Commonsense Psychology

In FN3, the structure of generic scenarios related to human cognition is based on the book **A Formal Theory of Commonsense Psychology - How People Think People Think, Andrew S. Gordon and Jerry R. Hobbs (2017)**. This texts presents a summary of the main ideas used to create the scenarios discussed on next chapters.

Introduction: The Objectives of Commonsense Psychology

Commonsense Psychology (CSP), as articulated by Gordon&Hobbs, aims to formalize the implicit understanding that ordinary people have about how the mind works and how mental states interact with actions and the world. This endeavor is not about proposing new psychological theories, but rather about explicating the foundational, unstated assumptions we all operate with in daily life when we attribute beliefs, desires, intentions, and other mental states to ourselves and others.

The primary objective is to create a computationally tractable model of this everyday psychological reasoning. Such a model is crucial for various applications, particularly in Artificial Intelligence (AI) and Natural Language Processing (NLP), where systems need to understand and generate human-like behavior, interpret natural language, and engage in meaningful interaction. Without a formalized understanding of commonsense, AI systems struggle with context, nuance, and the basic motivations behind human actions and communication. CSP work seeks to lay bare the "axioms" of human commonsense about the mind, much like physics describes commonsense about the physical world.

CSP Method: Formalization in First-Order Logic

CSP's distinctive method for formalizing commonsense psychology is rooted in **abductive inference** and **first-order logic**. He posits that much of human understanding and interpretation, especially of language, involves abductively inferring the most plausible explanation for observed phenomena (e.g., why someone said something, or why an event occurred).

This approach involves:

- **Axiomatization:** Expressing common-sense knowledge as a set of logical axioms and predicates. These axioms are fundamental truths or rules that govern the relationships between various concepts. For instance, an axiom might state that if an agent perceives something, they come to believe it.
- **Abduction as Inference:** Using abduction as the primary mode of reasoning. In abduction, if we have a set of axioms (P Q) and an observed phenomenon (Q), we can hypothesize P as an explanation. This contrasts with deduction (P Q, P, therefore Q) and induction (observing many Q after P, inferring P Q). CSP frames understanding as finding the best

abductive explanation for observed facts, such as interpreting an utterance by inferring the speaker's intentions and beliefs that would make the utterance relevant.

- **Minimal Cost Abduction:** When multiple explanations are possible, abduction seeks the "best" explanation, often defined as the one that involves assuming the fewest new, unproven propositions (i.e., minimal cost). This principle guides disambiguation and plausible inference in complex situations.

This logical framework allows for precise, unambiguous representation of commonsense notions, enabling computational systems to "reason" with them. The focus is on capturing the *implications* and *connections* between concepts that underlie our everyday inferences, rather than building a comprehensive ontology of every possible mental state.

Main Concepts and Their Interrelations

CSP explores several core concepts in commonsense psychology, defining them through their relationships with each other, forming a coherent, interconnected framework.

Belief and Knowledge

- **Belief (believe):** This is a foundational mental predicate. To "believe" something means an agent holds a proposition (a state of affairs or a statement) to be true. Beliefs are the agent's internal model of the world, shaping their understanding and informing their actions. Beliefs can be formed through perception, communication, or inference.
- **Knowledge (know):** Knowledge is formalized as a special, stronger form of belief. If an agent "knows" a proposition, it implies that they believe it, and that proposition is also objectively true (within the system's defined reality). The distinction often lies in justification or reliability.
- **Relation:** Knowledge is a subset of belief. `know(A, P) => believe(A, P)`. Beliefs are constantly being formed and revised based on new information.

Desire and Intention

- **Desire (want):** Desire represents an agent's motivational state – a wish or preference for a certain proposition or state of affairs to be true in the future. Desires drive an agent's goals and provide the impetus for action. Unlike beliefs, desires are not about what *is* true, but what the agent *wants* to be true.
- **Intention (intend):** Intention is a stronger commitment than mere desire. An agent "intends" to do something when they have a desire to achieve a goal and commit to performing a specific action or series of actions to achieve it. Intention links desires directly to action and planning.

