pérez, Mason, and Taylor [3] observed that given an overarching plan "*details of geometry and [...] error characteristics [...] must [...] be constructed anew for each task.*" Their approach synthesizes compliant motion from general geometric models and then adapts them using task-specific error estimates. Mason [4] also shows how physical interaction itself can fill in the specific information, which "*eliminate[s] uncertainty [...] by purely mechanical means,*" while Hogan [5] frames impedance control as a way to express general task goals that adapt fluidly during execution. For locomotion, Kajita et al. [6] leveraged Model Predictive Control (MPC) with simplified models that capture the task-relevant aspects of the full system dynamics [7]. Instance-specific completion is achieved by iteratively solving a constrained optimization problem.

These early systems achieved notable robustness in unstructured environments not by relying on complete policies but by separating general priors from online adaptation. However, their general policies did not scale well to more complex manipulation or long-horizon tasks. Hence, as computational resources increased, more structured approaches emerged. Model-based planning and TAMP [8], [9] enabled symbolic and geometric reasoning but assumed complete world models, producing policies (or plans) that are fully instantiated and not adapted during execution.

Learning-based methods promise to avoid the dependency on complete world information. However, they often introduce a different form of over-commitment: Deep reinforcement learning [10]–[12] relies on accurate simulators and enormous computational resources, resulting in policies with limited generalization. Similarly, learning from demonstration (LfD) [13], [14] tends to overfit to instance-specific details [15], [16] in the learned policy rather than capturing task-invariant structure and filling in the instance-specific details later.

A principled separation between reusable priors and execution-time completion enabled robustness in early robotic manipulation systems. We believe this strategy must be emphasized more in our attempts to advance robotic manipulation. Rather than merging learning and models into a single, static policy, we advocate for a two-stage process—policy templates completed online—as a more reliable path to generalization. We now demonstrate this principle through three examples of contact-rich manipulation.

III. In-Hand Manipulation

As a first example, we look at dexterous in-hand manipulation. Forming *complete* policies of in-hand manipulation a priori fails for several reasons. Firstly, accurate instance-specific information like object properties, contact forces, or disturbances are unknown before executing the manipulation. Second, even if the information were present a priori, our current contact dynamics models do not capture the complexity of the real world. Furthermore, many policies are built in simulation. Combining learning and model-based approaches here will likely not lead to success because they do not extract complementary information. Instead, they would extract the same regularities already present in the expertly crafted models in physics simulators.

A. Policy Splitting

Enabling Principle: The soft morphology of compliant hands implicitly completes general policy templates during execution. The ability of soft hands to store and release energy in the hand's deformation enables this online completion. By appropriately changing the hand's morphology via actuation, the hand can self-stabilize against disturbances [17]. The general policy template remains simple because the hand's self-stabilization compensates for variability in the execution and fills in instance-specific details.

General Policy Template: The general policy template sequences local actuation primitives that achieve object manipulation while maximally leveraging the hand's capabilities for instance completion. Thereby, it sequences environmental constraints while storing and releasing energy to improve self-stabilization. Our policy templates are general yet simple because they are based on kinematics and provide structure for motion generation while not directly modeling object interaction.

Instance Completion: Compliant hands passively shift our general policy templates to the closest feasible instantiation that fulfills physical constraints like attaining force equilibria and energy minimization. Thereby, the compliant hardware addresses low-level control aspects (e.g., adapting to the object's shape, balancing contact forces, and counteracting external disturbances) such that the policy template can ignore the complex contact dynamics and focus on task-level information. The policy completion happens instantly in the hardware and is not limited by the bandwidth of a controller.

B. Experimental Results

We built the RBO Hand 3 [18] with a dexterous and compliant morphology to apply the two-stage policy construction and achieved highly generalizable object manipulation across many dimensions. For example, we showed that a single open-loop general policy template adapts online to different object geometries, object poses, and execution speeds [19]. In the context of feedback control, we demonstrated that coarse, linear actuations toward a goal in the hand's sensor space are sufficient for closing a control loop [20]. The compliant hardware completes the coarse control during execution.

