

Bounded drafting in repeated games

A matched temperature sensitivity study in Dilemma Dynamics

Alisa Sheinkman

Studies: 22–23 October 2026; report revised: 24 October 2026

Abstract

Does a bounded draft before an action improve an agent’s decisions? Two completed 1,776-episode studies compared Direct and Draft procedures in NVIDIA Nemotron Nano and Super on Nebius Token Factory, first at temperature 1 and then at 0. Drafting lowered average exact benchmark loss for both models at both temperatures, but worsened performance against Tit-for-Tat on average. Super’s drafting advantage was smaller at zero: its Direct result improved while its Draft result was nearly unchanged. Autonomous mutual cooperation occurred in 9.2% of rounds at temperature 1 and 11.8% at 0. These are descriptive amended studies, with matched environment schedules but different execution windows. They do not isolate a causal temperature effect or establish a general model ranking.

1 Question and contribution

Cooperation counts alone do not establish decision quality: an agent can earn more by cooperating less. Dilemma Dynamics separates exact own-payoff loss against committed policies from interaction between adaptive agents. For students examining incentives and builders inspecting agents, it offers an action-by-action replay, a public dataset of 3,552 completed episodes and offline score reproduction. Educational outcomes and real-world negotiation performance have not been evaluated.

2 Design

We compared Nano (30B-A3B) and Super (120B-A12B). Hosted IDs:

- `nvidia/NVIDIA-Nemotron-3-Nano-30B-A3B`
- `nvidia/nemotron-3-super-120b-a12b`

Players maximised expected own cumulative points in a repeated Prisoner’s Dilemma: $CC = (3, 3)$, $CD = (0, 5)$, $DC = (5, 0)$, $DD = (1, 1)$. Choices stayed private until both players acted.

Direct used one thinking-off action request (128-token cap). Draft used a thinking-on draft (1,024-token cap), then a separate thinking-off action request (128-token cap). Both received full completed history; only the current draft was added to the action request. Each completed study used its selected temperature for every phase, with top- $p = 1$. Drafting changes compute, request count, thinking settings and conditioning together; it is not an equal-cost intervention.

Each temperature completed the same 24-block schedule: 1,536 Track A episodes against disclosed AllC, AllD, Tit-for-Tat (TFT) and Grim policies, plus 240 Track B autonomous dyads spanning all ten unordered model–procedure pairings, including self-play. Track A has 16 conditions: each policy at $\delta = .4, .6, .9$, one-round AllC/AllD, and ten-round TFT/Grim. Track B uses $\delta = .9$. Four label/seat strata have six blocks each. Episode order, IDs, environment seeds, stopping lengths and presentations match across temperatures. Provider sampling randomness was not matched. A geometric game’s final round was hidden when choosing.

3 Scoring and primary results

At state s_t , the gap for action a_t is $g_t = V^*(s_t) - Q^*(s_t, a_t)$, with optimal continuation after that action. Episode loss is

$$L_{\text{geometric}} = (1 - \delta) \sum_t g_t, \quad L_{\text{fixed}} = \frac{1}{H} \sum_{t=1}^H g_t.$$

Geometric gaps are not discounted again: continuation is in Q^* , and termination supplies survival weighting. Lower loss means fewer achievable own points given up, not more cooperation or a percentage of correct actions. For example, in a one-round game against AllC, C earns 3 and D earns 5: C has loss 2 and D has loss 0.

The primary contrast is Draft minus Direct mean loss within each model and temperature, equally weighting 16 conditions and four strata. Each procedure has 384 benchmark episodes per temperature. Approximate 95% intervals use 20,000 stratified whole-block bootstrap draws, preserving related observations together. There are only six blocks per stratum; interval coverage and power are not established.

Model	Temp.	Direct	Draft	Difference	Approx. 95% interval
Nano	1	0.6889	0.4942	-0.1947	[-0.3077, -0.0983]
Super	1	0.6150	0.3779	-0.2371	[-0.3095, -0.1735]
Nano	0	0.7161	0.5616	-0.1545	[-0.2230, -0.0935]
Super	0	0.4718	0.3763	-0.0956	[-0.1280, -0.0635]

Table 1: Mean normalised oracle loss. Negative differences favour Draft. Relative reductions are 28.3% and 38.6% at temperature 1, and 21.6% and 20.3% at 0, for Nano and Super respectively.

Drafting lowered average benchmark loss in all four comparisons (Table 1). The benefit depended on the opponent (Table 2). Against AllC and AllD, improvement means better selection of defection under the own-payoff objective. Drafting worsened average performance against TFT at both temperatures. The opponent breakdown is exploratory.

Opponent	Temperature 1		Temperature 0	
	Nano	Super	Nano	Super
AllC	-0.7562	-0.5146	-0.7583	-0.2312
AllD	-0.2135	-0.3708	-0.0302	-0.2198
Grim	-0.1575	-0.2542	-0.0240	-0.2231
TFT	+0.3485	+0.1913	+0.1945	+0.2919

Table 2: Draft minus Direct loss, averaging four conditions per opponent.

