

Designing Counterexample Sequences for Interpretive Revision in Philosophical Discussions of Promises

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Counterexamples can reveal a learner’s interpretation while also changing the distinctions available for subsequent inquiry. This paper examines the resulting sequencing problem in philosophical discussions of promises. It distinguishes a counterexample’s logical force, diagnostic reliability, and formative role, and specifies a learning target that permits reasoned disagreement about substantive moral conclusions. A finite episode model represents responses and interpretive changes jointly. A stepwise backward construction characterizes the existence of an observation-based policy that establishes a declared target within a bounded number of episodes while preserving specified conditions. Two opposed witnesses show how common preparation can make immediate diagnosis inefficient and how latent prerequisites can make diagnosis necessary. A sustained fictional dossier connects these results to promise formation, release, competing responsibilities, explanations, and transfer. Verified philosophical, educational, and planning antecedents locate the proposal without a claim of conceptual priority. The argument concludes with responsibilities for target selection and participation, serious objections, and a design for empirical evaluation. The formal guarantees remain conditional on the representation; educational effectiveness and wider applicability remain open.

Keywords: counterexample sequences; interpretive revision; philosophical dialogue; promise concepts; learning targets; diagnostic inquiry

Research Purpose and Scope

This note states the purpose of circulation, the status of the argument, and the role of research assistance. The paper presents current insights for criticism and wider discussion in public and scholarly settings. It makes no claim to conceptual priority. Similar concepts and arguments may already exist in other disciplines, languages, and intellectual traditions; further comparison may alter the terminology, scope, or contribution.

Definitions organize a bounded inquiry. Conditional formal results carry their stated assumptions, and illustrative cases support conceptual analysis. Empirical validation, wider applicability, and the adequacy of proposed normative premises remain open where the main text identifies those obligations. Relational analysis supplies a method of investigation while stronger ontological questions remain suspended.

AI assistance contributed to source discovery and pre-use verification, argumentative development, formal checking, organization, and drafting. The accompanying project records distinguish retained research insights, newly proposed analysis, and verified literature. The manuscript is prepared for the author's review; circulation and submission decisions remain with the author. Critical responses concerning antecedents, assumptions, counterexamples, and alternative formulations are welcome.

DRAFT

1 Introduction

This paper examines the design of counterexample sequences in philosophical discussions of promises. Its objective is to explain how an instructor can justify the order of examples when questioning both reveals and changes a learner's interpretation. The discussion first distinguishes educational targets and functions of counterexamples, then develops a conditional model of sequential teaching, and finally examines a sustained teaching dossier, pedagogical authority, and an empirical evaluation agenda. Conceptual argument, stipulated examples, and mathematical consequences perform different roles throughout the analysis.

A learner can classify several familiar cases correctly while interpreting the principle behind them in a way that becomes unstable in an unfamiliar situation. An instructor may then introduce a promise made under a misunderstanding, a promise released by its recipient, or a promise whose performance conflicts with an urgent responsibility. Such examples can disclose a previously unrecognized distinction. They can also change the learner's vocabulary, direct attention toward a feature selected by the instructor, or encourage a new account of the original case. The response therefore belongs to an evolving educational relation. Its evidential meaning depends partly on the teaching that preceded it and on the purpose for which the response is assessed.

This dependence creates a problem for a familiar design intuition: select the next example that most sharply separates competing interpretations. Sharp discrimination can be useful when the objective is to identify an interpretation that remains stable during questioning. Philosophical teaching adds a second objective. The learner is expected to become able to recognize, evaluate, and use distinctions, including distinctions introduced through the examples themselves. An episode that gives the instructor substantial information can leave the learner poorly prepared for the next challenge. Conversely, an explanation that produces little immediate diagnostic information can make a later counterexample intelligible. The educational value of an episode consequently depends on its position within a sequence.

The central proposal is a purpose-sensitive account of counterexample sequencing. A defensible sequence connects a warranted learning target to opportunities for the learner to examine a case, identify the premise under pressure, consider alternative repairs, and test a revised interpretation beyond the initial example. This proposal concerns the organization of inquiry. Its evaluation includes the learner's capacity to explain a distinction and challenge its application. Agreement with the instructor's preferred moral conclusion supplies only one possible response, whose significance requires further interpretation.

Promises provide a useful bounded setting because several questions can be separated without settling a complete theory of obligation. A discussion can examine whether an undertaking was formed, which considerations support carrying it out, how an authorized release changes the situation, and whether an overriding consideration leaves a reason for explanation or acknowledgment. The same final action can receive different descriptions depending on these distinctions. A learner who explains a difficult case through an articulated rival theory may demonstrate considerable understanding. An assessment that treats this response as simple failure would measure conformity to the teaching key alongside, or in place of, the intended conceptual competence.

The argument develops three contributions. The first is a distinction between a counterexample's logical function, its diagnostic function, and its formative function. A case can challenge a supporting premise while leaving a conclusion available, expose a learner's current classification, or prepare a new interpretation. These functions can coincide, yet each requires its own warrant. The second contribution is a finite model in which responses and interpretive changes are represented together. A stepwise derivation identifies conditions under which a sequence is guaranteed, within the model, to reach a declared target under uncertainty. The third contribution is a teaching dossier that connects those conditions to case selection, explanations, transfer, and the learner's standing to contest the target.

The mathematical construction is deliberately restricted. Its states summarize selected capacities and prerequisites; they supply an analytical representation of a task rather than an exhaustive description of a person. The model establishes consequences of a specified transition relation. Its educational interpretation requires evidence concerning whether the represented episodes and changes are plausible.

In particular, the existence of a formally successful policy establishes neither an empirically effective curriculum nor the legitimacy of its goal. Those questions receive separate treatment in the later sections.

Established work supplies important antecedents. Philosophical discussions of counterexamples already attend to the relation between conjectures and their supporting arguments. Educational research distinguishes the analysis of cases from the explanations that make those differences useful. Machine teaching and planning under partial information provide mathematical tools for representing ordered interventions and uncertainty. The present inquiry uses these resources to articulate a narrower design problem: the relation between interpretive revision, sequential preparation, and warranted evidence of philosophical learning. The comparison concerns clarification and application, with conceptual priority remaining open.

The practical motivation is an instructor's responsibility to make a sequence intelligible to its participants. A learner should be able to understand which distinction an example introduces, why a response invites revision, and how a later task tests the resulting understanding. This responsibility also extends to recognizing when an example fails, when direct explanation is appropriate, and when a learner has identified a weakness in the proposed lesson. Counterexample design becomes philosophically useful when it preserves these possibilities of correction within the teaching relation itself.

2 Antecedents in Philosophical and Educational Inquiry

This section locates the proposed account within verified work on counterexamples, learning activities, teaching sequences, and information-state planning. The objective is to identify the specific resources borrowed from each source and the additional judgments required by a philosophical application. The comparison proceeds by argumentative function, keeping historical interpretation distinct from the present design proposal.

Lakatos's fictional classroom distinguishes an objection to a supporting lemma from an objection to the main conjecture. The dialogue also presents itself as a rational reconstruction, with a different documentary role from an actual classroom transcript [2]. These features provide a useful antecedent for the present distinction between the target of a challenge and the form through which that challenge is presented. A teaching example can require revision of an argument without determining the fate of every claim associated with it. The present paper carries this distinction into an explicitly planned educational sequence; it leaves a comprehensive account of Lakatos's philosophy outside its scope.