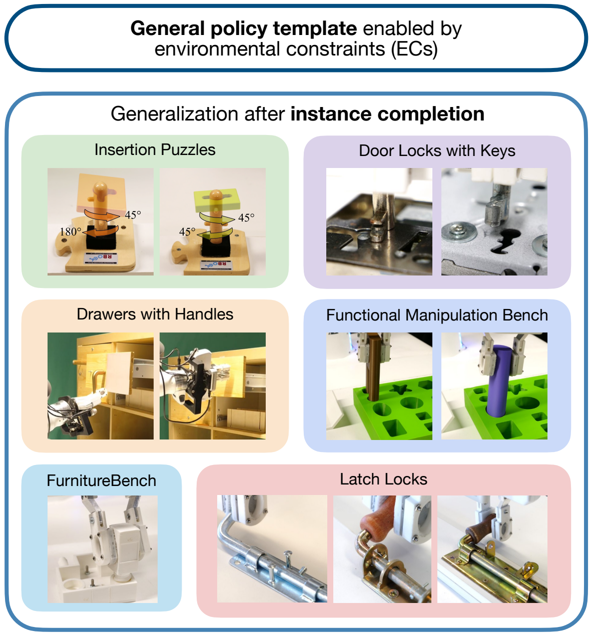

Fig. 2. Environmental constraint-based policy templates capture task-level structure that generalizes across object instances and scene variations for different manipulation tasks. From a single human demonstration, the policy infers a sequence of environmental constraints that abstract away from instance-specific geometry and appearance. During execution, this information is filled in for the specific object via contact feedback, enabling robust performance across diverse object instances. Example videos for the illustrated tasks are available here: insertion puzzles, door locks with keys, drawers with handles, Functional Manipulation Benchmark [15], FurnitureBench [16], and latch locks.

Fig. 1 illustrates another dimension of policy adaptation: A general open-loop policy template for continuous object rotation programmed in the Palm Up pose adapts to different wrist orientations of a robotic arm. Each wrist orientation represents a different problem instance, as each instance policy needs to counteract a different relative gravitational force. We achieve 14.65 complete object rotations on average, which is over an order of magnitude more than the state-of-the-art learning-based method averaged over the same wrist poses [1]. On top of this, our hand is entirely unsensorized while Yang et al. [1] use custom tactile sensors and feedback control. Note that we only evaluate our approach on one object—Yang et al. [1] test on ten objects. Our results show that the hand's soft morphology completes the simple and general policy templates with the specifics of various in-hand manipulation tasks.

IV. Learning from Demonstration Based on Environmental Constraints

Our second example examines the challenge of learning long-horizon, contact-rich manipulation tasks from human demonstrations. These tasks, as represented in the Functional Manipulation Benchmark [15] and FurnitureBench [16], combine free-space motion with forceful interactions and require robust long-horizon sequencing.

Such tasks are complex due to uncertainty in perception, object geometry variations, and noisy control, all of which can derail execution if not handled adaptively. Many learning-based approaches overfit to instance-specific details, replicating exact trajectories without capturing the underlying task structure. At the same time, fully specifying task-solving policies a priori, whether manually or through learned dynamics, is often infeasible. Key parameters like insertion depth or rotation axes are not reliably observable from vision alone, and execution is often disrupted by slippage or misalignment.

A. Policy Splitting

Enabling Principle: Environmental constraints (ECs) [21] govern how objects can be manipulated, for example, the prismatic constraint of a drawer or the revolute constraint of a door. These constraints abstract away from exact geometry and instead capture structure common to all instances within a category of objects. As such, they offer a compact representation of task-relevant regularities that remain consistent across different object instances.

