Comparative Study of Adaptive Robotic Arm Grasping with Human-Like Capabilities

Yunqi Huang
Mounds Park Academy

Abstract

In a bubble-tea shop scenario, robotic arms must grasp and transport liquid-filled cups made of different materials, such as paper, plastic, or glass. Due to material-dependent compliance and friction, together with the effects of liquid sloshing, the task can easily lead to cup deformation, micro-slippage, and spillage. This study aims to improve grasping stability by mapping human functional mechanisms—material intuition, commonsense-driven parameterization, and tactile feedback—into robotic manipulation strategies. Three methods are compared: (1) a fixed-parameter baseline, (2) a method that combines vision perception with cognitive reasoning and material understanding to perform parameterized settings, and (3) tactile-feedback closed-loop control. The evaluation protocol labels each trial as either task complete or failure-type labels (liquid spilled, misplaced placement, cup dropped), and a single trial may exhibit multiple failure types simultaneously. The experimental results show that introducing human functional mechanisms (physical material understanding, cognitive commonsense, tactile feedback) can improve grasping stability for cross-material liquid-cup manipulation and reduce common problems such as deformation, slipping, and spilling.

Keywords: bubble-tea shop robotics; liquid-filled cup manipulation; cross-material grasping; grasp stability; material intuition; commonsense-driven parameterization; tactile feedback; closed-loop control; failure mode analysis

1. Introduction

As robots transition from structured factory lines to open service environments, they increasingly face *non-standard* objects and dynamic disturbances. In a bubble-tea shop, a robotic arm must grasp and transport liquid-filled cups made of different materials—paper, plastic, and glass. These cups differ significantly in stiffness, friction, and deformation behavior, and condensation can further change surface friction, making stable grasping harder. Unlike industrial parts, bubble-tea cups are not only required to be "picked up," but also to be carried and placed accurately while minimizing sloshing-related risk: even when a cup does not drop, micro-slippage inside the gripper can change the cup's pose, amplify motion-induced oscillation, and lead to visible failures such as spillage or placement error. In this task, small errors in grip force, motion trajectory, and feedback adjustment can quickly escalate into real-world failure.

Many existing grasping solutions are unstable in this setting because fixed strategies do not generalize across materials. Vision systems can recognize the *appearance* of a "cup," but they typically cannot directly infer invisible physical properties such as friction, stiffness, and effective weight; as a result, the same grasp parameters may be too strong for some cups (causing deformation) and too weak for others (causing slip). More importantly, the risk for liquid containers does not come only from the initial grasp, but also from transportation dynamics and pose changes during motion, which requires the system to correct itself online.

In contrast, humans naturally employ a set of human functional mechanisms when handling liquid-filled cups: material intuition to anticipate softness and slip risk, commonsense-driven strategy selection to choose safer force and motion profiles before contact, and tactile closed-loop adjustment to sense micro-slip or deformation during contact and immediately fine-tune the grip. Motivated by this observation, this paper asks: How can we engineer and map these human functional mechanisms into robotic grasping strategies to improve stability for cross-material cup manipulation in bubble-tea shop scenarios?

To answer this question, we built a tabletop experimental platform based on a self-constructed 3-axis robotic arm equipped with a two-finger gripper. For perception, we used a DFRobot HuskyLens camera for vision-based recognition and installed two FSR pressure sensors on the gripper to provide tactile feedback. The tabletop setup emulates key elements of a bubble-tea shop workflow, including grasping cups, transporting them under motion-induced disturbances, and placing them at different target locations. We compare three strategies: (1) a fixed-parameter baseline that uses the same grasp force and motion trajectory for all cups; (2) a vision + commonsense parameterization method in which the camera identifies cup material/type and the system switches to corresponding safe grasp parameters before execution; and (3) a tactile closed-loop method that uses FSR feedback to adjust gripper angle/force during grasping and transportation to suppress slip and excessive deformation.

For fair comparison, we use a unified evaluation criterion and categorize outcomes as task complete or one of three failure modes: Spilled, Misplaced, and Dropped. “Spilled” is evaluated using a dedicated physical proxy: we place a small box on top of the cup as a spill indicator, and if excessive shaking causes the box to slide off, the trial is labeled Spilled. “Misplaced” is labeled when the cup is not placed at the designated target location. “Dropped” is labeled when the entire cup falls out of the gripper. For each cup material and each method, we conduct 50 trials. In logging, a successful trial receives a 1 in the success column; failures receive a 1 in the corresponding failure column(s). Notably, a single trial may exhibit multiple failure modes simultaneously, and in such cases multiple failure labels are recorded to capture the compound nature of liquid-cup manipulation failures.

The contribution of this work is a controlled comparative framework centered on *human functional mechanism mapping* for liquid-filled cup manipulation. Using a consistent failure taxonomy, we analyze the strengths and limitations of each strategy: vision + commonsense primarily improves pre-contact parameter selection, while tactile closed-loop control primarily improves post-contact execution stability and online correction. These findings provide practical evidence to support future deployment of robotic grasping systems in modular service scenarios, including bubble-tea shops and similar retail environments.