The philosophical importance of this antecedent concerns the granularity of revision. A learner who encounters an objection needs to determine whether it challenges a term's extension, an inference, a background description, or a substantive principle. Treating every difficult example as grounds for abandoning an entire position can obscure the exact work performed by criticism. Preserving a position regardless of every challenge creates the opposite difficulty. The proposed teaching task asks the learner to locate the point of pressure and explain the consequences of a chosen response. That task permits several defensible revisions, provided their commitments remain explicit.

Schwartz and Bransford studied sequences in which learners analyzed contrasting cases and subsequently encountered an explanatory text or lecture. Their comparisons included repeated case analysis without the lecture, and their general discussion assigns complementary roles to differentiated knowledge and explanatory organization [4]. The study settings and assessments differ from the philosophical setting considered here. The relevant implication for design is therefore bounded: comparing examples and supplying explanations can perform distinct educational functions. The present proposal treats their relation as a question to investigate, with the effectiveness of a particular promise-discussion sequence remaining open.

A contrasting case becomes educationally useful through the distinction a learner can extract and the reasoning through which that distinction is evaluated. Merely presenting two descriptions side by side leaves these operations unspecified. An explanation can make a relevant difference salient, connect it to an argument, or clarify the purpose of the comparison. The instructor may also need to explain why an initially striking difference is irrelevant to the current claim. Case selection therefore involves both positive

preparation and the correction of misleading salience. This is a reason to include explanatory episodes in a sequence even when they provide limited information about the learner's initial interpretation.

Chi's conceptual framework distinguishes observable learning activities from the cognitive processes those activities may indicate. Constructive and interactive participation receive more specific descriptions than the mere presence of overt activity [1]. For the present inquiry, this distinction supports a cautious treatment of dialogue records. A learner's speaking, agreeing, or answering several questions creates observable material for assessment. The relation between that material and understanding still requires interpretation of the reasons expressed. A conversational format can support conceptual work, while its visible form alone supplies an incomplete basis for evaluating that work.

Zhu and colleagues distinguish sequential teaching from batch teaching and discuss uncertainty about a learner's model, including arrangements that interleave teaching and probing [5]. Their overview supplies an established comparison for optimizing a sequence relative to a specified learner and target. The present proposal adds a philosophical question about the warrant and revisability of that target. A target that efficiently induces assent may remain educationally questionable. The mathematical problem of producing a selected state and the normative problem of supporting a learner's inquiry therefore need separate formulations, even when the same episode contributes to both.

LaValle develops information-state planning in which a set of possible states can guide action and a task can be achieved while the precise underlying state remains unresolved. The treatment includes updates from possible transitions and observations, together with goal conditions expressed through containment of the possible-state set [3]. These are the principal mathematical antecedents of the finite construction below. The derivation restates the relevant steps for teaching episodes whose responses and effects belong to a joint relation. The transfer is analytical: the planning result acquires educational significance only through a separately defended interpretation of its variables and goals.

These comparisons clarify the scope of the proposed contribution. Counterexample analysis, explanations, interactive activity, adaptive teaching, and partial-information planning already possess substantial literatures. The present paper organizes their interaction around a specific philosophical problem. An instructor must select a next episode while remaining uncertain about both current interpretation and possible revision; the learner must retain opportunities to question the resulting direction. The resulting framework evaluates a sequence through its target, its possible developmental paths, and its evidence of transferable understanding. Its contribution can be assessed without assuming that these concerns originate in the present terminology.

3 Learning Targets for Discussions of Promises

This section specifies the educational object before introducing a selection rule. The objective is to distinguish competent interpretation from answer matching and to explain how a promise-discussion target can accommodate substantive disagreement. The analysis develops a working definition, identifies several dimensions of performance, and examines the limits of a finite representation.

Definition 3.1 (Interpretive learning target). An interpretive learning target is a declared repertoire of distinctions, reason-giving activities, and revision capacities that participants have reason to regard as relevant to a specified inquiry. Its assessment identifies the case family, permitted support, evidential standard, and range of defensible responses.

The definition treats a target as a proposal requiring justification. A course can reasonably expect students to distinguish a conclusion from its supporting reasons, recognize an objection to a universal claim, and explain the commitments of a revised position. These expectations concern participation in an inquiry. A requirement to affirm the instructor's preferred moral theory would need a different justification. Keeping these objectives separate allows a learner to understand a position, apply its distinctions accurately, and criticize it. A record of disagreement can then supply evidence of learning when the reasons demonstrate the intended competence.

For the promise dossier, the initial target has four connected components. The learner should distinguish the formation of an undertaking from the later question of performance; distinguish a recipient's release from an overriding consideration that leaves the undertaking's formation intact; explain how a counterexample bears on a stated principle; and transfer the distinction to a new case whose surface details differ. These components are selected for their role in analyzing a family of arguments. They establish a local curriculum objective. Their usefulness beyond that family requires further educational and philosophical comparison.

The first distinction concerns formation. An apparent assurance may be quoted from a script, uttered as a joke recognized by both parties, or understood differently by the participants. Such variations can affect whether the description of the case supports the claim that an undertaking was made. An instructor who simply labels every utterance of a particular sentence a promise would build a controversial semantic premise into the assessment. The appropriate exercise can stipulate a shared understanding for one comparison and reopen that stipulation in a later case. The learner's task is to identify which assumption carries the classification.

The second distinction concerns changes after formation. A recipient can communicate a release within the exercise's agreed conditions. A separate emergency can provide a reason to depart from a commitment while leaving unresolved what acknowledgment remains appropriate. The educational point lies in recognizing that these are different routes to a final action. A learner can accept both distinctions while defending different judgments about the strength of the resulting reasons. The assessment should therefore record the learner's interpretation of the route, together with the reasons given for the final judgment.

The third component concerns the scope of refutation. Suppose a proposed principle states that every valid undertaking requires performance under every subsequent condition. An accepted example of release can challenge that principle under its stated interpretation. A different principle might already include release among its conditions. Repeating the same example against that revised principle would require additional argument. The learner needs to connect the case with the exact proposition at issue. Familiarity with an instructor's favored example can conceal a failure to make that connection, particularly when the wording of the target proposition changes.

The fourth component concerns transfer. A learner may repeat a successful explanation of a meeting arrangement while failing to recognize an analogous distinction in an equipment-sharing arrangement. Alternatively, the learner may correctly reject the analogy because a relevant feature differs. Transfer assessment must therefore allow two forms of success: applying a warranted distinction and explaining why an apparent transfer would be mistaken. A scoring key that rewards only the first form encourages indiscriminate extension. The quality of the reasoning should determine which performance the case supports.

Definition 3.2 (Target-admissible disagreement). A disagreement is target-admissible when the learner accurately identifies the relevant claims and case assumptions, gives reasons responsive to them, and preserves the distinction being assessed while disputing a substantive conclusion or its normative basis.

This definition leaves room for correction. A response that contradicts its own premises, overlooks an explicit condition, or misdescribes an interlocutor's argument remains open to criticism. The category of admissible disagreement protects reasoned alternatives; it supplies neither automatic acceptance of every response nor immunity from argument. Its practical use requires assessors to explain which requirement a response satisfies or fails. Borderline cases should remain visible in the evaluation record, especially when the disagreement concerns the adequacy of the learning target itself.

A finite model can represent these targets through a set of states that satisfy a declared criterion. Several states may qualify. One can represent a learner who favors a strongly commitment-centered position and another who assigns greater weight to conflicting responsibilities, provided both demonstrate the distinctions required by the exercise. This representation avoids forcing philosophical competence into a single endpoint. It still compresses a richer activity. The state labels depend on the available evidence, case descriptions, and interpretive judgments; a diagram cannot independently validate their meaning.