General Policy Template: Rather than memorizing geometric features, we encode the observed interaction during demonstration into a hybrid automaton shaped by the task's ECs. Because this policy template is very compact, it can be learned directly from a *single* demonstration [22], [23].

Instance Completion: At execution time, instance-specific geometric information gets filled in through sensor feedback, particularly from contact forces. The robot compliantly follows the expected EC while simultaneously estimating its parameters, even if the object pose has shifted or it is a novel object instance. Deviations, such as unexpected constraints, are detected via contact monitoring and can trigger targeted requests for human corrections, refining the policy for that instance without requiring extensive retraining.

B. Experimental Results

Using ECs, we achieve high success rates (90%) on two long-horizon contact-rich manipulation benchmarks: the Functional Manipulation Benchmark [15] and FurnitureBench [16]. This result is accomplished with only a *single* demonstration per task, whereas other methods typically require hundreds and still yield lower success. Beyond benchmarks, the same method generalizes to diverse tasks, including insertion puzzles, door locks with keys, drawers with handles, and latch locks, as illustrated in Fig. 2.

These experiments demonstrate that EC-based policy templates capture task-invariant structure that generalizes across object geometries and configurations [22], [23]. The resulting policy templates remain compact yet expressive by abstracting away from instance-specific motion, enabling effective learning from a single demonstration. Compliant execution and targeted human corrections further enable adaptation to unmodeled variations and disturbances. As a result, the system achieves robust generalization across a wide range of real-world scenarios without requiring extensive retraining.

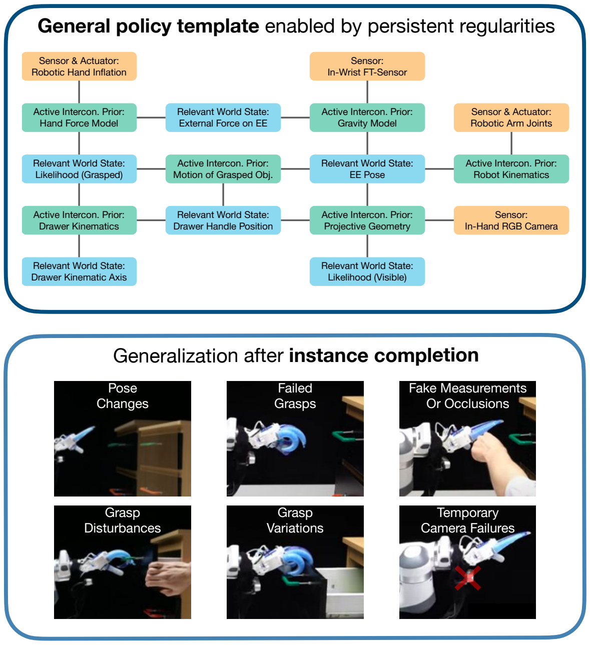

Fig. 3. AICON encodes task structure in policy templates by capturing persistent regularities as recursive estimators and linking those regularities using actively modulated interconnections. The AICON-ic policy template for partially observable drawer-opening using a robot sensorized with an RGB camera and force-torque inputs is shown on top. During execution, the policy is completed online via state estimation from sensory signals. Given these state estimates, actions are selected by following the gradient of the overall policy template. Since both estimates and gradients adapt continuously to feedback, the resulting behavior remains robust under various disturbances, as illustrated at the bottom. Watch how the gradient through the policy template adapts in real-time to disturbances here.

V. A General Action Representation Leveraging Instance-Specific Completion

In our third example, we consider the general challenge of sequential manipulation in real-world settings, where uncertainty and unexpected disturbances are the norm. Under this uncertainty, even simple contact-rich tasks like drawer opening can unfold in unpredictable ways.

Critical factors such as object geometry, friction, and occlusions are often unknown and may change during execution. This makes precomputing policies impractical and renders both handcrafted and learned policies brittle when deployed outside their design assumptions, whether those stem from engineering foresight or training distributions. To address this, we need a suitable representation that enables the separation of general structure from instance-specific details.

