Pre-registered computational re-analysis: setup-level inconsistencies in an experimental paper and a null result in mW water (the Mpemba effect)

Baturhan Bıktım · 29 November 2026 · Non-peer-reviewed draft (pre-preprint)

Abstract

The Mpemba effect is defined as hot water reaching 0 °C before cold water (definition 1) or beginning to freeze before it (definition 2). Without running any new physical experiment, this study asks three computational questions with predictions, rejection criteria and decision rules written down in advance. (i) In a noiseless one-variable cooling model the hot sample never arrives first under any condition; once container differences, supercooling and thermometer offset are added, the false "Mpemba" rate rises to 28–36 % when the initial gap is small (5 °C) and falls to 0–5 % at 25 °C and above. (ii) Examining the published numbers of Janni et al. (2026) setup by setup, container-to-container spread in cooling rate alone explains only one of five setups (EPS tray: 20.5 % observed, 20.0 % modeled); for the baking-sheet setup the model gives 22 % against 61.8 % observed, and this result is robust to counting unit, distribution and variance-share assumptions. (iii) In the mW water model, crystallization times after an instantaneous quench to 200 K from 400 K and from 280 K starts could not be told apart (N = 216: log-rank p = 0.41, hazard ratio 0.71 [0.34–1.42]; N = 512: p = 0.40, ratio 1.61 [0.55–5.35]). This null result conflicts with Ghosh et al., who report faster crystallization from the hot start in TIP4P/Ice; the difference may come from model non-equivalence, from the regime, or from insufficient power (5–13 events per arm). The results do not decide whether the effect exists or not in water.

1. Introduction

The two definitions require different physics, and much of the contradiction in the literature comes from there. Burridge and Linden (2016) argued, on the basis of controlled experiments, that there is no evidence for a meaningful effect; Janni et al. (2026) reported 46 and 58 cases under the two definitions out of 176 measurements in five setups, and attributed the effect to airflow fluctuations and differences in supercooling. On the computational side, Jin and Goddard (2015) and Ghosh et al. (arXiv 2024; Commun. Phys. 2025) reported faster crystallization from a hot start; Teza et al. (2026) give the general framework. This study asks three questions: (1) which measurement artifacts produce a "false Mpemba" without any real effect, and how often; (2) can the rates reported by Janni et al. be explained by container cooling-rate spread alone; (3) is a start-temperature-dependent difference in crystallization seen in mW water after a rapid quench?

2. Methods

Pre-registration principle. For the Janni test and the mW study, the prediction, rejection criterion and decision rule were written before running anything and sealed with SHA-256; whatever came out was reported. Deviations were recorded in addenda (A1, A2) before any results were seen.

2.1 Toy cooling model. Newtonian cooling $dT/dt = -k\,(T - T_b)$, $T_b = -18$ °C, cold sample at 25 °C. The spread of k between containers was 0–15 %; the supercooling threshold was N(−5; 2 °C); the thermometer

offset was N(0; 1 °C); temperature dependence of k and evaporation were tried. The parameters are assumptions.

2.2 Janni test. For each setup, the paper's reported k spread (std/mean), the number of trials and samples; equally spaced initial temperatures of 22–72 °C, with a ≥ 3 °C rule; counting unit U1 ("reaching 0 °C before any colder sample") and U2 ("before the coldest reference in the trial"); k normal or log-normal; within-trial variance share f = 1, 0.7, 0.5; a k–T_0 coupling γ (relative change in k per 10 °C). Simulation with 4000 repetitions; rejection criterion: the observed rate falls outside the 95 % interval of the model. The counting unit and the design are my interpretation.

2.3 mW water MD. Molinero–Moore mW, LAMMPS, dt = 10 fs, NPT at 1 atm. The first energy-drop nucleation criterion failed its validity check (pilot record, pre-registration v2); it was replaced by the per-molecule Lechner–Dellago $\bar{q}6$ (4 neighbors), with the threshold θ = 0.815 calibrated only on reference liquid and cubic-diamond crystal (balanced accuracy 0.9994). Nucleation: the 100 ps running mean of the ice-like fraction φ reaching 0.30 and being sustained for 0.3 ns. T_b = 200 K was chosen with the pilot. Main experiment: N = 216, T_0 = 400 and 280 K (16 structures each), secondary 340 K (16), upper limit 15 ns; the N = 512 repeat used 12 structures each, 10 ns. Decision: censored log-rank test, two-sided; α = 0.01 (N = 216), α = 0.05 (N = 512).

