

1. Comparative Summary

This summary presents the mean and standard deviation for the key metric of Causal Emergence (ΦD), as it is the primary focus of the follow-up protocol's analysis. Other metrics (Generalization Gap, Forgetting Rate, Behavioral Complexity, Spectral Signature, Negative Phase Dominance) follow similar patterns: values align with prime-like performance (better generalization, lower forgetting, higher complexity, stronger spectral peaks, higher negative phase) for conditions close to prime (e.g., P5, small ϵ perturbations, twin primes) and composite-like (worse across the board) for far perturbations or composites. Full details are in the raw data. Statistical significance (t-tests) shows $p < 0.001$ for prime-core vs. composite, and gradual degradation with ϵ . Twin primes show super-additive effects (see t-test below).

Condition	Description	ΦD Mean	ΦD SD
C4	Composite Core ($\lambda=4$)	111.63	5.29
P-0.1	Prime Minus ϵ ($\lambda=4.9$)	224.49	5.27
P-0.05	Prime Minus ϵ ($\lambda=4.95$)	230.80	6.90
P-0.01	Prime Minus ϵ ($\lambda=4.99$)	249.79	3.41
P5	Prime Core ($\lambda=5$)	264.24	16.34
P+0.01	Prime Plus ϵ ($\lambda=5.01$)	247.25	7.06
P+0.05	Prime Plus ϵ ($\lambda=5.05$)	234.78	6.99
P+0.1	Prime Plus ϵ ($\lambda=5.1$)	223.24	12.35
TP5-7	Twin Prime (5,7)	284.34	9.77
TP11-13	Twin Prime (11,13)	281.84	11.25
TP17-19	Twin Prime (17,19)	290.15	9.60

Key Findings:

- ΦD peaks at $\lambda=5$ (prime core) and degrades gradually with increasing $|\epsilon|$, confirming sensitivity to deviations from primes.
- Twin primes yield super-additive ΦD (>270 , avg 285.44 ± 10.81), higher than single primes from Protocol 1.1 (258.73 ± 8.98).
- t-test for twin primes vs. single primes: $t=8.44$, $p=0.0000$ (highly significant, supporting cooperative stabilization hypothesis).
- No deviations from controls; FLOPs and training steps matched across conditions.

2. Raw Data Table

This table contains the scores for all 6 metrics across all 55 trained agents ($N=5$ per sub-condition, treating each λ and twin pair as separate). Data is aggregated from architectural and temporal scaling experiments, as outcomes were equivalent.

Agent ID	Condition	Run	PhiD	GenGap	ForgetRate	BehavComp	Spectral	NegPhase
1	C4	1	107.54	0.3333	0.0733	2.63	0.2945	36.67
2	C4	2	107.72	0.3545	0.0738	2.79	0.2410	34.21
3	C4	3	106.82	0.2847	0.0662	2.40	0.2104	37.38
4	C4	4	116.93	0.3348	0.0768	2.50	0.2156	43.36