A. Policy Splitting

Enabling Principle: It is infeasible to enumerate all possible physical parameter combinations or disturbance scenarios ahead of time. However, persistent regularities hold across task instances, such as how force and motion interact or how objects project into the image frame. By encoding these structural relationships between relevant state variables we can reason about how to act given instance-specific physical parameters and unexpected events.

General Policy Template: Active InterCONnect (AICON) [24], [25] encodes these regularities as a network of actively interconnected estimators. Each estimator models the temporal evolution of a state variable, while the connections between them capture structural relationships among several variables. These interconnections are *active*, meaning their influence is modulated dynamically based on the state estimates, allowing the same structure to adapt flexibly to different task instances. As this dynamic network reflects the structure of the task, it can also be used to derive signals that represent the desired actions to achieve goals simply by propagating gradients through the network.

Instance Completion: During execution, instance-specific information is continuously integrated through sensory feedback. This refines both the estimators and their interconnections in real-time. Further, actions required to complete the task can be derived from the network's gradients that are shaped by the current estimates and the regularities encoded in the network's structure. This allows the system to adapt its behavior smoothly to instance-specific conditions and disturbances without the need for discrete planned stages and specifying trajectories.

B. Experimental Results

AICON-based behavior exhibits strong robustness by adapting continuously to sensory feedback. The system maintains accuracy across varying objects, lighting, and viewpoints in perception tasks such as kinematic structure estimation [26] or object segmentation [27]. More critically, the control behavior emerging from the system's gradients shows robustness to sensory noise, model uncertainty, and various unmodeled disturbances [25].

For instance, in a real-world drawer-opening task using only a wrist-mounted RGB camera and a force-torque sensor in Fig. 3, AICON enables successful execution despite challenges such as varying drawer poses, disturbance forces, grasp errors, and various sensor disturbances. As a result, it outperforms planning-based methods in success rate and robustness under uncertainty and disturbances [25].

Beyond robotics, AICON's ability to adaptively complete the gaps for a specific instance has proven valuable for biological modeling. For example, it has been used to model aspects of human vision beyond previously explored parameter ranges [24] and has generalized effectively to unseen real-world video stimuli [28].

VI. Call to Action

We argued that progress in robotic manipulation depends less on whether policies are learned, engineered, or both but more on how they are structured. Based on our previous research, as well as seminal works [2]–[7], we advocate for a two-stage policy construction process that separates general, task-invariant structure from instance-specific details. Strong inductive biases capture this general structure, while instance-specific properties complete the remaining information during execution. Whether learning-based or model-based methods are used in this process should be guided by the characteristics of the task to be solved.

References

[1] M. Yang, C. Lu, A. Church, Y. Lin, C. J. Ford, H. Li, E. Psomopoulou, D. A. Barton, and N. F. Lepora, "AnyRotate: Gravity-invariant in-hand object rotation with sim-to-real touch," in *Conference on Robot Learning (CoRL)*, 2024.

[2] W. Choi and J.-C. Latombe, "A reactive architecture for planning and executing robot motions with incomplete knowledge," in *IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS)*, vol. 1, 1991, pp. 24–29.

[3] T. Lozano-Pérez, M. T. Mason, and R. H. Taylor, "Automatic synthesis of fine-motion strategies for robots," *The International Journal of Robotics Research*, vol. 3, no. 1, pp. 3–24, 1984.

[4] M. Mason, "The mechanics of manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, 1985, pp. 544–548.

[5] N. Hogan, "Impedance control: An approach to manipulation," in *American Control Conference*. IEEE, 1984, pp. 304–313.

[6] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped walking pattern generation by using preview control of zero-moment point," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, 2003, pp. 1620–1626.

[7] R. Full and D. Koditschek, "Templates and anchors: Neuromechanical hypotheses of legged locomotion on land," *The Journal of Experimental Biology*, vol. 202, pp. 3325–32, 2000.