3. Results

3.1 The false Mpemba rate depends on the initial gap. Even with a temperature-dependent cooling rate added, trajectories cannot cross in an autonomous one-variable equation; in the noiseless model the rate is 0 %. With a 10 % container difference, supercooling σ = 2 °C and offset 1 °C (n = 300, margin of error ≈ ±3 %):

Hot sample start (cold at 25 °C)	0 °C criterion	Freezing-onset criterion
30 °C	28 %	36 %
40 °C	2 %	11 %
50 °C	0 %	5 %
70 °C	0 %	1 %

3.2 The Janni et al. data at setup level. U1, f = 1, normal distribution; model mean [95 % interval]:

Setup	Observed	Model	Decision
DI, EPS tray	20.5 %	20 % [9–30]	consistent
DI, baking sheet	61.8 %	22 % [12–32]	rejected
DI, open beaker	27.8 %	45 % [22–67]	consistent (wide interval)
Tap, EPS tray	17.5 %	48 % [35–60]	rejected (over-prediction)
EPS holder	10.0 %	31 % [18–42]	rejected (over-prediction)

For the baking sheet, none of the 12 combinations of unit/distribution/variance covered 61.8 % (model 0–22 %). A k–T_0 coupling (γ ≈ +0.12–0.15 per 10 °C, fine scan exploratory) closes the baking sheet but the same γ breaks the EPS tray (which needs γ ≲ +0.04). A single γ does not explain the setups together. This does not

mean the paper is wrong; an additional mechanism specific to this setup is needed, and its identity could be determined from the raw data.

3.3 mW water.

Group	Events / n	KM median (ns)	Event rate (1/ns)
N = 216, 400 K	11/16	8.55	0.069
N = 216, 280 K	13/16	7.79	0.097
N = 216, 340 K	12/16	7.51	0.089
N = 512, 400 K	7/12	7.40	0.079
N = 512, 280 K	5/12	undefined	0.049

Log-rank (hot vs cold): N = 216 p = 0.41, hazard ratio 0.71 (bootstrap 95 % CI 0.34–1.42); N = 512 p = 0.40, ratio 1.61 (0.55–5.35). By the pre-specified decision rule, “indistinguishable” at both sizes. The direction is inconsistent between the two sizes; 340 K gave the fastest median. The thermostat reached the target within the first 100 ps, and there was no trigger faster than 0.3 ns. Secondary sensitivity $\varphi^* = 0.20$: N = 216 p = 0.97; $\varphi^* = 0.40$ could not be evaluated (the recorded traces stop at the 0.30 confirmation).

4. Discussion

Shown: a small initial gap and reasonable measurement noise can produce 28–36 % false Mpemba outcomes (the same magnitude as Janni’s pooled 26–33 % rates, although the setup-level agreement does not hold); Janni’s baking-sheet setup cannot be explained by container-rate spread; in mW the difference after an instantaneous quench could not be distinguished.

Not shown: whether the effect exists or not in water. The Janni results were not refuted; only the authors’ own explanation did not by itself match most of the setups.

Conflict with Ghosh et al. Ghosh, Pathak, Chatterjee and Das reported that hot-start samples crystallize faster in TIP4P/Ice (arXiv:2407.06954; Commun. Phys. 2025, doi 10.1038/s42005-025-02251-6; I read only the preprint version: 96 molecules, $T_f = 230$ K, 8 runs per T_s). Our null result does not refute it. Possible reasons: (a) the model — mW has no hydrogen-bond or orientation information, whereas that study ties the effect to the orientation/H-bond state of the metastable liquid; (b) quench temperature and thermostat (200 K versus 230 K); (c) power — for N = 512 the hazard-ratio interval is 0.55–5.35, so a large difference in favor of the hot start is not excluded. This study cannot separate the three.

