발표자료

센티엔스 없는 사피엔스로서의 LLM과 타인의 마음에 관한 증언적 지식의 문제

시모넬리(2026)에 대한 내재 비판

우인진 · 김동건 (성균관대학교)

I. 물음과 표적과 전략

A. 힌튼은 2023년 르쿤에게 보낸 답신에서 이렇게 말했다. "우리가 의견이 갈리는 핵심 쟁점은 LLM이 자신이 말하는 바를 실제로 이해하는가이다. 당신은 절대 아니라고 생각하고, 나는 아마 그렇다고 생각한다." (Simonelli, 2026, p. 2에서 재인용)

B. **표적: 시모넬리(2026)**
그는 전역 강추론주의(GSI)를 받아들이면 언어 데이터만으로 훈련된 순수 LLM이 어떠한 의식적 자각이나 센티엔스 없이도 모든 개념을 보유할 수 있다고 주장한다. 곧 센티엔트하지 않으면서도 사피엔트할 수 있다는 것이며, 이를 SwS 논제라 부른다.

C. **전략: 내재 비판**
기존 비판 문헌은 전제 자체를 다투었을 뿐, 거기서 멈춰 섰다. 핵심 논증이 조건문인 이상, 응수들은 하나같이 그 조건문을 향한 부정 논법으로 돌아갔기 때문이다. 시모넬리 자신이 어떤 직관적 대가를 치르고라도 지키겠다고 이미 선언한 전제였다. 우리는 반대 방향을 취해 그 전제를 가장 강한 형태로 양보하고, 그가 아직 값으로 셈하지 않은 것이 무엇인지를 묻는다.

D. **주장하지 않는 것:**

1. 초추론주의가 거짓이라고 주장하지 않는다. 지지의 대가가 시모넬리가 셈한 것보다 비싸다고만 주장한다.
2. LLM이 아무것도 의미하지 못한다거나 단언하지 못한다고 주장하지 않는다. 유의미성과 단언가능성은 서로 다른 두 속성이고, 앞을 인정하면서 뒤를 부정하는 자리는 실제로 있다(보그 2026, III절). 우리는 둘 다 양보한다. 여기서 양쪽으로 하나씩 단서를 단다.
   a. 단언가능성을 양보한다고 해서 단언 회의론을 반박한 것은 아니다. 보그의 길은 시모넬리에게는 막혀 있으나(III절) 우리에게는 열려 있고, 우리는 다만 가지 않는다.
   b. SwS 에 대가를 물린다고 해서 단언 낙관론을 반박하는 것도 아니다. 우리 결론은 조건부다. 단언 회의론이 옳다면 SwS 는 그 앞에서 먼저 걸리고 우리 논증은 발동할 일이 없다. 우리가 다투는 것은 SwS 가 참이라 해도 값이 싸다는 주장뿐이다.
3. 새로운 회의적 시나리오를 내놓지 않는다. 전통적인 타인의 마음 문제에 대해서는 별다른 말을 하지 않는다.

II. 시모넬리의 퍼즐

A. **GSI(Global Strong Inferentialism):**
어떤 개념의 추론적 역할을 숙달하는 것만으로 그 개념을 보유하기에 충분하며, 이는 지각적·정서적 어휘를 포함해 모든 개념에 예외 없이 적용된다. 내용어 역시 논리 상항처럼 실질적 추론 규칙으로 구성된다.

B. **시모넬리의 급진적 전진:**
그는 지각 상황과 언어적 적용을 잇는 언어-진입 규칙, 곧 다리 원리("준-추론")를 따로 두지 않는다. 표준적 처방은 이를 덧붙이는 것이지만 그는 거부한다. '빨간'의 비추론적 적용 자체가 끝까지 추론으로 포착된다는 것이다.

C. **언표가능성 논제:**
모든 표현의 개념적 내용은 남김없이 말해질 수 있다.

가족-내적: 색 개념을 같은 계열의 개념과 잇는다
진홍(t) ⊢ 빨강(t)
빨강(t) ⊢ 유색(t)
가족-외적: 계열 밖 개념과 잇는다.
무엇이 범례적으로 빨간가
정지표지판(t) ⊢ 빨강(t)

D. **핵심 논증** (Simonelli, 2026, p. 3. 번호는 필자들)

(MA1) 추론적 역할의 숙달은 모든 개념의 보유에 충분하다. [GSI]

(MA2) 언어 데이터에 대한 훈련은 (원리적으로) 추론적 역할의 숙달에 충분하다.