[8] L. P. Kaelbling and T. Lozano-Pérez, "Hierarchical task and motion planning in the now," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1470–1477.

[9] M. Toussaint, "Logic-geometric programming: An optimization-based approach to combined task and motion planning." in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015, pp. 1930–1936.

[10] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, *et al.*, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research (IJRR)*, vol. 39, no. 1, pp. 3–20, 2020.

[11] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal, "Visual dexterity: In-hand dexterous manipulation from depth," in *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

[12] J. Pitz, L. Röstel, L. Sievers, D. Burschka, and B. Bäuml, "Learning a shape-conditioned agent for purely tactile in-hand manipulation of various objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 13 112–13 119.

[13] A. O'Neill, A. Rehman, A. Gupta, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, *et al.*, "Open X-embodiment: Robotic learning datasets and RT-X models," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 6892–6903.

[14] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, *et al.*, "π_0: A vision-language-action flow model for general robot control," *arXiv preprint arXiv:2410.24164*, 2024.

[15] J. Luo, C. Xu, F. Liu, L. Tan, Z. Lin, J. Wu, P. Abbeel, and S. Levine, "FMB: a functional manipulation benchmark for generalizable robotic learning," *The International Journal of Robotics Research (IJRR)*, 2023.

[16] M. Heo, Y. Lee, D. Lee, and J. J. Lim, "FurnitureBench: Reproducible real-world benchmark for long-horizon complex manipulation," in *Robotics: Science and Systems (RSS)*, 2023.

[17] A. Sieler, A. Koenig, and O. Brock, "What is the key to dexterous manipulation: Learning or compliance?" in *German Robotics Conference (GRC)*, 2025.

[18] S. Puhlmann, J. Harris, and O. Brock, "RBO Hand 3: A platform for soft dexterous manipulation," *IEEE Transactions on Robotics (T-RO)*, vol. 38, no. 6, pp. 3434–3449, 2022.

[19] A. Bhatt, A. Sieler, S. Puhlmann, and O. Brock, "Surprisingly robust in-hand manipulation: An empirical study," in *Proceedings of Robotics: Science and Systems (RSS)*, 2021.

[20] A. Sieler and O. Brock, "Dexterous soft hands linearize feedback-control for in-hand manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 8757–8764.

[21] C. Eppner, R. Deimel, J. Alvarez-Ruiz, M. Maertens, and O. Brock, "Exploitation of environmental constraints in human and robotic grasping," *The International Journal of Robotics Research (IJRR)*, vol. 34, no. 7, pp. 1021–1038, 2015.

[22] X. Li and O. Brock, "Learning from demonstration based on environmental constraints," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 10 938–10 945, 2022.

[23] X. Li, M. Baum, and O. Brock, "Augmentation enables one-shot generalization in learning from demonstration for contact-rich manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3656–3663.

[24] A. Battaje, A. Godinez, N. M. Hanning, M. Rolfs, and O. Brock, "An information processing pattern from robotics predicts properties of the human visual system," *bioRxiv*, no. 2024.06.20.599814, 2024.

[25] V. Mengers and O. Brock, "No plan but everything under control: Robustly solving sequential tasks with dynamically composed gradient descent," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

[26] R. Martín-Martín and O. Brock, "Coupled recursive estimation for online interactive perception of articulated objects," *The International Journal of Robotics Research (IJRR)*, vol. 41, no. 8, pp. 741–777, 2022.

[27] V. Mengers, A. Battaje, M. Baum, and O. Brock, "Combining motion and appearance for robust probabilistic object segmentation in real time," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 683–689.

[28] V. Mengers, N. Roth, O. Brock, K. Obermayer, and M. Rolfs, "A robotics-inspired scanpath model reveals the importance of uncertainty and semantic object cues for gaze guidance in dynamic scenes," *Journal of Vision*, vol. 25, no. 2:6, 2025.