2. Literature Review

2.1 Overall Approaches in Robotic Grasping (Current State of the Art)

Robotic grasping research can be broadly grouped into four approaches: fixed-parameter grasping, vision-based grasping, tactile/force-feedback grasping, and soft/compliant grasping (Soft Grippers / Compliant Design). Fixed-parameter grasping relies on preset angles and constant gripping force for standardized objects; vision-based grasping uses cameras to obtain appearance and geometric information to estimate grasp points and poses; tactile/force feedback uses sensors to detect contact and adjust during execution; soft grippers expand the safe grasping range through flexible materials and compliant structures.

2.1.1 Fixed Grasping

Fixed-parameter grasping has practical engineering advantages for standardized objects or repetitive tasks: it is simple, predictable, and easy to reproduce. However, it implicitly assumes that an object's physical properties and contact conditions are sufficiently stable. When material, friction, and stiffness vary significantly, or when the task involves dynamic disturbances, a single set of parameters often fails to satisfy both "grip securely" and "avoid damage to the cups," leading to systematic failures such as crushing, slipping and dropping, and spillage when handling liquid containers.

2.1.2 Vision-based Grasping

Vision-based grasping relies on cameras to obtain an object's appearance and geometric information. It can recognize what the object is / which category it belongs to, and adjust key geometric decisions for grasping, including grasp position, grasp angle/direction, grasp speed, and the gripper's approach pose, thereby improving grasp success and stability.

However, vision information cannot directly provide the cup's underlying physical properties such as friction, stiffness, hardness, and weight, which are invisible from appearance alone. Therefore, in cross-material tasks, a vision-based method may know it is a cup, but it does not know the physical differences among cups made of different materials or how they respond to force, which can still lead to mismatched strategies and failures such as deformation, slippage, or dropping.

2.1.3 Tactile Feedback

Tactile/force feedback uses sensors to detect contact states and adjust the grasp in real time during grasping and transportation, making it a key direction for improving grasp stability. This approach emphasizes online correction during execution. for example, compensating when slip tendencies appear, thereby reducing the risk of eventual dropping, and it can also help avoid applying excessive pressure to fragile objects.

At the same time, tactile/force feedback typically focuses on execution-phase stability and usually becomes effective after contact occurs. It is a reactive correction mechanism and does not replace pre-contact strategy selection and planning.

2.1.4 Soft Grippers

Soft grippers expand the safe grasping range through flexible materials and compliant structures. They are particularly suitable for fragile or easily deformable objects, reducing damage risk by enabling gentler contact. However, for tasks that require precise placement, resistance to dynamic disturbances, and stable pose maintenance, compliance often introduces trade-offs in controllability and precision.

2.2 Key Gaps in the Bubble-Tea Cup Task (Current Issues)

Liquid-cup manipulation combines cross-material differences with liquid sensitivity in a single task chain, exposing more concentrated gaps in general grasping approaches. These gaps can be summarized in three points:

(1) Limits of vision perception: shape is visible, key physical quantities are not. Cameras can provide appearance and geometric information, but they cannot directly obtain invisible physical properties such as weight, friction, and stiffness.

(2) Lack of material intuition: inability to understand how different materials respond to force. Without some human common sense understanding of material differences, robots can struggle

to reliably distinguish the consequences of applying force to paper cups VS glass cups, which makes cross-material tasks prone to deformation or slippage failures.

(3) “With liquid vs. without liquid” requires different motion strategies. The speed cannot be the same. Fixed algorithms often fail to adapt to changes in an object’s internal state. If a system does not distinguish whether a cup contains liquid and still uses the same strategy, spillage or damage becomes more likely. Intuitively, when holding a drink with liquid, people naturally move more slowly and steadily, because moving too fast increases sloshing and the risk of spillage. Liquid-cup grasping similarly needs to treat whether liquid is present as an important factor for motion-strategy differences.

2.3 Positioning of This Work: Evaluating Stability Improvements from Mapping Human Functional Mechanisms in Cross-Material Liquid-Cup Grasping

This study looks at cross-material grasping of liquid-filled cups in bubble-tea shop tasks. It asks whether incorporating human functional mechanisms into robotic-arm grasping strategies—such as physical material understanding, cognitive commonsense, and tactile feedback—can make cup grasping and transportation more stable. The study also provides reusable guidance for similar service scenarios.