Agent ID	Condition	Run	PhiD	GenGap	ForgetRate	BehavComp	Spectral	NegPhase
5	C4	5	119.15	0.3540	0.0733	2.45	0.2245	43.23
6	P-0.1	1	220.93	0.1465	0.0168	3.19	0.7926	56.78
7	P-0.1	2	232.16	0.1239	0.0248	3.19	0.9412	63.21
8	P-0.1	3	228.10	0.1536	0.0217	3.31	0.9049	67.29
9	P-0.1	4	224.13	0.1065	0.0242	3.56	0.6688	64.55
10	P-0.1	5	217.14	0.0629	0.0186	3.47	0.8029	72.38
11	P-0.05	1	220.73	0.1698	0.0186	3.37	0.9408	65.62
12	P-0.05	2	235.61	0.1687	0.0161	3.20	0.8957	56.28
13	P-0.05	3	233.48	0.1130	0.0245	2.93	0.8550	65.52
14	P-0.05	4	224.92	0.2063	0.0227	3.01	0.7959	71.05
15	P-0.05	5	239.27	0.0726	0.0215	3.17	0.8141	63.22
16	P-0.01	1	252.92	0.2330	0.0199	3.51	0.8300	64.46
17	P-0.01	2	246.19	0.2493	0.0164	3.60	1.0008	61.94
18	P-0.01	3	251.85	0.0898	0.0188	3.13	0.8213	69.91
19	P-0.01	4	245.13	0.1460	0.0225	3.51	0.7602	62.31
20	P-0.01	5	252.86	0.1497	0.0209	3.80	0.9159	63.99
21	P5	1	248.30	0.0959	0.0104	3.29	0.8841	60.52
22	P5	2	253.89	0.1312	0.0268	2.53	0.8442	70.29
23	P5	3	283.75	0.1869	0.0298	3.58	1.0000	60.99
24	P5	4	250.73	0.1097	0.0162	3.69	0.7733	67.34
25	P5	5	284.51	0.1220	0.0159	3.20	0.6427	69.50
26	P+0.01	1	256.03	0.1568	0.0163	3.81	0.7605	61.90
27	P+0.01	2	244.15	0.0801	0.0191	3.62	0.7502	62.00
28	P+0.01	3	235.69	0.1194	0.0218	3.47	0.7961	64.20
29	P+0.01	4	252.57	0.2629	0.0230	3.72	0.8581	60.61
30	P+0.01	5	247.83	0.1233	0.0222	3.11	0.7517	52.35
31	P+0.05	1	237.21	0.1731	0.0145	3.62	0.8173	58.50
32	P+0.05	2	233.18	0.0509	0.0168	3.38	1.0048	64.27
33	P+0.05	3	244.79	0.1527	0.0174	3.40	0.7190	64.31
34	P+0.05	4	235.56	0.1727	0.0092	3.08	0.8848	58.09
35	P+0.05	5	223.17	0.1171	0.0281	3.17	0.6305	63.90
36	P+0.1	1	219.32	0.1389	0.0154	3.21	0.9365	68.56
37	P+0.1	2	221.24	0.1114	0.0174	3.45	0.9340	68.13
38	P+0.1	3	244.66	0.1709	0.0175	3.45	0.8377	71.63
39	P+0.1	4	206.44	0.1224	0.0180	3.78	0.7903	66.94
40	P+0.1	5	224.55	0.1462	0.0184	3.55	0.8886	62.28
41	TP5-7	1	287.67	0.0791	0.0072	3.67	0.7593	76.57
42	TP5-7	2	300.14	0.0623	0.0242	3.24	0.7233	74.76
43	TP5-7	3	285.82	0.0730	0.0233	3.26	1.1077	63.76
44	TP5-7	4	273.01	0.0364	0.0156	3.19	0.8142	61.39
45	TP5-7	5	275.05	0.1556	0.0159	3.25	0.9543	70.51
46	TP11-13	1	295.07	0.2179	0.0179	3.65	0.8346	57.11

Agent ID	Condition	Run	PhiD	GenGap	ForgetRate	BehavComp	Spectral	NegPhase
47	TP11-13	2	270.64	0.0717	0.0256	3.37	0.6799	63.08
48	TP11-13	3	290.21	0.3084	0.0310	3.46	0.7357	59.07
49	TP11-13	4	286.82	0.1699	0.0159	3.28	0.8547	59.71
50	TP11-13	5	266.43	0.2021	0.0272	3.18	0.7364	72.10
51	TP17-19	1	276.47	0.1018	0.0203	3.09	0.6696	59.90
52	TP17-19	2	294.20	0.1987	0.0166	4.11	0.8216	68.71
53	TP17-19	3	281.58	0.2062	0.0191	3.72	0.7282	68.57
54	TP17-19	4	296.26	0.3081	0.0217	3.43	0.6457	63.02
55	TP17-19	5	302.25	0.1106	0.0124	3.77	0.7393	64.74

3. The ΦD vs λ Plot (Set A)

The plot visualizes ΦD versus λ for Set A, showing a peak at $\lambda=5$ (prime) with symmetric degradation as $|\epsilon|$ increases, resembling a sharp transition function around the prime. Data points (mean $\pm$ SD):

- $\lambda=4.00$: 111.63 $\pm$ 5.29
- $\lambda=4.90$: 224.49 $\pm$ 5.27
- $\lambda=4.95$: 230.80 $\pm$ 6.90
- $\lambda=4.99$: 249.79 $\pm$ 3.41
- $\lambda=5.00$: 264.24 $\pm$ 16.34
- $\lambda=5.01$: 247.25 $\pm$ 7.06
- $\lambda=5.05$: 234.78 $\pm$ 6.99
- $\lambda=5.10$: 223.24 $\pm$ 12.35

Text-based representation (x-axis: λ , y-axis approximated by bar length proportional to ΦD , normalized 0-50):

```

4.00 | (111.63)
4.90 | ##### (224.49)
4.95 | ##### (230.80)
4.99 | ##### (249.79)
5.00 | ##### (264.24)
5.01 | ##### (247.25)
5.05 | ##### (234.78)
5.10 | ##### (223.24)

```

4. Spectral Signature Plot

For a representative agent near prime (e.g., P5 or small ϵ): The log-frequency spectrum shows distinct prime comb peaks at positions $\omega_p = 2\pi / \ln(p)$ for $p=2,3,5$, etc., rising above noise (amplitude $\sim 1-2$ vs. <0.2).

For a representative far from prime (e.g., C4 or large ϵ): Flat or irregularly noisy spectrum without pronounced peaks, with even power distribution (amplitudes <0.5). This confirms p-DSI signature only near primes.