The target also depends on permitted support. A learner who uses a written case summary, translation, visual diagram, or assistive technology may demonstrate the intended distinction with substantial support. If the educational objective concerns philosophical interpretation, removing that support can introduce an unrelated burden. A later assessment of independent recall would concern an additional objective whose justification should be stated. The model below therefore permits support to be included in an episode and in the description of the target. Competence is evaluated within declared conditions of participation.

Finally, target revision belongs to the inquiry. A learner may identify a missing category, such as a difference between releasing an undertaking and denying that it ever existed. The instructor can then expand or revise the case family. The earlier sequence retains its original purpose and record, while the revised inquiry receives a new target. This procedure preserves conceptual continuity without treating a finite teaching model as the boundary of legitimate philosophical thought.

4 Warrant and Function of Counterexamples

This section explains the reasoning that licenses a counterexample and the additional conditions required for its educational use. The objective is to keep logical force, evidential reliability, and formative value distinct. The discussion proceeds from a minimal logical pattern to disagreements about case description, auxiliary premises, and appropriate revision.

Definition 4.1 (Counterexample relative to a claim). A counterexample to a claim is a case whose warranted description is incompatible with that claim under the interpretation and background assumptions being tested. Its force is relative to the claim's quantification, domain, and stated conditions.

Consider a proposition of the form “every case satisfying conditions A also satisfies condition B.” A refutation requires a warranted instance that satisfies A and fails B. Finding an instance that merely resembles an A-case leaves the first requirement unresolved. Finding an unusual B-case leaves the second requirement unresolved. In a philosophical discussion, these obligations often carry more argumentative weight than the final contradiction. Participants can disagree about whether a commitment was formed, whether a release was authorized, or whether a background description excludes a competing interpretation.

The instructor should make this inferential sequence explicit. First state the claim in a form that reveals its domain and conditions. Next identify the relevant features of the proposed case. Then explain the interpretation under which those features satisfy the antecedent and conflict with the consequent. Finally consider whether a disputed background assumption supports a different diagnosis. This order allows the learner to locate disagreement. It also prevents the vividness of a story from functioning as a substitute for the argument that connects the story to the target claim.

An example can challenge a supporting premise while leaving a conclusion available through another route. Suppose a learner argues that a commitment requires performance because every recipient continues to want the promised action. A release case challenges the supporting generalization about recipients' wants. It leaves open whether some other reason supports a residual obligation. The appropriate response may therefore revise the argument while preserving part of the conclusion. The educational aim is to identify this difference, including the additional reasons needed to defend what remains.

An apparent counterexample may also reveal a shifting concept. A learner can respond to an emergency case by narrowing the meaning of a valid promise until the case falls outside the domain. This move can have a principled justification, such as an explicitly defended condition of feasibility. It can also serve only to protect a preferred conclusion. Assessment requires comparing the revised definition with the original motivating cases and with the reasons offered for the change. A refinement becomes philosophically useful when it clarifies the explanatory role of the concept and survives renewed scrutiny. Its formal ability to exclude a troublesome instance is only one part of that assessment.

Claim 4.2 (Distinct warrants for educational use). The logical force of a counterexample, its reliability as evidence of a learner's current interpretation, and its value in supporting a later interpretation require separate justification. Success in one role leaves the other roles open.

The claim follows from differences in the relevant objects. Logical force concerns a relation among a proposition, assumptions, and a case. Diagnostic reliability concerns the process through which a response is produced and interpreted. Formative value concerns what the learner becomes able to understand or do through the episode and its continuation. A logically decisive example can produce an ambiguous response because the learner has misunderstood its language. A response can reveal that misunderstanding while the episode leaves it unresolved. A carefully supplied explanation can then support learning even though it reveals little about the learner’s initial interpretation. These possibilities establish the need for separate assessments.

The instructional record should therefore include the function assigned to each case. A formation contrast can expose an unnoticed presupposition. A release contrast can distinguish two routes to an action. A conflict case can test how the learner locates a defeasible reason. A transfer case can examine whether the distinction travels across descriptions. A learner-generated counterexample can test both understanding and the adequacy of the instructor’s case family. The same example may perform several of these functions, but its place in the sequence should make the intended function intelligible.

The reliability of a response also depends on the interaction. A learner may infer that the instructor expects a particular answer and provide it without endorsing or understanding the reasoning. Another learner may resist a case because it omits a relevant feature, even while understanding the intended principle. The instructor needs a route for requesting reasons, inviting clarification, and recognizing that a disputed case can expose a weakness in the lesson. Such activities add time, but they improve the interpretability of the teaching record. Their value cannot be measured solely by the number of candidate interpretations eliminated.

Revision after a counterexample has several legitimate forms. Participants may narrow a claim, weaken its conclusion, revise an auxiliary assumption, distinguish previously merged cases, or abandon an unhelpful division. Some revisions merge cases that an earlier theory separated. Others simultaneously merge one pair and separate another. A pedagogy that rewards only an increasing number of distinctions can encourage unnecessary complexity. The criterion should concern what a revision explains, preserves, and makes available for criticism. The later model treats the target as a declared repertoire of capacities, allowing the content of those capacities to include justified simplification.

The resulting account gives counterexamples a place within a wider practice of explanation. Their force comes from an articulated relation to a claim, and their educational role comes from what participants can make of that relation. A sequence becomes reviewable when each episode records those connections and retains a route for correcting them. This prepares the distinction between stationary diagnostic selection and the development-sensitive construction that follows.

5 Selection under a Stationary Interpretation Model

This section reconstructs the limited setting in which selecting a highly discriminating question has a clear formal meaning. Its objective is to identify the assumptions that support hypothesis elimination and the boundary between that result and educational improvement. The construction uses a finite candidate family, then considers coding dependence and the effects of revision during inquiry.

Let H be a finite, nonempty set of candidate interpretations, U a set of permitted questions, and Y a response domain. Each candidate h supplies a deterministic response function $f_h : U \rightarrow Y$. A target $g : H \rightarrow V$ identifies the distinction the instructor wishes to infer. Several candidates can have the same target value, for example when they differ in details irrelevant to the intended comparison. For this section alone, questions leave the relevant response functions unchanged and the observed response is assumed accurate under the model.

After recording question–response pairs (u_s, y_s) through episode t , the remaining candidate set is defined by Equation 5.1.

$$H_t = \{h \in H : f_h(u_s) = y_s \text{ for every } s \leq t\}. \quad (5.1)$$

If the actual candidate belongs to H and generates every response through its declared function, it belongs

to H_t . Adding a response adds a constraint, so the next candidate set is contained in the preceding one. The target is exactly identified at a record when g has a single value on that set. This conclusion concerns identification relative to a correct candidate family. It leaves the learner’s development outside the construction because the learner’s relevant interpretation was assumed stationary.

Proposition 5.1 (Target separation under stationary responses). *A finite question family identifies the target for every candidate exactly when each pair of candidates with different target values receives different responses to at least one question in that family.*

Proof. Suppose a target-opposed pair receives the same response to every question in the family. Either candidate can then generate the complete response record, so that record admits two target values. Identification fails. Conversely, suppose every target-opposed pair is separated. Any two candidates compatible with the complete response record agree on every question in the family. They therefore cannot be a target-opposed pair. All compatible candidates have the same target value, which gives identification under the stated model. $\square$

One possible selection score is the size of the largest response class. For a current candidate set H_t , define the score in Equation 5.2; only responses produced by at least one current candidate enter the maximum.

$$W_t(u) = \max_{y \in f_{H_t}(u)} |\{h \in H_t : f_h(u) = y\}|, \quad f_{H_t}(u) = \{f_h(u) : h \in H_t\}. \quad (5.2)$$

Minimizing Equation 5.2 minimizes the largest number of candidates that could remain after one response. The meaning of this score is exact and narrow. It counts candidate descriptions, treats the next response as stationary evidence, and considers one step. It supplies neither a probability of learning nor a measure of the philosophical importance of a distinction.