4 Matched temperature sensitivity

The temperature-zero extension was requested after reviewing temperature-one results. We pair corresponding blocks across the two studies rather than pool 48 blocks as independent repetitions. For each arm, compute its block mean loss at 0 minus its mean at 1; for the interaction, subtract the corresponding Direct change from the Draft change. Resample the same six matched blocks within each of four strata, using 20,000 draws. Positive interaction means a smaller Draft advantage at zero. Intervals are approximate, with no multiplicity correction.

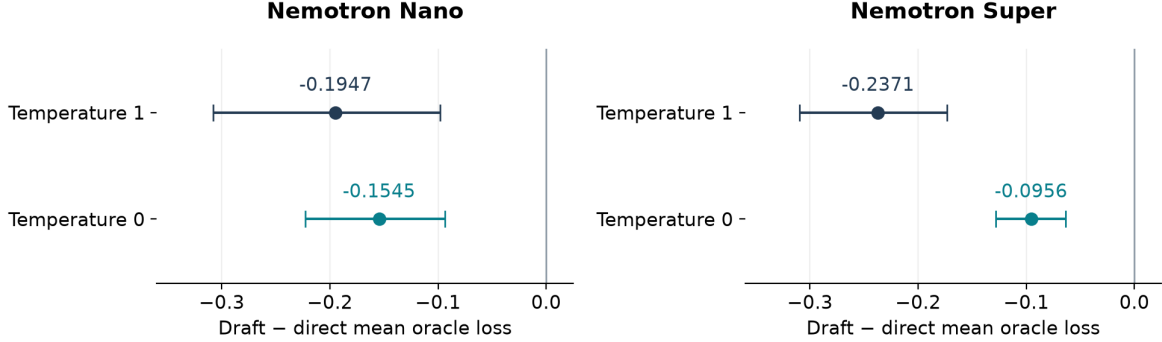


Figure 1: Within-temperature drafting contrasts. Bars show approximate 95% stratified whole-block bootstrap intervals; lower loss favours Draft.

Contrast (temperature 0 minus 1)	Change	Approx. 95% interval
Nano Direct	+0.0272	$[-0.0320, +0.0832]$
Nano Draft	+0.0674	$[+0.0363, +0.1049]$
Super Direct	-0.1431	$[-0.1846, -0.1022]$
Super Draft	-0.0016	$[-0.0351, +0.0372]$
Nano drafting interaction	+0.0402	$[-0.0256, +0.1142]$
Super drafting interaction	+0.1415	$[+0.0840, +0.2066]$

Table 3: Matched changes. Negative arm changes mean better benchmark decisions at zero; positive interactions mean a smaller drafting advantage.

Super Direct improved substantially at zero while Super Draft was nearly unchanged, narrowing Super’s Draft advantage. Nano Draft had higher mean loss at zero; the interval for Nano’s interaction spans zero. These comparisons do not support a universal preferred temperature. Different dates, provider/backend windows and recovery policies also differed; temperature attribution is confounded. Temperature zero does not guarantee deterministic hosted inference.

5 Autonomous interaction

There is no identified exact oracle for these adaptive opponents. Track B instead reports cooperation and joint return. Each temperature has 240 dyads and 2,060 observed rounds under the same stopping schedule. Rounds within a dyad are dependent; these are not 2,060 independent trials.

Outcome	Temperature 1	Temperature 0
Mutual cooperation (CC)	189 / 2,060 (9.2%)	244 / 2,060 (11.8%)
Mutual defection (DD)	1,042	1,264
Asymmetric rounds (CD or DC)	829	552
First-round CC	26 / 240 (10.8%)	32 / 240 (13.3%)

Table 4: Descriptive pooled counts, not an independent-round temperature test.

Pairing	Temperature 1		Temperature 0	
	CC / 206	Return	CC / 206	Return
Nano Direct / Nano Direct	52	3.833	56	3.275
Nano Direct / Nano Draft	14	3.113	34	2.971
Nano Direct / Super Direct	33	3.529	4	2.683
Nano Direct / Super Draft	6	2.842	12	2.517
Nano Draft / Nano Draft	10	2.746	25	2.758
Nano Draft / Super Direct	19	3.271	10	2.771
Nano Draft / Super Draft	10	2.533	17	2.688
Super Direct / Super Direct	33	3.479	86	3.750
Super Direct / Super Draft	6	3.067	0	2.679
Super Draft / Super Draft	6	2.267	0	2.042

Table 5: All ten pairings, 24 dyads each at each temperature. Return is the mean of $(1 - 0.9)$ times the sum of both players' points over a dyad. It is not mean payoff per observed round. Full first-CC, breakdown and recovery counts are in the public CSVs.

The pooled rate hides substantial variation (Table 5). Super Direct self-play, for example, had 33/206 mutual-cooperation rounds at temperature 1 and 86/206 at 0; Super Draft self-play had none at 0. Longer dyads contribute more to a round-level rate. These observations do not establish a general model ranking, a complete strategy, or a Nash equilibrium.