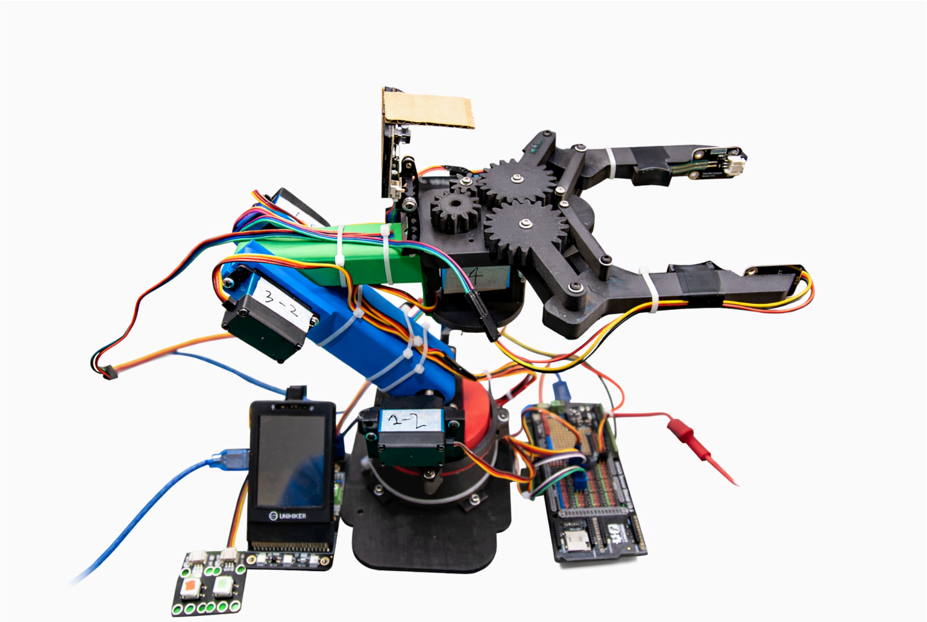
The goal of this study is to compare three strategies under the same experimental platform and the same evaluation criteria: a fixed-parameter baseline method; a method that combines vision perception, reasoning, and material knowledge to set parameters; and a closed-loop control method based on tactile feedback. A labeling scheme is used to record each trial as either “task complete” or one or more failure-type labels: Liquid Spilled, Misplaced, and Cup Dropped. A single trial may exhibit more than one failure type at the same time. The Results and Discussion sections then examine the contributions of mapping human functional mechanisms by analyzing two aspects: success rate and failure-type distribution. This work also provides reference for deploying grasping systems in modular environments such as shops.

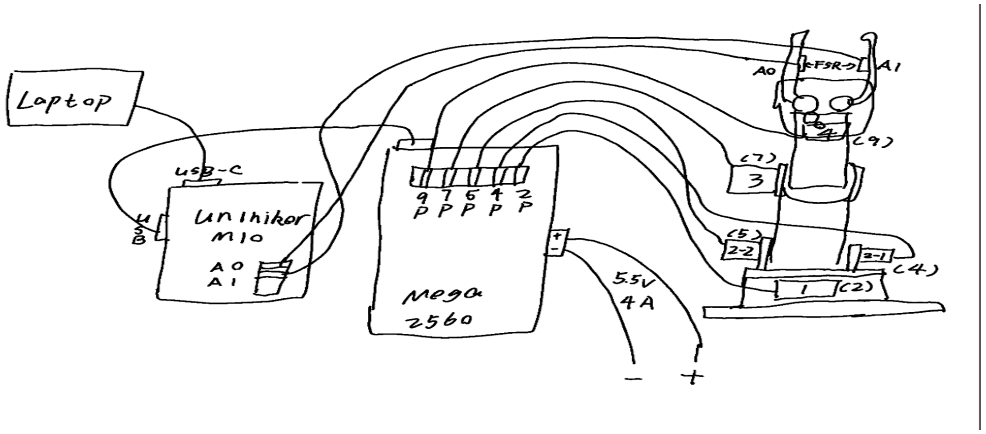
3. Methodology

3.1 Experimental Platform and Hardware Setup

This study uses a self-built three-axis robotic arm as the experimental platform. The joints are driven by 35 kg-class servos. The middle arm and upper arm adopt a synchronous dual–35 kg servo drive structure to obtain larger output torque, thereby improving posture-holding stability during transportation. The end-effector is a two-jaw, two-finger gripper.

For perception, a DFRobot HuskyLens camera is used for target identification, and two FSR pressure sensors are mounted on the two gripper sides for contact pressure sensing. The measurement range of the FSR pressure sensors is approximately 30 g to 1.5 kg. The experimental objects include three common cup materials in bubble-tea shops: paper cups, plastic cups, and glass cups.





3.2 Control Architecture and Action Triggering

The control system follows an “upper-level decision + lower-level execution” structure. The upper-level controller is a Xingkong M10 running Python, responsible for task flow control and strategy selection. The lower-level controller is an Arduino Mega2560, responsible for driving servos to execute arm and gripper motions. During operation, the M10 sends serial commands to the Mega2560 to trigger actions and complete the grasp-and-place procedure.

3.3 Task Definition and Single-Trial Procedure (Motion Sequence)

Each trial follows a unified motion sequence:

1. the robotic arm starts from the initial position;
2. it moves to the grasping position and grasps the cup;
3. it moves to the water-filling position for filling;

4. it moves to the designated platform position;

5. it places the cup on the platform;

6. it returns to the initial position, ending the trial.

