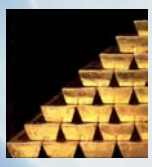
Gold's Theorem in Cognitive Science

Kent Johnson



Psych 245A Presentation

Erin Andrew

Does Gold's Theorem relate to language acquisition?

One path from the Logical Problem of Language Acquisition (LPLA) to innate knowledge:

Language learners never get negative evidence, so it is logically possible for them to a language just like the one they're learning, but containing a few extra sentences, like **John Mary kicked.* (LPLA) It's claimed that Gold's Theorem provides a nice formal proof of this.

So how is it that children don't learn languages like this?

Answer: They must have some sort of (presumably) innate language-learning mechanism that helps them limit the possible languages they can learn.

But *does* the argument from LPLA get support from Gold's theorem?

Gold's Theorem: A class of languages is unlearnable if it has Gold's Property

There have been many inaccurate statements made about Gold's theorem from both sides of the debate, if one understands Gold's theorem and its proof, it is clear that the theorem does not support nor dismantle the argument from LPLA, the two claims are independent, and only look similar when one is confused about the notion of learnability.

Structure of the Paper

- First we look at the formal proof of Gold's Theorem.
  - We will pay special attention to the formal definitions within the theorem, as they come up later in the paper.
- Then we will examine some of the false claims made about Gold's Theorem by people on both sides of the innate knowledge debate.
  - The point isn't that one side is right and the other is wrong, but that all the authors cited misunderstand the applicability of the theorem to natural language acquisition.
- Finally we will examine just what significance Gold's Theorem has for natural language acquisition.
  - Not very much, we will see that it's only when the notion of learnability is ambiguous that the theorem looks interesting to cognitive science.

Gold's Theorem Proof

Definitions:

- *Language*: A language L is any subset of Σ^* where Σ^* is the set of all finite sequences of some finite alphabet Σ (notice that we have not specified what this finite alphabet consists of).
  - So, $\Sigma=\{a_1,a_2,\ldots a_n\}$, and $\Sigma^*=\{\langle a_1\rangle, \langle a_{27},a_{32},a_3\rangle, \ldots,\langle a_2,a_3\rangle\}$, and $L=\{\langle a_1\rangle, \langle a_{27},a_{32}\rangle, \ldots, \langle a_2\rangle\}$
- *Environment*: an infinite sequence of sentences from the target language to be learned, with the provision that all sentences from the target language appears at least once in the sequence.
- *Learner*: A learner is a function ϕ that takes finite initial segments of the environment into the set of languages.
  - Intuitively, a learner gets some information about its environment, and makes a guess about the language it's speaking.
  - $\phi(\langle a_1,a_2,\ldots,a_n\rangle)=L_m$

Gold's Theorem Proof

More Definitions:

- A learner *learns a language L given an environment E* iff at some time, and for all times afterward, the learner correctly outputs the target language. (identification in the limit)
- A learner *learns L* iff the learner learns L given any Environment E (of sentences of L) whatsoever.
- A *learner learns a class C* iff the learner learns every language L such that $L\in C$
- A *class of languages C is learnable* iff there is a learner that learns C
  - Notice, that if a class of languages fails to be learnable then there is no single learner that learns all the languages in L, this does not imply that all learners fail to learn all languages in C.

Gold's Theorem Proof

Gold's Property: A class C of languages has Gold's property iff C contains an infinite subset of languages $\{L_\infty, L_1, L_2, \ldots\}$

such that $L_1 \subset L_2 \subset L_3 \subset \ldots$

and, for all sentences a, $a\in L_\infty$ iff $a\in L_i$ for some $i>0$

Gold's Theorem: Any class of languages with the Gold Property is unlearnable.

The Proof

- Basically we pick any arbitrary learner ϕ, and give ϕ sentences from L_1, until it converges on L_1, then we give it sentences from L_2, until it converges on L_2, and so on. So, ϕ either converges on a language in C, and doesn't learn any new languages, or it does not. If it does not, it doesn't learn a language, by definition. If it does converge, one of two things is true, either ϕ converges on L_i for some $i>0$, or it converges on L_∞. If it converges on L_i, then it cannot learn L_{i+1} and if it converges on L_∞, then there was some L_i that it failed to learn.

Misinterpretations of Gold's Theorem

- They come in many flavors:
  - Confusion about Language
  - Confusion about Learners
  - Confusion about Environment
  - Taking Gold's Theorem to apply only to 'external' languages, and not 'states of mind'
  - Claiming Gold's Theorem ignores the constituent structure of language
- Basically these criticisms try to show that the results of Gold's Theorem don't successfully disprove whatever it is that the authors want to argue, and in the process they show a fundamental misunderstanding of the theorem.