Candidate representation matters. If one interpretation receives several equivalent encodings, those encodings can alter a count-based ranking while leaving the substantive uncertainty unchanged. A target-sensitive design should therefore examine the distinctions represented by the candidate family, their relation to the learning objective, and the cost of each question. A justified probability distribution would support a different analysis, with additional assumptions. Calling a count an uncertainty measure without specifying these choices can hide the role of the instructor’s representation in the resulting sequence.

The stationary construction nevertheless has practical value. An instructor may wish to determine whether a learner understands a specific term before choosing an explanation. A short diagnostic comparison can narrow the plausible interpretations and prevent an unsuitable intervention. The result remains conditional on how the diagnostic episode operates. If the question itself introduces the relevant distinction, the recorded response may already reflect a changed interpretation. The instructor then needs an account of the episode’s effects, alongside an account of its informativeness.

This boundary is consequential for philosophical learning. A response incompatible with every listed candidate may reveal a new distinction, a misunderstood prompt, or a defective measurement assumption. Treating the empty candidate set as proof of learner irrationality would silently protect the model from criticism. The appropriate response is to identify which assumption failed and revise the record of possibilities. Likewise, successful target identification can establish what the instructor has inferred while leaving the learner’s capacity unchanged. A sequence designed for learning must represent that second object directly.

6 Learning Episodes and Evidence Updates

This section introduces a model in which an educational episode can both produce evidence and change the learner’s represented capacities. The objective is to specify the timing and dependencies needed for sequence analysis. Symbols, modeling assumptions, an update rule, and its limitations are developed in that order.

Let Z be a finite, nonempty set of modeled current states, U a finite set of possible episodes, and Y a finite response alphabet. A state summarizes the distinctions and prerequisites relevant to the local teaching task. It may also include aspects of the teaching relation, retained support, or prior preparation when those features affect subsequent responses. An episode can be a question, explanation, comparison, learner-generated objection, or supported transfer exercise. The finite description is a modeling choice; it leaves many features of actual philosophical activity unrepresented.

For every feasible pair (z, u) , let $R_u(z)$ be a nonempty subset of $Y \times Z$. A pair (y, z') in this set means that episode u , beginning at state z , can produce recorded response y and terminate at state z' . The relation permits several responses, several successors for one response, and changes occurring within the episode. Recording response and successor jointly avoids imposing an arbitrary independence assumption between them. It also makes clear that an answer expressed near an episode's end may already reflect its formative influence.

Assumption 6.1 (Adequacy of the episode representation). The modeled initial possibilities include the actual relevant state. Every actual response–successor pair produced by a permitted episode belongs to the stated relation. The represented state retains the history needed to determine the permitted response–successor possibilities for later episodes within the declared horizon.

This assumption is demanding. Its final clause requires the model to retain relevant dependencies rather than naming a compressed snapshot a state and assuming that it suffices. If fatigue, trust, prior explanation, or a changing interpretation affects the next episode, the model must represent that influence or widen its uncertainty relation appropriately. An observation outside the relation defeats the adequacy assumption for the current use. Mathematical success cannot independently establish this empirical premise.

A nonempty set $B \subseteq Z$ records the states compatible with the available evidence. For a feasible episode u and observed response y , Equation 6.1 defines the updated possible-state set.

$$\Phi(B, u, y) = \{z' \in Z : \text{some } z \in B \text{ satisfies } (y, z') \in R_u(z)\}. \quad (6.1)$$

The possible response set is $Y_u(B) = \{y \in Y : \Phi(B, u, y) \neq \emptyset\}$. This set is nonempty whenever the episode is feasible at every member of B , by the nonemptiness assumption on $R_u(z)$. An empty update after an actual response signals a conflict between the evidence and the model. It requires investigation of the candidate domain, episode relation, and response coding.

Proposition 6.2 (Conditional preservation of the actual successor). *Under the episode-adequacy assumption, if the actual state belongs to B , then the actual successor belongs to the updated set in Equation 6.1 for the observed response.*

Proof. Write the actual initial state as z and the actual response–successor pair as (y, z') . Adequacy gives $z \in B$ and $(y, z') \in R_u(z)$. These two facts satisfy the membership condition defining $\Phi(B, u, y)$. Repeating this argument after each episode establishes the same preservation property throughout any finite sequence for which adequacy continues to hold. $\square$

The update combines two operations conceptually. Evidence excludes initial possibilities that could not have produced the recorded response. The episode then carries the surviving possibilities into their permitted successors. Consequently, the new possible-state set can contain states absent from B . This differs from stationary elimination, where evidence only removes candidates from a fixed family. A decline in the number of possible current states can reflect convergence produced by teaching, stronger evidence, or both. It therefore requires interpretation before it is treated as a measure of what the instructor has learned about the earlier learner.

Let $S \subseteq Z$ be a set of states satisfying a declared preservation condition, and let $G \subseteq S$ be the local learning target. For example, the preservation condition may require retaining a previously established distinction while a new case is introduced. This formal condition captures only a specified pedagogical concern; the legitimacy of an episode also depends on consent, accessibility, and appropriate conduct.

Let $A(B)$ contain the episodes that are feasible at every state in B and satisfy the separately declared participation conditions. The model is allowed to return an empty permitted-episode set. That outcome identifies a limit of the represented design.

A successful terminal record must justify the claim that the current state lies in G . Under the exact possible-state interpretation, this requires $B \subseteq G$. The condition permits uncertainty among several target-admissible states. The instructor can thus have adequate evidence of the declared competence while remaining uncertain about the learner's preferred substantive theory. It is also possible for the actual state to belong to G while the available record remains compatible with states outside it. That distinction separates achievement from warranted certification of achievement.

The mathematical vocabulary does not decide the moral quality of the selected target or preservation condition. A restrictive target can be described precisely and pursued efficiently. The later discussion therefore subjects both G and $A(B)$ to philosophical criticism. Their explicit representation makes that criticism easier to locate: a disagreement may concern the modeled possibilities, the intended competence, the admissible intervention, or the inference from response to state. Keeping those disagreements separate is part of the proposed method.

7 Conditional Guarantees for Bounded Sequences

This section derives a finite criterion for the existence of a successful teaching sequence under the preceding model. The objective is to make the policy, uncertainty, safety, and stopping assumptions explicit. The argument constructs backward success sets, proves their interpretation by induction, and then examines computational and epistemic limits.

An observation-based policy selects a permitted episode from the history of recorded responses and earlier episodes. It cannot select an action by consulting the actual state when that state remains unobserved. A policy may stop when its current possible-state set is contained in G . A safe policy keeps the possible-state sets within the nonempty subsets of S until it stops. These conditions concern what is warranted uniformly across the modeled possibilities, rather than the performance of a policy along one favorable trajectory.

Let $\mathcal{B}_S$ denote the nonempty subsets of S . Define $\mathcal{W}_k$ as the family produced by Equation 7.1.

$$\begin{aligned}\mathcal{W}_0 &= \{B \in \mathcal{B}_S : B \subseteq G\}, \\ \mathcal{W}_{k+1} &= \mathcal{W}_k \cup \{B \in \mathcal{B}_S : \exists u \in A(B) \forall y \in Y_u(B), \Phi(B, u, y) \in \mathcal{W}_k\}.\end{aligned}\tag{7.1}$$

Equation 7.1 first identifies the records at which the target is already warranted. It then admits a possible-state set if one permitted episode sends every possible response branch into a previously admitted set. The existential choice belongs to the policy; the universal quantifier ranges over modeled response possibilities. Interchanging these quantifiers would allow a different first episode to be chosen after learning which response it would produce, which would misrepresent an executable sequence.