3.4 Outcome Definition and Failure-Type Labeling

To ensure comparability across methods, each trial is recorded as either task complete or assigned one or more failure-type labels; multiple failure types are allowed to occur in a single trial.

- Liquid Spilled (Spilled): A small box is placed on top of the cup as a spill indicator. If excessive shaking causes the box to slide off the cup top during transportation, the trial is labeled Liquid Spilled.

- Misplaced (Misplaced Placement): If the cup is not placed at the designated target location (the specified position on the platform), the trial is labeled Misplaced.

- Cup Dropped (Dropped): If the cup fully falls out of the gripper during grasping or transportation, the trial is labeled Cup Dropped.

3.5 Compared Methods (Method 1 / Method 2 / Method 3)

The comparison focus of this study is: under the same cup condition, comparing the performance differences among Method 1, Method 2, and Method 3. Therefore, for each cup, all three methods are executed and recorded under the same evaluation criteria.

3.5.1 Method 1: Fixed-Parameter Baseline (Assembly-Line Repetition)

Method 1 is a fixed-parameter baseline method, and its overall logic is similar to an assembly line: after presetting a fixed set of grasp parameters and a fixed motion procedure, the robotic arm repeats the same actions across all trials. This method includes no identification, no decision-making, and no grasp-stage fine-tuning, and the parameters remain unchanged during trials.

Because different cups have different sizes (e.g., diameters), before starting trials for a specific cup type, Method 1 uses a manually set “standard grasp opening / gripper angle” corresponding to that cup size; however, once a trial begins, the robotic arm still repeats the same fixed parameters and fixed motion procedure without any adjustment.

3.5.2 Method 2: Vision + Cognitive Reasoning with Material Understanding

Method 2 uses vision-based identification results and performs parameter settings by combining cognitive reasoning with material understanding. In this study, identification is implemented using QR codes, providing stable and accurate recognition. Based on the identified cup information, the system selects corresponding grasp and motion parameters to improve stability in cross-material and liquid-cup manipulation.

Specifically, the parameter settings in Method 2 are reflected in using different grasp forces and motion strategies for different cup materials. For example, paper cups are more likely to be crushed, so the grasp force needs to be controlled to be smaller; glass cups are usually heavier and more prone to slipping, and the material is relatively hard and not easily deformed by squeezing, so the grasp force can be appropriately increased to reduce the risk of dropping. Meanwhile, Method 2 also differentiates motion strategies based on object state: for example, it can use a slower moving speed for glass cups to reduce risk; it uses a slower moving speed for cups with liquid to reduce shaking and spillage, and a relatively faster moving speed for cups without liquid.

3.5.3 Method 3: Tactile-Feedback Closed-Loop Fine-Tuning

In Method 3, the tactile-feedback closed loop mainly occurs during the grasping stage: after moving from the initial position to the grasping position, the system reads the pressure values from the two FSR sensors in real time while the gripper closes and computes their average as the pressure indicator for that grasp.

Before the formal 50 trials, this study conducts approximately 10–15 pre-trials to establish a "standard grasp pressure value" for each cup (a reference standard based on the average of the two FSR readings). During the subsequent 50 trials for each cup, if a grasp shows changes in the average pressure value relative to the standard value due to grasp deviation or related factors, the system performs fine-tuning during the grasping stage based on this change, to improve grasp stability and reduce failures.

Item	**Method 1 (Baseline)**	**Method 2 (Vision)**	**Method 3 (Tactile)**
Object	Plastic, Glass, Paper Cups	Plastic, Glass, Paper Cups	Plastic, Glass, Paper Cups
Method	No Sensors / Blind Execution	**Camera + Cognitive Commonsense**	**Pressure Sensor**
Parameter	**Fixed Force & Trajectory** (Repeats same values regardless of material)	**Retrieve Parameters via Material ID** (Identifies cup type and loads specific force/motion values for that object)	**Adjust Degrees based on Pressure** (Detects pressure during the grasp and decides whether to modify gripper angle to optimize force)
Results	**Success:** Task Complete **Fail:** 1. Liquid Spilled 2. Misplaced 3. Cup Dropped	**Success:** Task Complete **Fail:** 1. Liquid Spilled 2. Misplaced 3. Cup Dropped	**Success:** Task Complete **Fail:** 1. Liquid Spilled 2. Misplaced 3. Cup Dropped

3.6 Trials and Data Recording

For each cup material and each method, 50 trials are conducted. After each trial:

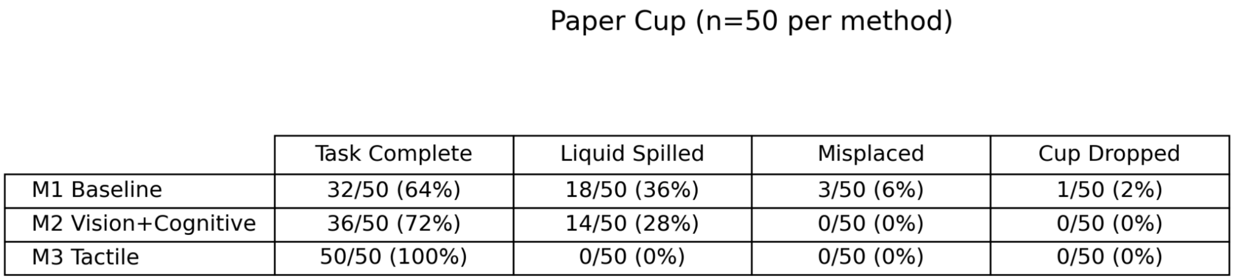
- If the task is completed, a 1 is recorded in the success field.
- If a failure occurs, a 1 is recorded in the corresponding failure-type field(s).
- A single trial may receive multiple failure labels if multiple failures occur simultaneously.

4. Results

This study compares three methods (Method 1/2/3) across three cup types (paper, plastic, and glass). For each “cup × method” condition, 50 trials are conducted. Each trial is recorded as either Task Complete (success) and/or labeled with one or more failure types: Liquid Spilled, Misplaced, and Cup Dropped.

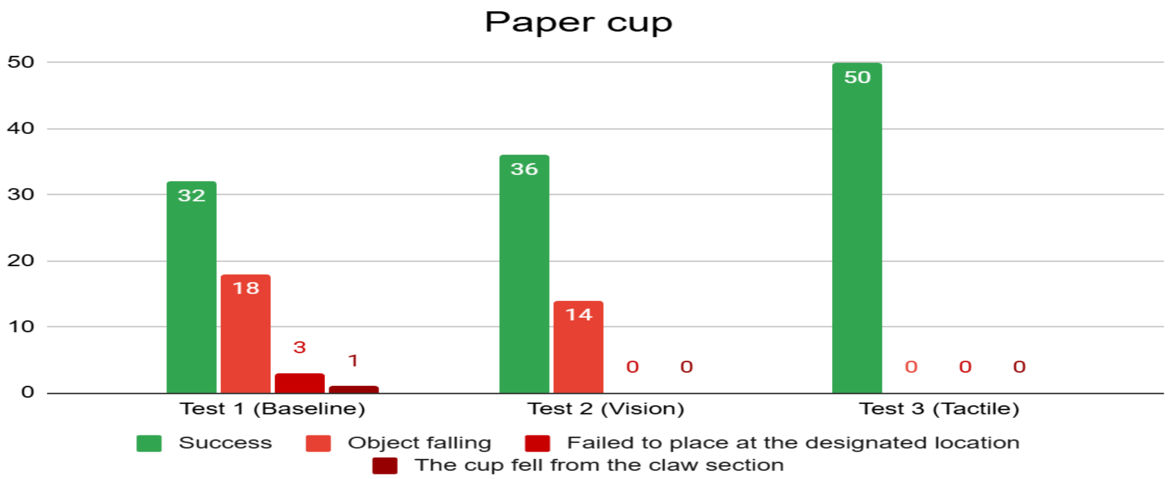
Note that a single trial may exhibit multiple failure types at the same time (for example, the same trial can be marked with more than one failure label). Therefore, the counts of failure labels can overlap and are not required to sum to 50.

4.1 Paper cup



Paper Cup (n=50 per method)

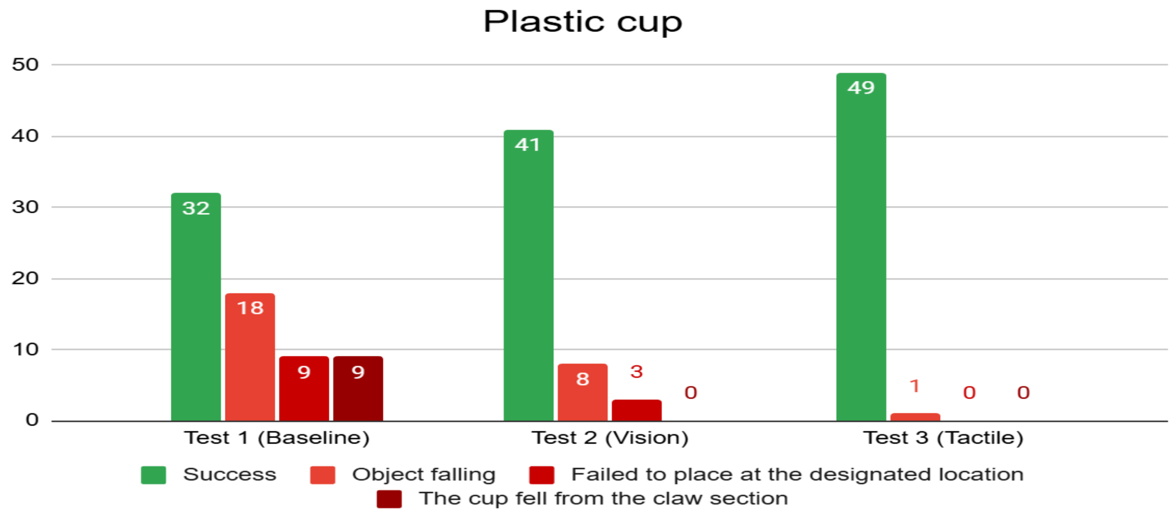
	Task Complete	Liquid Spilled	Misplaced	Cup Dropped
M1 Baseline	32/50 (64%)	18/50 (36%)	3/50 (6%)	1/50 (2%)
M2 Vision+Cognitive	36/50 (72%)	14/50 (28%)	0/50 (0%)	0/50 (0%)
M3 Tactile	50/50 (100%)	0/50 (0%)	0/50 (0%)	0/50 (0%)