- **Relation:** `intend(A, E)` (Agent A intends event E) implies `want(A, occur(E))` (A wants E to occur). Intentions are desires coupled with a commitment to act, guided by beliefs about the feasibility of actions.

Planning and Action

- **Planning (plan):** Planning is a cognitive process where an agent, holding an intention or desire for a future state, mentally constructs a sequence of actions believed to lead to that state. This involves reasoning about preconditions (what must be true before an action), effects (what will be true after an action), and the logical flow of events. A plan is essentially a method or procedure.
- **Action (act):** An action is an event caused by an agent, typically performed with an intention to achieve a desired outcome. Actions are the means by which agents directly interact with and change the world to fulfill their intentions and desires.
- **Relation:** `intend(A, E)` often leads to `plan(A, E')` (where E' is the set of actions constituting E). The execution of a `plan` involves performing a sequence of `actions`. Actions have `effects` which change the state of the world and are crucial for successful planning.

Perception

- **Perception (perceive):** Perception is the process by which an agent directly acquires information about the state of the world through sensory input. It's the primary channel through which external reality influences an agent's internal mental states.
- **Relation:** A fundamental axiom of commonsense psychology is that if an agent perceives a proposition to be true, they come to believe that proposition. Thus, `perceive(A, P) => believe(A, P)`. Perception is crucial for updating `beliefs` and providing the necessary information for `planning` and `action`.

Emotion

- **Emotion (emotion):** CSP addresses emotions as internal states of an agent that are often triggered by `beliefs` and `desires`. Emotions can influence an agent's `reasoning` and `actions`, and can contribute to longer-lasting "moods". While not as deeply formalized as other concepts, their causal role in psychological life is acknowledged.
- **Relation:** `emotion(A, Type, Cause)` implies that a `Cause` (often a `Belief` or a `Desire` state) leads to an `Emotion` of a certain `Type` in `Agent` A. Emotions can affect `desires`, `intentions`, and subsequent `actions`.

Communication

- **Communication (communicate):** This involves one `agent` (the speaker) performing an `action` (an utterance) with the `intention` of causing another `agent` (the hearer) to come to `believe` a certain `proposition` (the message). Effective communication often relies on `shared context` and `abductive inference` by the hearer to understand the speaker's intent.
- **Relation:** Communication is an `action` driven by `intention`. Its success is measured by the change in the hearer's `beliefs`. Specific types of communication, like promising (a `speech act`), can create `obligations` between agents.

Causation

- **Causation (cause):** CSP views causation as a fundamental relationship where one event or state of affairs (the cause) brings about another (the effect). This is critical for understanding how `actions` produce results and for `planning` (predicting the effects of planned actions).
- **Relation:** `cause(E1, E2)` means event E1 leads to event E2. `Agentive actions` are often `causes` of changes in the world. `Preconditions` for actions are states that must `cause` the action to be possible.

Time and Space

- **Time:** All `events`, `processes`, and `states` occur over time. CSP implicitly incorporates temporal reasoning, as `actions` have durations and sequences, and `causal` relationships imply temporal precedence.
- **Space:** Similarly, `entities` and `agents` exist in `locations`, and `actions` occur at specific places. Spatial relations provide the fundamental physical context for `perception` and `action`.
- **Relation:** `Time` and `Space` are implicit but essential dimensions for grounding `actions`, `perceptions`, and `causal` chains.

Mind-Body Interaction

- **Mind-Body Interaction:** While not a separate concept, CSP inherently models the interaction between mental states and the physical world. `Perception` is the link from the body/environment to the mind (physical input leads to mental `beliefs`). `Action` is the link from the mind to the body/environment (mental `intentions` lead to physical `behavior`).
- **Relation:** This forms a continuous feedback loop: `Perception` informs `Beliefs`, `Beliefs` influence `Desires` and `Intentions`, `Intentions` drive `Actions`, and `Actions` change the `Environment`, leading to new `Perceptions`.