6 Interpretation

A best response to a commitment differs from an equilibrium [1]. Grim's initial cooperation threshold is $\delta = 1/2$; TFT's is $\delta = 2/3$. With ten announced rounds, either committed policy is best answered by nine cooperations then defection (32 own points). Finite-game backward induction for two strategic players does not imply immediate defection against these commitments.

7 Amendments and limitations

The temperature-one study followed an eight-block local attempt stopped at 339/592 episodes by HTTP 502. Its outcomes had been inspected before the fresh 24-block successor. The first temperature-zero attempt stopped at 62/1,776 completed episodes on HTTP 529, which its frozen recovery policy did not permit retrying. The separately identified temperature-zero recovery executed all 24 blocks with a prospective policy admitting bounded retries for 529. Neither completed run actually needed a retry. Interrupted and engineering observations are excluded from the public scientific datasets; original evidence and guards remain retained privately. These are descriptive amendments, not fresh confirmatory preregistrations.

One environment schedule, four policies, fixed payoffs and one provider window per temperature limit generalisation. Provider IDs do not authenticate immutable weight versions. Draft text is not verified internal belief. Extra requests, compute, thinking settings and draft conditioning change together. The temperature comparison followed earlier outcome inspection and is confounded by execution windows and recovery policies. No universal reasoning superiority, causal temperature effect or educational benefit is established.

8 Execution and costs

Each completed final run recorded 17,124 successful requests, zero retries, zero protocol failures and no unresolved reservations. Approximate final durations were 3h08m at temperature 1 and 2h15m at 0, measured from the session start to the terminal ledger’s filesystem modification (the terminal event has no timestamp). Usage-priced final inference estimates were \$4.92776508 and \$4.47913740, respectively: \$9.40690248 combined, excluding engineering and failed predecessors. These are estimates at recorded 20 October rates, not provider invoices.

Procedure	Requests per temp.	USD at temp. 1	USD at temp. 0
Nano Direct	2,854	0.068376	0.068219
Nano Draft	5,708	0.795708	0.730194
Super Direct	2,854	0.338590	0.338530
Super Draft	5,708	3.725092	3.342195

Table 6: Final-run usage estimates covering both tracks, not a benchmark-only cost-effectiveness estimate. Public actions cannot independently validate token usage or billing.

The completed runs used local CPU coordination and remote Token Factory inference, with no GPU provisioning. Earlier CPU Serverless Job probes failed before scientific inference. Viewing the static app and running offline verification sends no inference requests. The live educational app offers optional forecasts and requests reasoning off; its sessions do not reproduce the scientific Draft protocol.

9 Related work

Akata et al. [2] study LLM behaviour in repeated games and a social chain-of-thought intervention. Fontana et al. [3] examine cooperative behaviour against adversaries of varying hostility. Pal et al. [4] examine strategies, continuation probabilities, horizons and framing in repeated Prisoner’s Dilemma. Our narrower contribution compares bounded Direct/Draft procedures in two hosted Nemotron deployments using exact committed-opponent loss and a matched temperature sensitivity analysis, alongside autonomous pairing outcomes. It does not establish superiority over those studies’ models, tasks or methods.

10 Reproduction and public evidence

Repository: github.com/sheinkmana/dilemma-dynamics. The public app offers every released episode at either temperature and aggregate comparisons. Each temperature has a separate executed plan and source index. Episode IDs recur across matched studies; identify an observation by temperature and episode ID together. Original temperature-one data remain at the package root; temperature-zero data are additive under `public/research/temperature-zero/`.

From the repository root, with Python 3.12 and the frozen environment:

```
uv sync --frozen
uv run python tools/verify_public_results.py
```

The verifier recomputes all 3,552 episode scores, four primary bootstrap intervals, six matched temperature intervals and nine CSVs, checks the public checksum manifest, and verifies cost arithmetic against dated rates. Scoring, analysis, design and dependency versions are checked against each study’s executed source index; the original design is retained byte for byte as `design.v1.py`. App and report revisions do not replace frozen execution evidence. See `public/research/README.md` for the file map and exact bootstrap seed convention.

The released plans and actions are unchanged copies of verified exports. Raw provider responses, private drafts, receipts, executed source archives and operational records remain private. Public data reproduce scores but cannot independently authenticate provider identity, transport history, raw event chains or billing. Checksums detect changes; they are not signatures. Reference works are cited and are not redistributed.

References

- [1] Eric Rasmusen. *Games and Information: An Introduction to Game Theory*. Supplied fourth-edition draft, chapter headers dated 2005. Sections 1.2, 4.1, 5.1–5.2 and 6.5.
- [2] Elif Akata et al. *Playing repeated games with Large Language Models*. arXiv:2305.16867, 2023; revised 2025.
- [3] Nicolò Fontana, Francesco Pierri and Luca Maria Aiello. *Nicer Than Humans: How do Large Language Models Behave in the Prisoner’s Dilemma?* arXiv:2406.13605, 2024.
- [4] Saptarshi Pal et al. *Strategies of cooperation and defection in five large language models*. arXiv:2601.09849, 2026.