4.2 Plastic cup

Plastic Cup (n=50 per method)

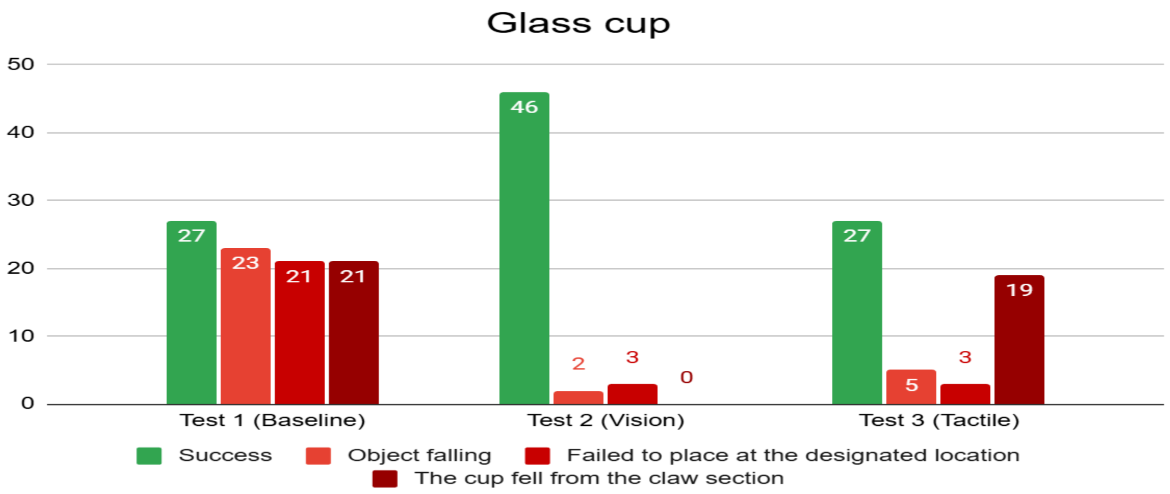
	Task Complete	Liquid Spilled	Misplaced	Cup Dropped
M1 Baseline	32/50 (64%)	18/50 (36%)	9/50 (18%)	9/50 (18%)
M2 Vision+Cognitive	41/50 (82%)	8/50 (16%)	3/50 (6%)	0/50 (0%)
M3 Tactile	49/50 (98%)	1/50 (2%)	0/50 (0%)	0/50 (0%)



4.3 Glass cup

Glass Cup (n=50 per method)

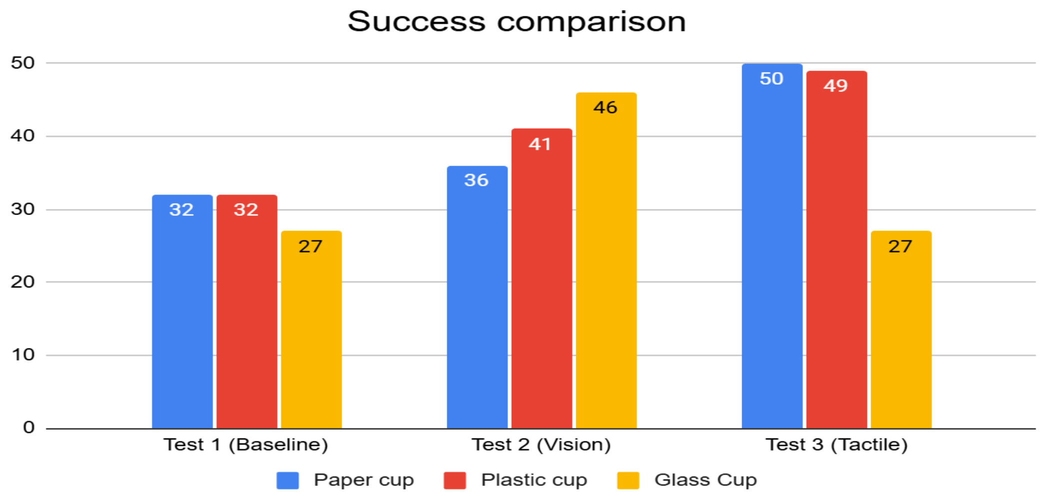
	Task Complete	Liquid Spilled	Misplaced	Cup Dropped
M1 Baseline	27/50 (54%)	23/50 (46%)	21/50 (42%)	21/50 (42%)
M2 Vision+Cognitive	46/50 (92%)	2/50 (4%)	3/50 (6%)	0/50 (0%)
M3 Tactile	27/50 (54%)	5/50 (10%)	3/50 (6%)	19/50 (38%)



