

A Formal Reconstruction of Lewis’s Theory of Common Knowledge

The Cubitt–Sugden Syntactic Approach in Lean 4

Abstract

This paper presents a machine-verified formalization of Cubitt and Sugden’s (2003) syntactic reconstruction of David Lewis’s theory of common knowledge. Using the Lean 4 proof assistant, we establish Lewis’s theorem in full generality: that reflexive common indicators generate an infinite hierarchy of higher-order reasons to believe. The formalization makes explicit the axiomatic foundations required for Lewis’s informal argument, particularly the crucial axioms A1 (modus ponens for indication) and A6 (transitivity through shared standards). We prove not only the first few levels of nested belief but the complete infinite hierarchy via structural induction.

1 Introduction

David Lewis’s (1969) theory of common knowledge addresses a fundamental puzzle in social epistemology: how can a population come to share infinitely nested mutual expectations from finite cognitive resources? Lewis’s answer appeals to “reflexive common indicators”—states of affairs that indicate not only some target proposition but also their own recognition by all members of a population.

Cubitt and Sugden (2003) provided the first rigorous reconstruction of Lewis’s informal argument, making explicit the logical assumptions required for the proof. This paper presents a machine-verified formalization of their approach in Lean 4, extending their work by proving Lewis’s theorem for the complete infinite hierarchy rather than just the first few levels.

2 Primitive Operators

The formalization employs two primitive relations that capture Lewis’s key concepts without internal structure.

2.1 Reason to Believe

The expression Rip represents “individual i has reason to believe proposition p .” Following Cubitt and Sugden’s interpretation, this means p is derivable within the logic of reasoning that i endorses. Crucially, this is normative rather than descriptive: it captures what i should believe given their evidence and logical standards, not what they actually believe.

2.2 Indication

The expression $IndAip$ represents “state of affairs A indicates to i that p .” Lewis defines indication counterfactually: A indicates p to i if, were i to have reason to believe A holds, i would *thereby* have reason to believe p . The word “thereby” suggests that the reason for p depends on the reason for A —a dependence that the primitive treatment cannot represent internally.

3 Core Axioms

Two axioms govern the interaction between the primitive operators. These must be assumed rather than derived because R and Ind lack internal structure.

Axiom 1 (A1: Modus Ponens for Indication). *If A indicates p to i , and i has reason to believe A holds, then i has reason to believe p :*

$$Ind A i p \rightarrow R i A \rightarrow R i p$$

This is Lewis's "principle of detachment," essential at every step of his argument.

Axiom 2 (A6: Transitivity Through Shared Standards). *If A indicates to i that j has reason to believe A , and i has reason to believe that A indicates u to j , then A indicates to i that j has reason to believe u :*

$$Ind A i (R j A) \wedge R i (Ind A j u) \rightarrow Ind A i (R j u)$$

This captures agents' ability to reason about each other's reasoning under shared inductive standards—what Lewis informally calls "ancillary premises about rationality, inductive standards, and background information."

4 Conditions for Reflexive Common Indicators

A state of affairs A is a *reflexive common indicator* in population P that φ if the following four conditions hold:

C1 (Publicity): Every member of P has reason to believe A holds. Formally: $\forall i, R i A$.

C2 (Mutual Awareness): A indicates to each member that every other member has reason to believe A . Formally: $\forall i j, Ind A i (R j A)$.

C3 (Content): A indicates φ to every member. Formally: $\forall i, Ind A i \varphi$.

C4 (Shared Standards): If A indicates u to i , then i has reason to believe A indicates u to j . Formally: $Ind A i u \rightarrow R i (Ind A j u)$.

5 The Iteration of Reasons to Believe

The central result shows how these conditions generate an infinite hierarchy of nested reasons to believe.

5.1 First-Order (L1)

Every individual has reason to believe φ . The proof is immediate: by C1, i has reason to believe A ; by C3, A indicates φ to i ; by A1, i has reason to believe φ .

5.2 Second-Order (L2)

Every individual i has reason to believe that every individual j has reason to believe φ . Here the proof requires A6: from i 's knowledge that A indicates φ to j (via C3 and C4), together with C2's guarantee that A indicates j 's recognition of A , we derive that A indicates to i that $R j \varphi$.

5.3 Higher Orders

The pattern continues: L3 establishes that every individual has reason to believe that every other has reason to believe that every other has reason to believe φ , and so on. Cubitt and Sugden prove the first few levels and note that “the pattern continues.” Our formalization makes this precise.

6 The General Theorem

Rather than proving each level separately, we define the R -closure of φ inductively and prove that every proposition in this closure is believed by everyone.

Definition 1 (R -Closure). *The R -closure of φ is the smallest set containing φ and closed under the operation “ j has reason to believe.” It contains: φ , $Rj\varphi$, $Ri(Rj\varphi)$, $Rk(Ri(Rj\varphi))$, and so on for all individuals and all finite nestings.*

Lemma 2 (Key Lemma). *If q is in the R -closure of φ , then A indicates q to every individual. The proof proceeds by structural induction: the base case uses C3; the inductive step applies C4 and A6 to lift indication from u to Rju .*

Theorem 3 (Lewis’s Theorem). *For any proposition p in the R -closure of φ , every individual has reason to believe p . The proof combines the key lemma with A1 and C1: since A indicates p to i and i has reason to believe A , by modus ponens i has reason to believe p .*

7 Philosophical Significance

This formalization vindicates Lewis’s claim that his account avoids standard objections to common knowledge:

No unbounded reasoning required. The proof uses only two inductive steps, not infinitely many mental operations.

No mind-reading required. Reasoning concerns the publicly observable state A , not private mental states.

Explains genesis. Shows how common knowledge arises from finite conditions, answering Skyrms’s question “Where does all the common knowledge come from?”

The answer: from four ingredients—a self-revealing state of affairs (C1), mutual awareness of that state (C2), indication of the target proposition (C3), and shared reasoning standards (C4, A6). These finite conditions generate infinitely many levels of nested belief.

8 Limitations and Extensions

While mathematically correct, the Cubitt–Sugden approach has philosophical limitations that motivate further work:

Unexplained axioms. A1 and A6 are assumed rather than derived. Why should modus ponens for indication hold? Why can agents reason about each other’s reasoning in this way?

Logical omniscience. If R captures derivability in an agent’s logic, agents have reason to believe all tautologies—an unrealistic idealization.

No conflicting reasons. The framework cannot represent agents with reasons for contradictory propositions (as in cases of conflicting evidence).

Vromen (2024) addresses these limitations using justification logic, where reasons are explicit terms rather than an implicit relation. In that framework, A1 and A6 become theorems derivable from more primitive principles about reason structure.

9 Conclusion

The Lean 4 formalization presented here establishes that Cubitt and Sugden's reconstruction of Lewis's theory is mathematically rigorous: the axioms A1 and A6, together with conditions C1–C4, suffice to prove Lewis's theorem for the complete infinite hierarchy of reasons to believe. The formalization makes explicit what Lewis left implicit and extends Cubitt and Sugden's partial proof to the general case.

This syntactic approach represents an important baseline for understanding Lewis's theory. While it assumes A1 and A6 as axioms, it provides a clear target for deeper foundations that derive these principles from more primitive assumptions about the structure of reasons.

References

- Cubitt, R. & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy*, 19, 175–210.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Vromen, H. (2024). Reasoning with reasons: Lewis on common knowledge. *Economics & Philosophy*, 40, 397–418.