Proposition 7.1 (Finite-horizon sequence criterion). *For every nonnegative integer k and nonempty $B \subseteq S$, membership $B \in \mathcal{W}_k$ is equivalent to the existence of an observation-based permitted policy that safely establishes the target within at most k episodes, for every trajectory allowed by the model from B .*

Proof. At horizon zero, the policy can only stop. Stopping is warranted exactly when $B \subseteq G$, which is the definition of $\mathcal{W}_0$.

Assume the equivalence holds for horizon k . For soundness at horizon $k + 1$, first consider $B \in \mathcal{W}_k$. The policy supplied by the inductive hypothesis already succeeds within the shorter horizon. Otherwise, membership in $\mathcal{W}_{k+1}$ supplies an episode $u \in A(B)$ for which every possible successor belief belongs to $\mathcal{W}_k$. Execute u . The recorded response identifies a branch y , and the inductive hypothesis supplies a successful continuation policy for $\Phi(B, u, y)$. The response alphabet is finite, so these branch policies can be assembled into one observation-based policy. Each successor lies within S , and the initial B also lies within S . The assembled policy therefore preserves safety and stops within $k + 1$ episodes.

For completeness, suppose a permitted policy safely establishes the target within at most $k + 1$ episodes. If it stops immediately, B belongs to $\mathcal{W}_0$. Otherwise its first episode is one fixed $u \in A(B)$, chosen from the available record. For each possible first response y , the policy's continuation safely succeeds from $\Phi(B, u, y)$ within at most k further episodes. The inductive hypothesis places every such successor belief in $\mathcal{W}_k$. The second clause of Equation 7.1 therefore places B in $\mathcal{W}_{k+1}$. This establishes both directions. $\square$

Several assumptions prevent misleading interpretations of the result. The response–successor relation is nonempty for feasible episodes, so a universal condition over possible responses cannot certify success merely because the model forgot to provide any response. The policy branches on observable records rather than hidden states. The target is a set of admissible terminal states, with stopping available once the evidence warrants containment. Safety applies throughout the represented sequence. Each of these conditions contributes to the proof; removing one changes the meaning of the conclusion.

For finite Z , the number of nonempty possible-state sets is at most $2^{|Z|} - 1$. The families $\mathcal{W}_k$ increase monotonically. Each strict increase adds at least one possible-state set; therefore the construction reaches a fixed point after finitely many strict increases. An equality of consecutive families is a fixed point because the same update operation is applied at every stage. This establishes termination of exhaustive backward analysis in the finite model. The size of the possible-state space can nevertheless make direct enumeration impractical. The proposition supplies a correctness criterion, while efficient computation for richer cases remains a separate question.

The smallest k for which B belongs to $\mathcal{W}_k$, when it exists, is a worst-case episode bound under the model. It treats each episode as one unit, which can be unsuitable when explanations, discussions, and supported exercises require different amounts of time or effort. A cost-sensitive variant must define costs and how they accumulate. The present analysis retains the unit-episode convention to isolate the logical difference between immediate information and eventual learning. The teaching dossier identifies where this convention would need refinement in practice.

Claim 7.2 (Sensitivity to enlarged uncertainty). If a revised model contains every response–successor possibility of an earlier model, retains the same target and preservation condition, and permits the transfer of a policy's admissibility to the earlier model, every policy guaranteed successful in the revised model is also successful in the earlier model. The converse can fail.

To justify the claim, hold the policy fixed as a function of observable history. Every earlier-model trajectory generated by that policy is also a revised-model trajectory: the initial possibilities are included, and each step uses a response–successor pair retained by the revised relation. A guarantee over the larger trajectory family therefore covers the smaller one. The admissibility-transfer condition ensures that the policy remains executable under the comparison. The converse fails when the revised relation adds a branch leading outside the target or preservation set. This result identifies a source of fragility: a guarantee can depend on excluding a response or learning path that actual participants can produce.

The criterion should consequently be used to expose assumptions and compare designs. Its pedagogical value lies partly in showing which uncertainty branches a proposed sequence has neglected. A policy that succeeds under a narrow model may remain worth studying, but its empirical claim should reflect that dependence. The formal result organizes a conditional argument; educational warrant still requires a defensible target, an adequate model, and evidence that the sequence supports the intended form of participation.

8 A Promise-Interpretation Teaching Dossier

This section develops a fictional teaching dossier that connects the formal distinctions to concrete philosophical episodes. The objective is to compare the roles of preparation, diagnosis, counterexample analysis, and explanation. The case sequence is followed by two opposed finite witnesses, showing how different prerequisite structures justify different orders.

Table 1: Case roles in the fictional promise-discussion sequence

Episode	Case variation	Interpretive objective	Evidence sought
Reference undertaking	Shared understanding and feasible performance	Articulation of initial reasons	Explanation of the judgment
Formation contrast	Rehearsal understood by participants	Distinction between wording and undertaking	Identification of the changed premise
Authorized release	Later release within stipulated authority	Separation of formation and subsequent status	Comparison with the formation contrast
Competing responsibility	Urgent conflict after formation	Analysis of overriding and residual reasons	A defended account of the route to action
Learner objection	Participant introduces a countercase	Revision of the case family or claim	Identification of a premise under pressure
Transfer exercise	New setting and altered surface details	Application or justified refusal of transfer	Reasons responsive to structural differences

The setting is a small undergraduate discussion in which participants examine an unrestricted claim about keeping promises. The instructor introduces an ordinary arrangement: one person undertakes to meet another at an agreed place and time, both understand the undertaking, and performance remains feasible. This case establishes the reference description. Participants state the reasons they think matter and identify which features their initial judgment relies upon. The exercise records those reasons without treating the first response as a complete account of the participant’s moral outlook.

The next episode changes one feature. The apparent undertaking is a line in a rehearsal, and both participants understand that context. The purpose is to distinguish words occurring in a description from the role those words perform in the relation. A learner who already makes this distinction can explain it quickly. A learner who treats the sentence alone as sufficient may need a clearer contrast, direct explanation, or an opportunity to supply another example. The record should distinguish a preexisting explanation from an explanation formed during this exchange.

A release case then preserves the initial undertaking and adds a later communication by the recipient, under stipulated conditions in which the recipient has standing to release it. This case separates a change after formation from a claim that formation never occurred. A further emergency case preserves the undertaking and introduces a competing responsibility. The instructor invites several accounts of the resulting reasons. Some learners may defend a residual need for acknowledgment; others may dispute its basis. The target concerns their ability to distinguish these routes and state the assumptions supporting a judgment.

Table 1 records the structural role of each episode. The table is a design proposal, with the cases supplying stipulated distinctions rather than observed learning outcomes.

The instructor supplies a short explanation after the comparison when it helps organize the distinctions. That explanation can state the difference between questioning whether an undertaking arose and questioning what follows after it arose. The learner then examines a new case in a different setting. This sequence makes room for telling, comparison, and independent reasoning. Its adequacy remains open to the learner’s response. If the learner identifies an ambiguity in the release stipulation, the instructor should clarify or revise the case before treating it as a successful test.