4.4 Experimental video

Youtube: https://youtu.be/gPxInTdssOs
Bililbilil:
https://www.bilibili.com/video/BV1dGzfBdEzm/?share_source=copy_web&vd_source=cda449dd0d81e71b2f82dfab0a08269f

5. Discussion



5.1 Paper cup: why tactile feedback improves stability on deformable objects

Paper cups are soft and deform easily, so their grasping stability strongly depends on keeping the clamp force within a safe range. Under Method 1, the robot repeats a fixed set of parameters like an assembly line. Because the preset gripper closing amount and grasp force may not match the actual cup condition, the cup wall can deform. This deformation can increase shaking during transport and raise the probability of spill-related failures. Method 2 improves the process by using vision identification together with cognitive reasoning and material knowledge to select more appropriate grasp and motion parameters for paper cups, reducing failures caused by parameter mismatch.

Method 3 shows the clearest improvement on paper cups because it maps human touch into a pressure-based feedback loop during the grasping stage. While the gripper closes, the two FSR sensors read contact pressure in real time, and the system uses the average value as the pressure indicator. The indicator is compared with a target pressure range established in pre-trials. If the pressure is too high, the risk of cup deformation increases. If the pressure is too low, the risk of slipping and later dropping increases. Method 3 fine-tunes the grasp in real time during closing so that the gripper reaches a state that avoids deformation while still preventing slippage. As a result, the grasp remains more stable during transportation and placement.

5.2 Plastic cup: transport shaking is the main cause of liquid spills

Plastic cups are less likely to be crushed than paper cups. In this task, the main problem for plastic cups is liquid spill caused by shaking during transport. Method 1 uses fixed motion and fixed settings, and it does not adjust the motion strategy based on whether the cup contains liquid. As a result, small shakes during movement can accumulate and cause the liquid to spill. Method 2 addresses this by using commonsense and material knowledge to set more suitable motion parameters. When the system judges that the cup contains liquid, it slows down and uses smoother motion to reduce shaking and lower the spill probability. When the cup is empty, the system can move relatively faster. Method 3 further helps by stabilizing the grasp at the start of transport. Using pressure feedback, the system reaches a more consistent grasp state, which reduces initial micro-slip and improves transport stability, weakening the shaking-to-spill chain.

5.3 Glass cup: strategy selection is critical, and grasp-stage tactile tuning has limits

Glass cups are heavier, harder, and more likely to slip, so they present a different stability profile. The results show that Method 2 provides the strongest improvement on glass cups and sharply reduces drop-related failures. This suggests that selecting parameters and motion strategies based on material understanding and commonsense is especially important for glass cups. In particular, using slower and smoother motion for fragile objects and liquid-filled conditions can reduce disturbance and lower the risk of slipping and dropping.

In contrast, Method 3 does not increase the task completion rate on glass cups, and the main remaining failure type is Cup Dropped. This reveals a limitation of the current Method 3 setup. Its tactile feedback loop is applied mainly during the grasping stage to align grasp pressure with a pre-established target range. However, for glass cups, instability can still develop after grasping during the transport stage, where slipping may occur due to inertia and changing contact conditions. Therefore, grasp-stage pressure alignment alone may not be sufficient to prevent later drops for high-slip objects such as glass cups.

5.4 Mechanism Interpretation and Integrated Analysis

Although the Results section reports the numbers, a mechanism-level summary helps connect the cup-specific findings to the central research question of mapping human functional mechanisms to robotic grasping. The evidence indicates that cognitive reasoning with material understanding is most effective for high-risk objects where correct strategy selection matters most, such as glass cups and liquid-filled conditions. Tactile feedback is most effective for objects that are sensitive to grasp-force errors, such as paper cups and plastic cups. Overall, different human-mechanism mappings provide different stability gains depending on object properties, and these findings offer practical guidance for service-scene grasping systems.

5.5 Limitations

This study has several limitations. First, the tactile feedback loop in Method 3 is mainly applied during the grasping stage and does not cover the full transport process, so it may be less effective for objects where instability develops after grasping. Second, vision identification uses QR codes. While QR codes make recognition stable and accurate, real service environments may not provide QR codes on cups, so the perception pipeline should be extended in future work. Third, the experiment is performed in a tabletop simulated setting. Real bubble-tea shops may introduce additional disturbances such as obstacles, sudden stops, and turning motions, which can increase liquid sloshing and slip risk. Fourth, liquid spill is represented by a small box sliding off the cup top, which is an indicator rather than a direct measurement of real liquid overflow. Finally, the tested cup set is limited in shape and size, and a wider range of cup geometries would improve generality.