Social Interaction and Obligation

- **Social Interaction / Obligation:** CSP extends commonsense psychology to social dynamics, particularly focusing on how communicative acts create social commitments. A `promise`, for instance, is a `speech act` that creates an `obligation` for the `speaker` to perform a future `action`, which the `hearer` then `believes` will occur.
- **Relation:** These concepts build upon `Communication`, and involve `beliefs`, `intentions`, and `actions` in an inter-agent context.

Coherent Framework: How Concepts Interrelate

CSP forms a tightly knit, coherent framework where concepts are not isolated but mutually defined and interlinked through axioms. The core of this coherence lies in the **Belief-Desire-Intention (BDI) model** of agency, implicitly formalized through his axioms:

- **Beliefs** provide the agent's model of the world (what is true).
- **Desires** provide the agent's goals and motivations (what is wanted).
- **Intentions** bridge the gap between desires and action, representing a commitment to achieve a desire by acting (what the agent commits to do).

This core BDI cycle is constantly informed by **Perception** (updating beliefs from the world) and realized through **Planning** and **Action** (acting on beliefs and intentions to change the world). **Causation** provides the fundamental mechanism for how actions lead to effects. **Communication** offers a means for agents to influence each other's beliefs and intentions, leading to complex **Social Interactions** and **Obligations**. Implicitly, **Memory** provides the persistence of beliefs and knowledge over time, allowing for learning and informed decision-making. **Emotions**, while less central to the logical structure, are acknowledged as influencing desires, beliefs, and actions. Finally, all these phenomena are grounded in **Time** and **Space**.

The strength of CSP framework for FrameNet Brasil lies in its systematic articulation of these interdependencies. By mapping these core commonsense psychological concepts to FN3 scenarios, we can build a robust, interconnected semantic network that better reflects how humans understand and interact with their world.

What this book is doing (in plain English)

- **Claim:** Humans already use a rich, informal theory to explain people in terms of **beliefs, goals, plans, and emotions**. If we want humanlike AI (or just clearer thinking), we need to **make that theory explicit and formal** in logic.

- **Deliverable (their side):** a large, first-order logical formalization: roughly **1,400 axioms**, arranged as **29 Commonsense Psychology theories** built on **16 "background" theories** (time, causality, etc.).

- **Method:** "**Successive formalization**" – get **breadth first** (cover everything you'll need), then iteratively harden it into rigorous axioms and make the modules consistent across the whole system.

- **Motivation:** People unavoidably **anthropomorphize** systems. If computers are going to live in our world, they should at least **behave in ways that line up with our commonsense expectations** about minds— or understand those expectations well enough to communicate around them.

The scaffolding they build on (Part II: 16 background theories)

These are the nuts-and-bolts modules that any psychological theory needs. In their system: **Eventualities & event structure**, **Time** (instants/intervals/durations), **Causality** (agents, ability, executability, difficulty), **Change of state**, **Space**, **Persons**, **Modality** (possibility/necessity/likelihood), **Logic reified** (reasoning about statements), **Defeasibility** (default reasoning), **Scales** (qualitative/ordered quantities), **Arithmetic** (measures/proportions), **Functions & sequences**, **Composite entities** (part–whole), **Traditional set theory**, and **Substitution / Typical elements / Instances**.

Why this matters: every "mind" predicate—believes, intends, remembers, fears—needs time, causality, scales (e.g., intensity, importance), and defeasible inference to make the everyday inferences we all casually make.

The psychology proper (Part III: 29 theories)

Here are the cores and how to think about them:

Representation & inference about minds

- **Knowledge Management:** objects of belief; belief vs. knowledge; degrees of belief; assuming; focus/attention; inference & justification; mutual belief.

- **Similarity Comparisons:** how we judge two structured things "alike."