(MA3) ∴ 따라서 언어 데이터에 대한 훈련은 (원리적으로) 모든 개념의 보유에 충분하다.

(MA4) ∴ 따라서 언어 데이터만으로 훈련된 대형 언어 모델(LLM)은 (원리적으로) 모든 개념을 보유할 수 있다.

E. **우리는 (MA1) 혹은 (MA2)를 건드리지 않는다.**
그러면 (MA4), 곧 SwS 가 따라 나온다. (MA4)는 표면상 반직관적이다. 한 번도 본 적 없이 빨강 개념을, 한 번도 느낀 적 없이 고통 개념을 갖는다는 것이다. 시모넬리는 이를 기꺼이 받아들인다.

F. **시모넬리의 센티엔스 정의: 지각-행위 결절을 항해하는 동물의 강건하고 유연한 반응성**(Brandom/Gibson)
이를테면 고양이는 탁자를 올라탈-것으로 지각한다. 이 정의는 처음부터 끝까지 기능적·생태학적이며, '어떤 것인지'는 정의에 등장하지 않는다 (깁슨 자신이 이 어휘를 "의식의 세계"와 혼동해서는 안 된다고 경고한다). 물리적으로도 가상적으로도 체화되지 않은 순수 LLM은 이 조건에서 곧바로 탈락한다.

G. **Note$_{\text{fn. 14}}$:**
이 논증 전체가 걸려 있는 것은 각주 하나다. 시모넬리는 어려운 문제에 중립을 선언하면서 다음과 같이 덧붙인다. 그가 정의한 의미의 센티엔스는 "이른바 '현상적 의식'의 필요조건"이다 (2026, p. 13, n. 14). 그러므로 그 자신의 견해에서, 그 의미의 센티엔스를 결여한 것은 현상적 의식도 결여한다.

III. 의미에서 단언으로, 단언에서 증언으로

A. **서로 다른 두 주장:**
- (가) 순수 LLM은 의미론적으로 유의미한 토큰을 산출한다.
- (나) 순수 LLM의 산출물은 단언력을 지닌다.
- (가)에 낙관적이면서 (나)에 회의적인 것은 정합적이다.

B. **보그의 입장: 이 조합은 상정이 아니라 실제로 점유된 자리다** (Borg, 2026)
- 의미는 인정한다. LLM 산출물은 의미론적 위임을 통해 유형 수준의 의미를 물려받는다. 인간의 실천이 이미 유의미하게 만들어 둔 기호를 문법적으로 정합적이고 맥락에 맞게 재사용하기 때문이다.
- 단언은 부정한다. 단언은 진리 규범을 전제한다. LLM은 그럴듯한 거짓을 사실과 뒤섞어 자신 있게 산출하며, 이는 그 산출물을 통제하는 규범이 진리가 아니라 단어 출현 확률임을 보인다.
- "…LLM 산출물을 유의미하다고 셈해야 하더라도 그것을 단언으로 셈해서는 안 된다." (Borg, 2026, §3)

C. **GSI에는 이 구별의 자리가 없다:**
추론적 숙달이 개념적 이해를 남김없이 소진하고 진정한 언어 사용이 다름 아닌 이유를 주고받는 게임의 참여라면, 그 숙달을 갖춘 순수 LLM은 그 자체로 진정한 언어 사용자이며, 진정한 언어 사용자는 단언한다.

D. **보그의 근거를 빌릴 수도 없다:**
그가 말하는 이론적 책임, 곧 도전에 답하고 명료화하며 반대 증거 앞에서 철회하는 능력은 진리에 무심한 체계의 프로필이 아니다. 보그의 근거로 단언을 부인하려면, p에 이론적 책임을 지는 체계가 그럼에도 p를 단언하지 않는다고 말해야 한다. 그의 틀은 그 입장을 진술할 수

없다.