5.6 Future work and extension experiments

Future work can expand both system capability and experimental coverage.

Extension Experiment 4 combines Method 2 and Method 3. In this extension, the robot integrates human functional mechanisms including commonsense reasoning, understanding of material properties, and tactile feedback in the same workflow. The system uses the Method 2 part to make judgments based on commonsense and material understanding, and then uses the Method 3 part to apply tactile feedback for fine-tuning during grasping. In this way, judgment and grasp adjustment are combined to improve stability for cross-material, liquid-cup grasping and transport tasks.

Extension Experiment 5 adds multimodal learning based on successful grasp cases. Building on the baseline framework, the system records visual information together with grasp-time parameters such as FSR pressure indicators and gripper closing profiles. Using successful trials as training examples, the system learns a mapping from object information and grasp-time signals to more stable parameter settings or fine-tuning actions. This aims to reduce reliance on manually designed rules and improve generalization across cup types and liquid states.

Future work can also test more cup types and surface conditions, such as different diameters, lids, and condensation. In addition, tactile feedback can be expanded from grasp-stage adjustment toward transport-stage stability monitoring to better address high-slip cases.

6. Conclusion

This study conducts a comparative evaluation of cross-material, liquid-cup manipulation in a bubble-tea shop scenario, focusing on task performance across paper, plastic, and glass cups. Using a unified evaluation framework, the study records overall task completion (Task Complete) and the occurrence of multiple failure types, including Liquid Spilled, Misplaced, and Cup Dropped. To ensure comparability, all methods are tested on the same experimental platform, under the same motion procedure, with the same number of trials per condition, and with consistent data-recording rules.

The results show that mapping human functional mechanisms into robotic grasping strategies improves overall task performance for cross-material liquid cups, but different mapping routes provide different benefits depending on the cup material. Method 2 uses vision identification together with cognitive commonsense and material understanding to set parameters, allowing the system to select more appropriate grasp and motion settings for different materials and object states. This produces consistent improvements in both success rate and failure-type control, with the most significant increase in task completion observed on high-risk objects such as glass cups. Method 3 introduces a tactile-feedback closed-loop fine-tuning process during the grasping stage based on FSR readings, bringing grasp pressure closer to a safe range. As a result, Method 3 significantly increases task completion and reduces slip- and deformation-related failures for cups that are more sensitive to grasp-force errors, such as paper and plastic cups. At the same time, the glass-cup results reveal a clear boundary of the current implementation: grasp-stage pressure fine-tuning alone may be insufficient to eliminate slip and drop risk that develops during transportation.

In summary, the main contribution of this work is a reusable and interpretable comparative evaluation framework for service-scene manipulation. Through systematic experiments across three methods and three cup materials, the study clarifies how different human-mechanism mappings affect task completion and failure-type distributions. These findings provide experimental evidence and practical reference for designing and selecting grasp strategies in modular service environments such as shops. Future work will extend the study through additional experiments that combine Method 2 and Method 3, integrating cognitive commonsense, material understanding, and tactile feedback into a single workflow, and by exploring multimodal training based on successful grasp cases to improve performance and failure suppression in more complex real-world settings.

References

Mahler, J., Liang, J., Niyaz, S., Laskey, M., Doan, R., Liu, X., Ojea, J. A., & Goldberg, K. (2017). Dex-Net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. *Robotics: Science and Systems*.

Calandra, R., Owens, A., Jayaraman, D., Lin, J., Yuan, W., Malik, J., Adelson, E. H., & Levine, S. (2018). More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robotics and Automation Letters, 3*(4), 3300–3307.

Yuan, W., Dong, S., & Adelson, E. H. (2017). GelSight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors, 17*(12), 2762.

Moriello, L., Biagiotti, L., Melchiorri, C., & Paoli, A. (2018). Manipulating liquids with robots: A sloshing-free solution. *Control Engineering Practice, 78*, 129–141.

Zhu, Y., Yang, D., & Lee, Y. (2025). Deformable and fragile object manipulation: A review and prospects. *Sensors, 25*(17), 5430.

Chu, X., Deng, J., You, G., Liu, W., Li, X., Ji, J., & Zhang, Y. (2025). GraspCoT: Integrating physical property reasoning for 6-DoF grasping under flexible language instructions. *arXiv preprint* (arXiv:2503.16013).

机器人大讲堂. (2025, November 7). 告别机器人“手残”！中国团队研发六自由度机械臂 GL-Robot, 能捏鸡蛋还能举哑铃！ 腾讯新闻.

Ikemura, K., Dong, Y., Blanco-Mulero, D., Longhini, A., Chen, L., & Pokorny, F. T. (2025). Sim-to-real gentle manipulation of deformable and fragile objects with stress-guided reinforcement learning. *arXiv preprint* (arXiv:2510.25405).