Example 8.1 (Preparation before a diagnostic episode). Let $Z = \{a, b, p, g\}$, let every state satisfy the preservation condition, and let the goal be $G = \{g\}$. States a and b represent two initial interpretations, p a shared prepared state, and g the declared target. A diagnostic episode d reveals the state and leaves it unchanged. A preparation episode e takes both a and b to p with the same response. An application episode c takes p to g and leaves a and b unchanged. Unspecified actions preserve g , and the preparation preserves p . Every episode costs one unit.

From $B_0 = \{a, b\}$, the sequence e, c reaches the target in two episodes. The preparation supplies no discrimination between the initial candidates because both produce the same response. The diagnostic

episode separates them perfectly, but its successor remains either $\{a\}$ or $\{b\}$. Each of those states still requires preparation and application, so the diagnostic-first route needs three episodes. In terms of the stationary response-class score, the diagnostic has a largest initial-candidate class of one while preparation has a class of two. That comparison favors immediate diagnosis even though the bounded teaching task favors preparation.

The example distinguishes information about the initial interpretation from uncertainty about the current state after an intervention. Preparation merges the modeled initial possibilities into p , so a score based only on the number of current successor states would evaluate a different object. The case therefore refutes a specific universal claim about maximizing immediate diagnostic discrimination. It establishes neither the superiority of explanations nor a general rule to avoid diagnosis. Its educational interpretation would require evidence that a common preparation genuinely serves both represented initial interpretations.

Example 8.2 (Diagnosis before tailored preparation). Let the initial possibilities again be $\{a, b\}$. A diagnosis reveals the initial state without changing it. Preparation e_a takes a to p_a and takes b to a state d outside the preservation set. Preparation e_b takes b to p_b and takes a to d . An application episode takes either prepared state to g and leaves unprepared states unchanged. The target is $\{g\}$, and all states except d satisfy the preservation condition. Remaining transitions preserve the relevant state.

In this second model, applying either tailored preparation before diagnosis admits a branch outside the preservation set. Diagnosis followed by the matching preparation and application succeeds in three episodes. A two-episode sequence cannot safely guarantee success: an initial preparation is unsafe, an initial application leaves both unprepared possibilities, and an initial diagnosis leaves two further steps necessary on either branch. The order follows from a different prerequisite structure. The model supplies a countercondition to the first example's pedagogical interpretation.

The state d represents a stipulated failure to preserve a specified distinction or condition. It is neither a clinical diagnosis nor a general description of distress. Its role is to show how an intervention can be unsuitable for one represented starting point. In practice, a teacher would need evidence concerning that risk, a proportionate interpretation of its significance, and a route for offering alternative support. The finite example gives those assumptions a visible place in the design argument.

Together, the witnesses locate the central design problem. The relevant next episode depends on what the instructor is trying to establish, what the learner may currently need, and how the episode can change the possibilities available afterward. The same observable response can occur along different histories, and different responses can lead into a shared prepared state. A justified sequence therefore requires an account of developmental transitions alongside an account of response discrimination. The dossier offers a structure for obtaining and criticizing that account, while actual effectiveness remains an empirical question.

9 Revision, Explanation, and Transfer

This section examines the evidence needed to interpret a successful response as part of learning. Its objective is to distinguish revision of the learner's position, revision of the instructor's model, and transfer of a distinction to a different case. The discussion develops a record of reasons and then considers retention, support, and changes in the case domain.

A useful revision record states the earlier formulation, the challenge, the premise affected, the current formulation, and the reasons for retaining or changing each part. This record can be brief. Its purpose is to make the inferential movement available for scrutiny. In the promise dossier, a learner might initially claim that the same words create the same obligation in every context. The rehearsal case can motivate a distinction between an utterance's wording and its understood role. The learner can then state which contextual conditions matter and examine whether the revised account handles the reference undertaking. The assessment concerns this movement in reasons, rather than a requirement to repeat a prescribed sentence.

The record should also preserve a justified refusal to revise. An instructor may propose a supposed release case in which the person issuing the release lacks the authority stipulated in the earlier lesson. A learner who identifies that difference can reasonably reject the case's intended force. Such a response can demonstrate the target distinction more clearly than immediate agreement. The instructor's own interpretation then requires revision. A sequence that permits this outcome treats philosophical learning as a practice of accountable reasoning shared across unequal roles.

Definition 9.1 (Revision record). A revision record is an account connecting a prior formulation, a specified challenge, an identified inferential or descriptive pressure, and a current formulation with its supporting reasons. It distinguishes an observed change in response from an inferred change in understanding and records uncertainty about that inference.

The definition separates a documentary object from the event it represents. A polished retrospective account can explain how a position could be revised while leaving the actual history of revision unresolved. An instructor should identify such a reconstruction when it is used. Equally, a faithful response transcript may provide limited evidence of the conceptual movement if it records only short answers. The evidential task can require follow-up explanation, comparison, or a learner-generated case. The record's usefulness depends on how it supports the inference under examination.

Transfer adds a further requirement. A learner should encounter a case whose relevant structure is related to the earlier examples but whose incidental cues have changed. For instance, an undertaking to attend a meeting can be followed by an undertaking concerning shared equipment. The assessment can ask the learner to identify which features justify transferring a distinction about formation or release. The learner should also encounter a near case in which an apparently similar communication has a different role. This pair tests both successful application and restraint in application.

The transfer task should allow the learner to explain a relevant difference that the assessor has missed. Otherwise the exercise risks rewarding a predetermined analogy. A defensible evaluation can separate three judgments: whether the learner identified the intended structural feature, whether the feature is actually present under the case description, and whether the learner's conclusion follows from the stated account. Assessors may disagree on any of these judgments. Recording the source of disagreement improves the interpretation of the result and can reveal that the case bank itself requires revision.

Retention concerns another time scale. A distinction used immediately after an explanation may be difficult to reconstruct later. A delayed exercise can provide evidence about whether the learner can recover and apply it under declared support conditions. The delay should be justified in relation to the educational aim and the opportunities for further learning between assessments. A changed response after later discussion may reflect development, forgetting, reinterpretation, or a new objection. The evaluation should examine the reasons supplied rather than treating every deviation from the earlier answer as deterioration.

Support can legitimately remain part of the target. A learner may use a distinction map or a written argument to organize a response. If the learning aim is competent analysis using available resources, that performance directly bears on the aim. An additional test without those resources concerns the capacity to reproduce or regenerate the analysis under different conditions. The two performances can inform each other while retaining their different evidential meanings. This distinction is particularly important when access needs make a supposedly neutral unsupported task systematically harder for some participants.

The formal model can incorporate transfer by representing a final episode whose response and successor depend on the relevant distinction in a changed setting. Such a representation is meaningful only if the transfer episode actually discriminates the intended competence. Stipulating that every state-labeled competent produces a successful transfer response merely defines the model's behavior. It leaves the empirical relationship to be tested. The model is most useful when it makes alternative possibilities explicit, such as cue-dependent performance that succeeds on a familiar case and fails on a structurally related one.

A learner's new question can also expand the target. The distinction between release and retrospective denial, for example, may introduce historical questions absent from the original case family. The

instructor can preserve the original target and add a new phase of inquiry with an explicit connection to it. This approach avoids interpreting every expansion as failure to finish the lesson. It also prevents the earlier finite success guarantee from being silently transferred to a changed objective. The guarantee concerns the original model; the expanded inquiry requires renewed analysis.

10 Pedagogical Authority and Conditions of Participation

This section addresses the normative premises that a sequencing model leaves open. Its objective is to explain the teacher's responsibility for target selection, intervention, and assessment while preserving the learner's standing to question the inquiry. The argument connects the purposes of philosophical education to transparency, accessibility, and revisable authority.