Candidate explanations for the baking sheet (untested): coupling of k with starting temperature or shelf position; a difference in counting unit; thermal contact of the metal tray. A preprint (arXiv 2508.05607) reports that, under the definition involving supercooling, rankings favoring the hot sample were seen when the freezer temperature is close to the nucleation temperature; this is qualitatively consistent with the finding in 3.1 that the “freezing-onset criterion produces more false positives”. Being a preprint, it is context, not evidence.

5. Limitations

- The toy model and the Janni test rest on assumptions (normality of k, equally spaced T_0, counting unit); no raw data were available.
- The Janni paper contains text–table inconsistencies (EPS holder 4/40 is 12.5 % in the text, correctly 10 %; baking sheet k std 1.0×10^{-4} versus 4.7×10^{-5}; the last setup is tap water in one place and deionized in the tables); the test was run with the table values. The authors' 2024 APS abstract reported 136 measurements and 52/63 cases; the paper has 176 and 46/58.
- mW is coarse-grained (no hydrogen bonding/electrostatics); N = 216 and 512 with 12–16 structures per arm; a very fast quench (at target within 100 ps) does not represent real air cooling and is a regime in which the start-temperature difference is erased. The threshold φ* and T_b were chosen with a small pilot (4 structures, T_0 = 320 K). The reference crystal is perfect cubic diamond.
- Two invalid steps were documented: the first N = 1000 rate calculation was wrong (Addendum A1) and the energy-drop criterion produced false triggers (pilot record and pre-registration v2); both were documented before any nucleation result was seen.
- In LJ-like systems there was no effect after an instantaneous quench (N = 108, does not crystallize); this does not show that water is not special.
- The peer-review status of the preprints was not verified.

6. Future work

1. A slow-cooling regime in mW (e.g. 1–10 K/ns), pre-registered. (2) 40+ structures per arm and N = 1000 to narrow the hazard-ratio interval. (3) Calibrating φ* with defective and sub-critical-nucleus references; recording traces also after the confirmation. (4) The same protocol with TIP4P/Ice. (5) Within-trial position/temperature data from Janni et al.

Data and code availability

The code, the pre-registration documents (v1, v2, Addendum A1, A2) with their SHA-256 digests, the outputs and the raw run results (`main_results.jsonl`) are in the folders scripts/, out/ and mw/ of the accompanying code archive. Scripts: sim1_artifact.py, sim1b_gap_sweep.py, sim2_lj_quench.py, test_janni_prediction.py, janni_sensitivity.py, test_janni_extended.py, mwlib2.py, q6lib.py, run_main.py, analyze_main.py.

References

- Burridge, H. C. and Linden, P. F. (2016). Questioning the Mpemba effect: hot water does not cool more quickly than cold. Sci. Rep. 6:37665. https://pmc.ncbi.nlm.nih.gov/articles/PMC5121640
- Janni, A., Botero Ampudia, S. and Dahlberg, E. D. (2026). An experimental investigation of the Mpemba effect. RSC Appl. Interfaces 3:442. https://pubs.rsc.org/en/content/articlehtml/2026/lf/d5lf00296f
- Ghosh, S., Pathak, P., Chatterjee, S. and Das, S. K. (2025). Commun. Phys.; arXiv:2407.06954. https://arxiv.org/abs/2407.06954
- Jin, J. and Goddard, W. A. (2015). J. Phys. Chem. C 119:2622. https://researchportal.ip-paris.fr/en/publications/mechanisms-underlying-the-mpemba-effect-in-water-from-molecular-d/
- Gijón, A. et al. (2019). Phys. Rev. E 100:032103. https://www.doi.org/10.1103/physreve.100.032103

- Teza, G. et al. (2026). Speedups in nonequilibrium thermal relaxation: Mpemba and related effects. Phys. Rep. 1164; arXiv:2502.01758. https://arxiv.org/abs/2502.01758
- arXiv:2503.18820 (preprint); arXiv:2508.05607 (preprint).