Confusions about the language

Deacon 1997:

"[Gold] proved that, without explicit error correction, the rules of a logical system with the structural complexity of a natural language could not be inductively discovered, even in theory. What makes them unlearnable is not just their complexity but the fact that the rules are not directly mapped to the surface forms of the sentences."

- But, the internal structure of the language is irrelevant- all Gold needs for his theorem is that a class of languages has Gold's property, the complexity of the language is irrelevant.
- Deacon takes Gold's Theorem to be saying something about the complexity of a language, rather than saying something about the relations between languages.

Confusions about the Language

Elman et al 1996

"Interestingly, sentences with relative clauses possess exactly the sort of structural features which may make (according to Gold 1967) a language unlearnable"

But Gold's theorem is about classes of languages, an not individual language. After all, a class containing one language is trivial to learn- there exists the constant function that always chooses that language.

Further, what is meant by a "learnable language?" Gold only defines learnability for classes of languages.

Confusions about the Learner

- Hirsh-Pasek and Golinkoff 1996
  - "Gold's learner was an unbiased learner"
- Cowie 1999
  - "children are not Gold-style learners: they do not test every logically possible grammar against the data."
- These authors suggest that human learners don't learn languages like Gold's learning functions, so they can safely get around the implications of Gold's theorem.
- But, the results of Gold's theorem apply to all learners, in the proof there were no constraints put on the learner at all. We showed that every learner could not learn an infinite class of languages with Gold's Property.

Confusions about the Environment

- Hirsh-Pasek and Golinkoff 1996
  - "Gold's learner heard a series of sentence strings and had to induce the units and rules of language, as if in a vacuum"
- But, the only constraint Gold puts on the learner's environment is that all the elements of the language be finite sequences drawn from a finite set- we can think of these as words, or phonemes, or anything we like. We can even include context or semantics in the sequence.
  - A 'sentence' could code information about whatever we like, $\langle a_{1l}, \ldots, a_{1n}, a_{2m}, \ldots, a_{2p}, \ldots a_{jk}\rangle$ where syntax is described with the sequence a_{1l} to a_{1n}, and semantics, and other information relevant to understanding the sentence's meaning is described with a_{2m} to a_{jk}.

Gold's theorem is about sets of sentences, not minds

- Chomsky claims that Gold-style models concern only E-languages, not I-languages.
- But, there is no reason that we can't interpret Gold-style learning models as being about the mind.
  - A learner looks like:

$$\Phi(\langle s_1,\ldots s_n\rangle)=L$$

An acceptable interpretation: Φ represents a learning strategy for a particular learner such that when presented with an environment $\langle s_1,\ldots s_n\rangle$, her brain is configured (all things being equal) so as to instantiate language L- that is, to 'think' L is the language she's learning

Gold's theorem ignores the constituent structure of language

- Chomsky claims that Gold's theorem only concerns the surface structure of grammar, and ignores the constituent structure of natural language. (Weak vs. Strong generative capacity.)
  - Ex. The rule X→aXa is different from the rule X→Xaa but both can produce the string aaaa, but one's constituent structure looks like [a[aa]a] and the other looks like [[aa]aa]. Gold's theorem, claims Chomsky only concerns the level that produces aaaa, and is not sensitive to the distinction between [a[aa]a] and [[aa]aa].
    - So, here we see a difference between strong and weak generative capacity- a language with strong generative capacity can discriminate between [a[aa]a], and aaaa, while a language with weak generative capacity cannot.

- But, Gold's theorem can be interpreted to be "directly" about classes of languages with the strongly generative feature. There is no reason why the learner couldn't get sentences that contain information about its constituent structure, or its generative history. This solution is a lot like the one we saw in response to the objection that Gold-style learners get less information about their environment than real kids.

Another Response

Gold's Theorem shows us that if a class C has Gold's Property, and the languages in C are differentiated by their weak generative capacity, then C is unlearnable.

If C is unlearnable and C ⊆ C*, then C* is unlearnable

C is learnable when the languages are individuated by their strong generative capacity, only if C is learnable when they are individuated by their weak generative capacity.

That is, all learners that can learn strongly generative languages can learn weakly generative languages.

Hence, if for each language in C, there are several languages in C* differentiated by their strong generative capacity, then C* is unlearnable.

Basically, the weakly generative languages are a subset of the strongly generative ones, so this response won't help get around Gold's theorem.

Claims about Gold's theorem we've seen

- Yang 2008
  - It is well known that under the Gold framework, context-free languages are not learnable…One conclusion does emerge convincingly from both frameworks : learning is not possible unless the hypothesis space is tightly constrained by prior knowledge, which can be broadly identified as Universal Grammar.
- Elman 1999
  - "Gold (1967) was able to show that formal languages of the class which allow embeddings…cannot be learned inductively on the basis of positive input only."