The proposed normative starting point is that a philosophical educational relation should support the learner's participation in giving, examining, and revising reasons. This premise is defended through the activity's declared purpose. If a course presents itself as teaching philosophical inquiry, then success should include a capacity to assess a claim's grounds and limits. Producing a selected answer can contribute to that purpose, but an arrangement that systematically excludes examination of reasons would inadequately realize it. The responsibility follows from the educational commitment being undertaken, together with the learner's standing within that commitment.

An instructor may possess greater knowledge of a literature, stronger command of an argument, or responsibility for organizing limited class time. These asymmetries can justify choosing an initial sequence and directing attention to neglected distinctions. Their justification remains connected to the educational purpose. A teacher should be able to explain the role of a case, acknowledge a relevant objection, and distinguish a curricular requirement from a disputed philosophical conclusion. This form of authority can guide inquiry while remaining answerable to the reasons through which the guidance is defended.

Claim 10.1 (Accountability of the learning target). When a sequence is justified by its contribution to philosophical inquiry, its target and assessment criteria should remain open to reasoned challenge. The teacher's control of the sequence supplies a responsibility to explain and reconsider those criteria when a material objection arises.

The claim rests on two premises. First, the declared aim concerns the learner's ability to participate in critical reasoning. Second, target selection and assessment materially shape which responses can count as successful participation. If the target were insulated from relevant criticism, the teacher would reserve exemption for the very judgment organizing the inquiry. A proportionate challenge route therefore belongs to realizing the stated aim. This conclusion establishes a pedagogical responsibility within the specified relation; it leaves wider political and legal questions to separate arguments.

Target openness need not mean that every lesson continually renegotiates every background assumption. An instructor can stipulate a local model to examine its consequences, identify the stipulation clearly, and reserve broader objections for a later discussion. The learner should understand that proving a conditional result within that model leaves its application open. The same principle applies to the finite sequencing construction. Students can examine which policy succeeds under a stipulated transition table while also questioning whether that table represents learning adequately. These are connected levels of inquiry with different immediate objectives.

The choice of counterexamples also distributes attention. A case bank can repeatedly foreground one kind of relationship and leave other experiences difficult to articulate. The issue cannot be resolved simply by adding more examples. The instructor should examine which distinctions the selected cases make available, which perspectives may be missing, and whether learners can introduce alternatives. Such an examination treats the case domain as a revisable research object. It avoids assuming that a finite menu captures every legitimate way of understanding a promise or responsibility.

Accessibility is part of this responsibility because a response can be shaped by burdens unrelated to the target. Complex language, inaccessible presentation, unfamiliar institutional details, or time pressure

can obscure the distinction under study. A teacher who interprets those effects as philosophical failure may select progressively harder counterexamples that intensify the wrong difficulty. An appropriate adjustment can simplify the wording, provide translation, offer a diagram, or permit another response mode while preserving the conceptual target. The role of the adjustment should be recorded so later assessments use comparable conditions.

Direct explanation can likewise support participation. A learner may reasonably request an explicit account after several unsuccessful comparisons. The teacher can provide that account and then invite analysis, criticism, or transfer. Withholding explanation until the learner supplies the expected phrase can turn questioning into a guessing exercise. The relevant criterion concerns whether the intervention helps the learner occupy a meaningful role in the inquiry. Surface features such as silence, question count, or unaided effort provide incomplete indicators of that role.

The learner's choice also matters. A reference request, a practical need for an immediate answer, and participation in a philosophical course have different purposes. The present pedagogical principle applies where an educational relationship and aim have been established. Extending it to every request for assistance would require further argument. Even within a course, participants can reasonably negotiate the timing, intensity, or format of an exercise. A sequence should have alternatives when a case's personal content or presentation creates an avoidable barrier.

Preservation conditions in the formal model therefore require interpretation beyond their symbols. A set S can represent retention of an established distinction, but it cannot exhaust every relevant interest of a learner. A mathematically safe transition may still occur through a coercive or disrespectful interaction omitted from the state description. The model's admissibility conditions should make important constraints explicit, and the practical use of the model must remain subject to scrutiny beyond those constraints. This is an implication of the limited representation, rather than a reason to abandon formal clarity.

An accountable sequence combines guidance with a route for correction. The teacher organizes a tractable inquiry; the learner encounters reasons and alternatives; both can identify when the examples or assessment misrepresent the intended distinction. This arrangement gives the formal analysis an appropriate role. It supports comparison of conditional paths while leaving the authority to choose and revise those paths grounded in the educational relationship and the reasons offered within it.

11 Objections and Alternative Designs

This section tests the proposal against concerns about formal reduction, target dependence, efficiency, and evidence. Its objective is to identify conditions under which the framework should be revised or set aside. Each objection is examined through the specific claim it challenges and the narrower conclusion that remains defensible.

The first objection concerns the finite state model. Philosophical learning can transform the concepts through which its own progress is understood, whereas a finite transition relation presupposes a fixed description. This is a substantial limitation. The model can represent a bounded phase of an inquiry after its target and relevant alternatives have been provisionally specified. It can also contain an explicit outcome indicating that the available categories require revision. Once that outcome is reached, the earlier guarantee has exhausted its scope. The next phase needs a new model and an account of how its target relates to the earlier one.

The second objection concerns the strength of the learning target. A teacher can declare a goal so weak that nearly every sequence succeeds, or so restrictive that only a favored theory qualifies. The formal result remains true under either choice. Its educational value therefore depends on an independent target argument, including content review and the treatment of defensible disagreement. The proposed promise target is offered for criticism precisely at this level. If its distinctions prove irrelevant, incoherent, or unduly restrictive, a successful computation supplies little reason to retain it.

The third objection is that counterexample sequences may encourage excessive revision. A learner can become skilled at finding exceptions while losing confidence in warranted generalization. The frame-

work should therefore include cases that support a claim under its stated conditions, explanations of the costs of unnecessary distinctions, and opportunities to defend a position against an unsuitable objection. The desired competence concerns calibrated revision. It includes recognizing when a challenge changes a premise outside the current scope and when a simpler account remains adequate.

A fourth objection concerns efficiency. Exhaustive possible-state analysis can be disproportionate to an ordinary classroom decision, particularly when the transition model is uncertain and time is scarce. The paper's proposal can be used at several levels. A teacher may employ the conceptual distinctions without computing a policy. A curriculum designer may use a small state model to expose an overlooked prerequisite. A research study may compare fully specified policies within a narrow task. The computational apparatus is justified only where it clarifies a consequential choice that simpler analysis leaves obscure.

A fifth objection concerns assessment reactivity. Asking for reasons or a transfer response can itself change the learner's interpretation, making it difficult to separate learning from measurement. The joint episode relation explicitly allows that possibility. The empirical study must still identify which target it is estimating: initial understanding, understanding after a sequence, or responsiveness to a further explanation. A delayed transfer task may be informative about the last two while leaving the first unresolved. The assessment should state its temporal target and retain the intervention history needed to interpret it.

A sixth objection concerns the realism of a worst-case guarantee. Actual learners can produce unexpected responses, misunderstandings, or creative distinctions outside any finite relation. A guarantee across the represented possibilities can therefore coexist with failure in practice. The appropriate use is conditional: the model reveals what would follow if its representation were adequate. An observed violation prompts revision of the relation and a reevaluation of the guarantee. For some applications, a probabilistic model with calibrated uncertainty may be more useful, although it introduces additional evidential obligations and remains subject to the same target problem.

A seventh objection is that a learner can produce convincing explanations through imitation. This possibility limits every assessment that relies on outward performance. The framework responds by using several kinds of evidence: reasons for a judgment, application under changed surface features, identification of an invalid transfer, and explanation of a rejected formulation. These sources can constrain the imitation hypothesis without eliminating every alternative account. The resulting inference remains fallible and purpose-relative. A finite set of successful responses should be reported as evidence for a specified competence, with the limits of that inference visible.