- **Memory:** storing/retrieving; accessibility; associations; remembering/forgetting; prospective memory ("remembering to do").

- **Envisioning:** "thinking of," causal systems, mental simulation, how envisionment interacts with belief.

- **Explanation:** what counts as an explanation, when it fails, the process of moving from mystery → hypothesis.

- **Managing Expectations:** priors, surprise, and norm violations.

- **Other-Agent Reasoning:** tracking others' goals/beliefs.

Goal–plan–action pipeline

- **Goals** (and **Goal Themes** like thriving, pleasure/pain, short- vs long-term).

- **Threats & Detection** (what counts as a threat, how serious, managing it).

- **Plans** as mental entities; **Plan Elements**; **Planning Modalities** (counterfactual, hypothetical); **Planning Goals** (constraints, preferences, enabling/blocking, minimizing/maximizing, locating instances); **Plan Construction** and **Plan Adaptation**.

- **Design** (artifacts, designing).

- **Decisions** (choice sets, deliberation, justifications, consequences).

- **Scheduling** (simultaneity, calendars, pending/scheduled plans, preferences).

- **Monitoring** (watching processes, characteristics like periodicity).

- **Execution Modalities** and **Execution Control** (start/stop, progress, costs, outcomes, abstraction/instantiation, aspect, distraction).

- **Execution Envisionment** (mentally simulating success/failure).

- **Causes of Failure** (taxonomy, causal complexes, explanation patterns).

- **Repetitive Execution** (iteration patterns).

- **Mind–Body Interaction** (perception, bodily action, levels of capability/activity, consciousness).

- **Observation of Plan Executions** (instructions, performances/specs, skill, evaluation).

- **Emotions** (general structure, intensity/arousal, happiness/sadness and "shades," "raw" emotions, cognitively elaborated emotions like hope/fear/joy after success, envy/jealousy, liking/disliking, appraisal & coping).

Their presuppositions & design choices (the "how")

- **Stick with first-order logic**, not modal slogans—then reify the right bits (events, propositions) so the system can **talk about its own beliefs**, justifications, etc.

- **Breadth first**: cover the whole day-in-the-life mind pipeline (belief → goal → plan → schedule → execution → monitoring → explanation → adaptation), even if early axioms are rough.

- **Defeasible reasoning everywhere**: normal human inferences are **default** and break under new info.

- **Quantitative hooks** via **scales & arithmetic**: importance, cost, difficulty, intensity are **ordered**, sometimes numeric, so you can compare and reason about tradeoffs.

- **Time & event structure** are first-class: states, events, subevents, iterations, aspect—because every mental predicate plays out over time and in parts.

How all the pieces interconnect (and your diagram)

I mapped dependencies from the 16 background theories into the 29 commonsense-psych modules, plus the key flows among the psychology modules (e.g., **Goals → Plans → Execution → Observation/Monitoring → Explanation/Adaptation**). You can open it and zoom around:

Download the concept map (PNG)

Quick legend:

- **Left column (squares):** background theories (time, causality, scales, etc.).

- **Right column (circles):** psychology theories.

- Arrows show "supports" or "feeds into" relations (e.g., **Causality → Explanation / Threats / Emotions**; **Time → Scheduling / Execution Control**).

- Internal arrows on the right show the mind-pipeline (e.g., **Plans → Execution Envisionment → Execution Modalities → Execution Control**, and **Causes of Failure → Plan Adaptation**).

TL;DR of the big ideas

- Humans share a **robust, tacit folk theory** for explaining behavior; it's not "cute"—it's operational, and we use it constantly.

- To make software truly human-compatible, we need a **formal, machine-readable version** of that theory that lines up with how people **think people think**.

- That means **lots of meticulous axioms** covering everything from **belief and memory** to **scheduling and emotions**, all grounded in **time/causality/defeasibility/scales**.

- Do **coverage first**, then harden and reconcile the logic so the modules compose.