Significance of Gold's Theorem

- "In sum, Gold's Theorem appears interesting to cognitive science when identifiability and acquirability are confused. When we distinguish these notions, we undermine the argument that Gold's Theorem is supposed to support."
- "In general, the relation of Gold's Theorem to normal child language acquisition is analogous to the relation between Gödel's first incompleteness theorem and the production of calculators."

Identifiability vs. Acquirability

- A class of languages is identifiable iff there exists a function Φ such that for any environment E for any language L in C, Φ permanently converges on the target L after some finite time.
  - ∀e∃t(a learner learns L for any environment e in some time t)
- A class C of natural languages is acquirable iff given almost any normal human child and almost any normal linguistic environment for any language L in C, the child will acquire L (or something sufficiently similar to L) as a native language between the ages of one and five years.
  - ∃t∀e(a learner learns L at time t given any environment e)

Identifiability vs. Acquirability

- In some respects it is easier for a class of languages to be identifiable than acquirable.
  - There is a larger class of learners able to identify languages, than acquire them. Acquirability quantifies only over normal human children, whereas Identifiability quantifies over all learners.
- In other respects, it's easier for a class of languages to be acquirable than identifiable
  - Identifiability quantifies over all environments, so it's easier for a problematic environment to crop up for identifiability than acquirability.

Argument to innate knowledge revealed!

- 1. If there are no constraints on language acquisition, then either children have access to negative data or natural languages are unlearnable.
- 2. If they exist, the constraints in question must be innate.
- 3. Children don't have access to negative data.
- 4. Natural languages are learnable.
- 5. ∴There are innate constraints on language acquisition.

Where this argument fails:

- "We can now see the dilemma for interpreting (1)-(5): should learnability be interpreted as identifiability or acquirability? If we use acquirability, then (4) looks true, but no argument whatsoever has been offered for (1). (1) is suggested by Gold's Theorem, but Gold's Theorem is about identifiability, which we've seen is strikingly different from acquirability. If we interpret (1) and (4) in terms of identifiability, then it's unclear why (4) should be true. Just because a class C of (natural) languages is acquirable by children doesn't mean that there couldn't exist a collection of logically possible but highly abnormal environments that would fool all learning functions... In general, the relation of Gold's Theorem to normal child language acquisition is analogous to the relation between Gödel's first incompleteness theorem and the production of calculators."

Gold's Theorem vs. LPLA

- Indirect negative evidence:
  - Failed predictions might be able to serve as indirect negative evidence to help get around LPLA, but it's irrelevant for Gold's Theorem.
- On the other hand, direct negative evidence is relevant for Gold's Theorem, and it would also help get around LPLA, so Gold's theorem is stronger in this respect.
- So, we see that Gold's Theorem is a different beast than LPLA, and not, in fact a logical formulation of the problem.

Appendix

- (A1) *Gold's classes of languages*: Finite ⊂ Superfinite ⊂ Regular ⊂ Context-free ⊂ Context-sensitive ⊂ Primitive Recursive ⊂ Recursive ⊂ Recursively Enumerable.
- (A2) If C⊆C`, then if C` is learnable, so is C. Similarly, if C is not learnable, neither is C`.
- (A3) If a learner learns a class of languages with the tester naming relation (characteristic function), then there is a learner who learns it with the generator naming relation (the turing machine that spits out a list of all the sentences in a language).

Appendix

- (A4) A learner learns a language using a class E of environments only if it learns it using any subset of E.
- (A5) *Learning with informant*: Using any form of informant and either form of learner (i.e. ones that guess with testers or generators), the class of Primitive Recursive languages is identifiable in the limit, but the class of Recursive languages is not.
- (A6) *Learning with text*: Using any form of text and either form of learner except the combination of Primitive Recursive text and the generator naming relation, the class of finite languages is identifiable in the limit, but the superfinite class of languages is not.
- (A7) *An anomaly*: Using Primitive Recursive text and the generator naming relation, the class of Recursively Enumerable languages is identifiable in the limit.

Appendix

- (A8) *Theorem I.8.* Using information presentation by Recursive text and the generator-naming relation, any class of languages which contains all finite languages and at least one infinite language L is not identifiable in the limit. (Gold 1967, 470)
- (A9) A class *C* of Recursively Enumerable languages is identifiable using arbitrary text iff every language *L* in *C* has a finite subset *T* such that for all $L` \in C$, if $T \subseteq L`$ then $L` \not\subset L$ (Angluin 1980, 121- 22).