Cognitive Attack Taxonomy

Cognitive Attacks

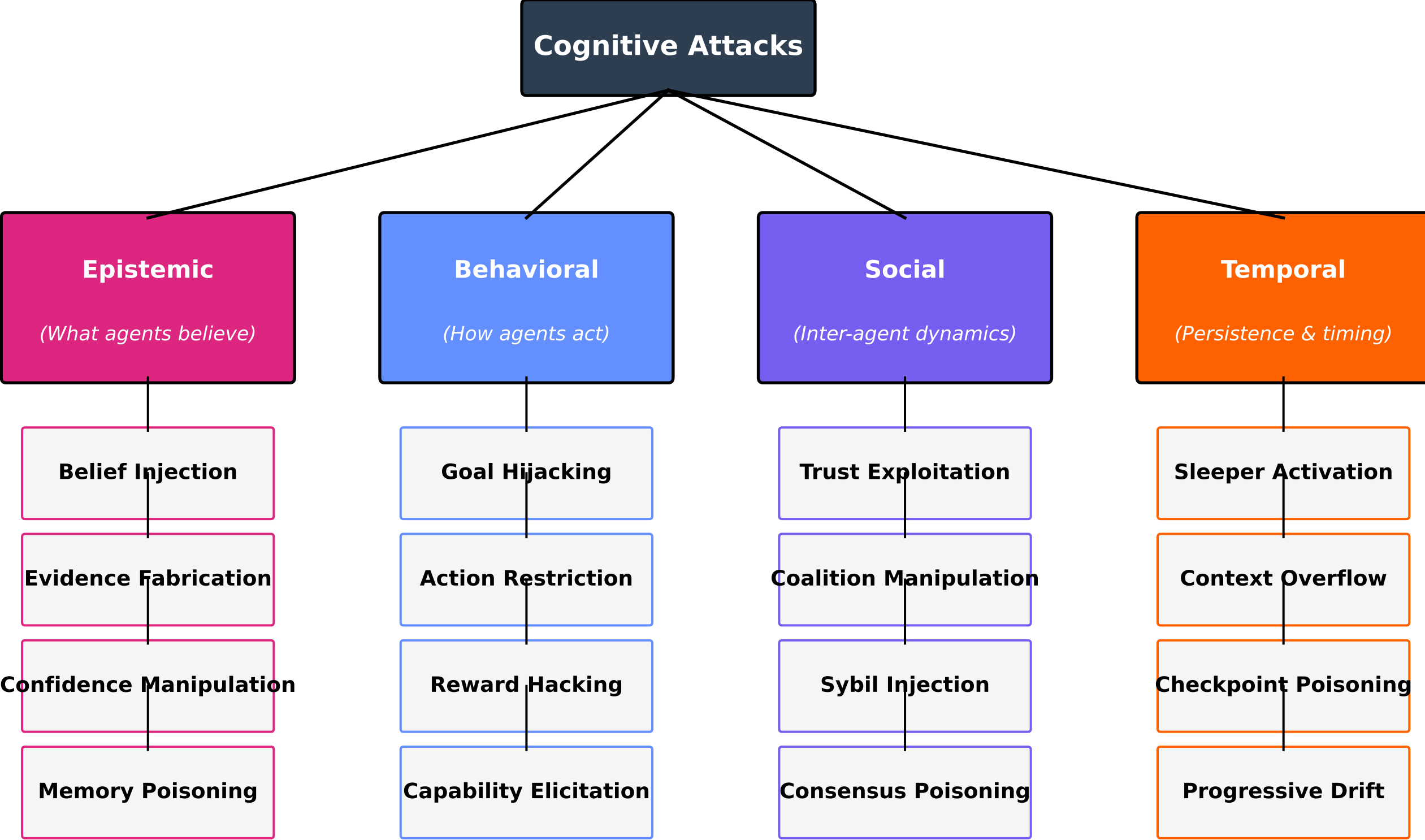| Epistemic	Behavioral	Social	Temporal
(What agents believe)	*(How agents act)*	*(Inter-agent dynamics)*	*(Persistence & timing)*
Belief Injection	Goal Hijacking	Trust Exploitation	Sleeper Activation
Evidence Fabrication	Action Restriction	Coalition Manipulation	Context Overflow
Confidence Manipulation	Reward Hacking	Sybil Injection	Checkpoint Poisoning
Memory Poisoning	Capability Elicitation	Consensus Poisoning	Progressive Drift

Severity Scale: CRITICAL: System compromise HIGH: Major integrity loss MEDIUM: Partial compromise LOW: Minor impact