An eighth objection concerns peer inquiry. The model centers an instructor selecting episodes, while philosophical discussion can proceed through reciprocal questioning among learners. A joint state can include several participants and an episode can be chosen through a shared procedure. That extension changes the authority structure and may alter the objective: success could require a shared capacity to articulate disagreement rather than an individual terminal state. The present derivation transfers only after those changes are specified. Peer discussion is therefore a substantive extension, with its own representation and evaluation needs.

These objections support a limited conclusion. The framework is useful when a bounded target can be defended, episode effects can be represented with meaningful uncertainty, and the resulting analysis clarifies a choice among sequences. It becomes less useful when the case descriptions lack warrant, the objective confounds competence and assent, or the transition relation merely encodes the desired conclusion. Those conditions provide revision and abandonment criteria for the proposed design, alongside its positive possibilities.

12 Open Questions and Evaluation Design

This section outlines the research needed to assess the proposal beyond its conditional arguments. The objective is to connect conceptual validation, empirical comparison, and open philosophical questions without presenting a planned study as completed evidence. The discussion specifies the target review, comparison conditions, assessment strategy, and circumstances that would require revision.

The first stage is a content review of the promise case bank. Philosophers and educators should examine whether each case supports the distinction assigned to it, whether the background assumptions are intelligible, and whether the range of accepted responses includes serious alternatives. Participants with different linguistic and educational backgrounds can help identify ambiguities or culturally specific presuppositions. Review disagreements should be retained and classified by their object: wording, case facts, inferential role, or normative interpretation. A case whose disagreement concerns the target itself may require revision before it is suitable for measuring that target.

Open Problem 12.1 (Content validation of interpretive targets). Develop a review procedure that can distinguish a learner's failure to use a target-relevant distinction from a defensible objection to the distinction, the case description, or its assigned educational role.

A comparative study could then examine several teaching arrangements. One condition supplies the argument and an explanation directly. Another uses a fixed order of cases. A third selects cases for immediate diagnostic discrimination under a stated candidate model. A fourth uses an adaptive sequence that represents prerequisites and possible learning changes. These conditions should be compared under justified resource constraints, with the amount and kind of explanation recorded. The study should avoid attributing an advantage to adaptivity when it merely reflects more time, additional content, or a different assessment standard.

The primary outcome should concern the declared interpretive competence. A baseline task can collect initial reasons while recording that baseline questioning may itself be formative. Immediate assessment can examine a new case, and a delayed task can examine retention or regeneration of the distinction. Transfer materials should vary surface details and include at least one case in which transfer should be limited or disputed. The coding scheme should distinguish correct identification of a distinction, a defensible alternative interpretation, a material inferential error, and an unresolved response. Raters should record reasons for difficult classifications rather than force every response into a binary success label.

The transition model requires its own evaluation. Researchers can compare predicted response-successor possibilities with observed sequences and identify episodes that generate outcomes outside the model. A successor state should be inferred from evidence appropriate to the target, with uncertainty retained when several states remain plausible. If the model is revised using the same episodes later used to evaluate it, the evaluation should identify that dependence. A separate held-out case family or later cohort would provide a stronger test of its explanatory and predictive adequacy.

Open Problem 12.2 (Empirical identification of episode effects). Identify which combinations of repeated assessment, process evidence, comparison conditions, and model revision can support a defensible distinction between diagnostic information and interpretive change produced during an episode.

The study should also examine effects omitted by a narrow answer score. Participants may become more willing to articulate uncertainty, more capable of identifying an ambiguous premise, or more able to challenge an unsuitable example. Conversely, a sequence may produce correct answers while discouraging independent questions. These possibilities matter to the declared educational purpose. They require carefully interpreted evidence, including participant accounts and analysis of the reasons expressed, alongside task performance. A single composite score could conceal tradeoffs that the proposal aims to preserve.

Accessibility and participation conditions should be specified before comparison. Researchers should record permitted supports, response formats, time accommodations, and routes for declining a personally difficult example. These conditions can be incorporated into the definition of an episode or treated as explicit aspects of the study design. The aim is to assess the intended philosophical task under meaningful conditions of participation. A difference caused by an avoidable presentation barrier should be interpreted accordingly, rather than incorporated silently into a judgment of conceptual competence.

Open Problem 12.3 (Revision across changing case domains). Develop a comparison method for learning sequences whose participants introduce new distinctions or cases, so that progress can be assessed while preserving the genealogy and changing scope of the target.

Several outcomes would require revising the proposal. If reason-based coding repeatedly collapses into agreement with one assessor's theory, the target has been inadequately operationalized. If a simple explanatory condition performs comparably under the same relevant resources, the additional adaptive apparatus needs a clearer justification. If transition violations are frequent or systematic, the finite model lacks adequacy for that setting. If the sequence restricts meaningful participant questions, its relation to the stated educational purpose requires reconsideration even when task scores improve.

The broader philosophical question concerns the appropriate endpoint of a learning sequence. A bounded lesson needs a criterion for completion, while philosophical understanding can open further questions. The proposal treats completion as achievement of a declared local repertoire under specified conditions. Further inquiry can begin from that achievement without implying that the learner has reached a final interpretation. Empirical study and philosophical criticism should therefore examine both the usefulness of local stabilization and the opportunities it leaves for continued revision.

13 Conclusion

This section consolidates the argument's scope and its implications for counterexample design. The discussion returns to the distinction between diagnosing an interpretation and supporting its development, then identifies the role of the conditional formal result and the remaining research obligations.

A counterexample sequence can be evaluated through the claim it challenges, the response evidence it produces, and the capacities it helps a learner develop. These objects require different warrants. A strongly discriminating question can improve an instructor's estimate of an initial interpretation while consuming an opportunity needed for preparation. A common explanation can prepare several interpretations for a later challenge, while a different prerequisite structure can make prior diagnosis necessary. The two finite witnesses establish the dependence of sequence choice on the represented learning process.

The backward construction gives a precise criterion for success within a finite, partially observed episode model. Its proof makes the policy's information, response uncertainty, preservation condition, and stopping rule explicit. The result supports comparison of conditional designs. Its educational significance depends on a defensible target and evidence for the episode relation.

The promise dossier shows how those requirements can be connected to formation, release, competing reasons, and transfer. A learner's substantive disagreement can satisfy the intended competence when the response identifies the relevant distinctions and gives reasons responsive to them. The teacher's authority therefore includes responsibility for explaining and reconsidering the case bank, target, and assessment.

The proposed contribution is an account of how these responsibilities and modeling choices fit together in a bounded philosophical inquiry. Content validation, empirical comparison, accessibility, and the treatment of new learner-generated distinctions remain open. Their investigation can refine both the sequencing method and the account of understanding it is intended to support.

References

- [1] Michelene T. H. Chi. Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*, 1:73–105, 2009.
- [2] Imre Lakatos. Proofs and Refutations (I). *The British Journal for the Philosophy of Science*, 14(53):1–25, 1963.
- [3] Steven M. LaValle. *Planning Algorithms*. Cambridge University Press, 2006. Chapter 11. Author-hosted text: <https://lavalle.pl/planning/ch11.pdf>.
- [4] Daniel L. Schwartz and John D. Bransford. A Time for Telling. *Cognition and Instruction*, 16(4):475–522, 1998.
- [5] Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N. Rafferty. An Overview of Machine Teaching. <https://arxiv.org/abs/1801.05927>, 2018. Version 1, 18 January 2018.