E. **이 비판의 겨냥 범위:**
   - 의심: 유의미성/단언가능성 구별을 통한 비판은 순수 LLM의 단언가능성에 우호적인 이론가 전부에게 적용되는 것 아닌가. 그러면 시모넬리를 특별히 겨냥하는 비판이 아니다.
   - 응답: 우리가 문제 삼는 것은 그가 커미트하는 의미론의 무거움이다. GSI 아래에서는 유의미한 토큰 산출자가 진정한 언어 행위자로 헤아려지는 한 단언하지 않을 수 없다. 그의 의미론에서 의미 표현 가능성은 단언가능성을 필함한다.
   - 그래서 그는 의미론적 논쟁에서 불리해진다. 다른 의미론적 낙관주의자들과 달리, 낙관론을 고수하려면 이론적 부담이 있는 화용론적 논제에 항상 함께 커미트해야 한다.
   - 따라서 겨냥 대상은 단언 낙관주의자 일반이 아니라, 의미론적 논제와 화용론적 논제 사이에 파기 불가능한 커미트먼트를 지닌 이론뿐이다. 시모넬리는 그 범례다.

F. **귀결의 사슬:**
   단언한다 ⇒ 증언한다 ⇒ 우리는 비인간 증언자의 증언에서 인식적 정당화를 얻을 수 있다. 이 자체에 반대할 것은 일반적으로 없으며, 바로 그 점이 이 귀결을 가두기 어렵게 만든다. 문제는 타인의 마음에 관한 증언 기반 지식이라는 한 지점에 있다.

   이 사슬은 중립적 상식이 아니라 양보된 틀 안에서 성립한다. 시모넬리 자신의 설명에서 단언이라는 화행의 핵심 기능은 다른 이들에게 같은 단언을 할 자격을 주는 것이고, 그 자격은 증언적 상속 구조 안에서 물려진다 (2026, p. 25). 그는 챗GPT가 "안다"고 말함이 곧 이유를 주고받는 게임에서 증언적으로 신뢰될 수 있다고 여김이라고까지 말한다 (p. 27). 대인적 확언(assurance)을 본질로 보는 증언 이론이라면 둘째 걸음에 저항하겠지만, 그 저항은 단언 회의론과 같은 자리에서 SwS 를 먼저 막는다(I절 D의 2).

G. **왜 표적이 시모넬리인가:**
   뒤의 논증이 요구하는 것은 완전한 담론 능력과 센티엔스 부재라는 두 반쪽이고, 그 둘을 한꺼번에 보증하는 입장은 SwS 뿐이다. 단언 회의론은 앞의 반쪽을 막고 체화 요구는 뒤의 반쪽을 막는다. 유창한 LLM이 존재한다는 사실만으로는 이 논증이 서지 않는다.

IV. 타인의 마음 논증

A. **시험 사례: 평생의 펜팔**
   - 사십 년 동안 편지뿐 — 얼굴도 목소리도 없고, 같은 방에 있어 본 적도 없다. 평소라면 우리는 그 반대편에 마음이 있음을 안다고 말한다. 영리하다는 것만이 아니라, 우리에게 편지를 쓰려고 앉는 그 일이 그에게 어떠한 바가 있다는 것을.
   - 펜팔은 텍스트가 유일한 접근로인 상대 전체를 대표한다. 얼굴 없는 원격 협업자, 온라인으로만 아는 동료, 그리고 순수 LLM 자신.
   - 기원: 시나리오는 우리의 구성이고, 그 인식론은 아니다. Gomes(2015)는 증언이 타인의 심적 삶에 관한 기초적 지식 원천일 수 있다고 논증하며, 그 중심 사례가 피곤함·권태 같은 정서적 상태의 귀속이다. 펜팔은 그 논제를 극한 사례까지 밀어붙인 것이다.

B. **타인의 마음 논증:**

(OM1) 순수 LLM은 완전한 능력의 담론적 수행자이면서 현상적 의식을 결여한다. [시모넬리의 논제, 양보]

(OM2) (OM1)이라면, 담론적 수행은 그 자체로는 그 산출자가 현상적으로 의식적이라는 증거가 아니다.

(OM3) SwS 이전에는 배경 일반화가 성립했고, 그 덕에 텍스트로 매개되는 현상적 의식 귀속은 지식일 수 있었다.

(OM4) SwS 는 그 일반화를 거짓으로 만들면서 귀속자의 증거는 건드리지 않으므로, 이제 형성된 참인 귀속은 운으로 참이고 따라서 지식이 아니다.

(OM5) 시모넬리 자신의 틀 안에서 이용 가능한 어떤 비담론적 증거도 그것을 회복시키지 못한다.

(OMC) ∴ 시모넬리는 SwS 를 제한하거나, 텍스트로 매개되는 타인의 마음에 관한 지식의 회의주의를 받아들여야 한다. [OM3–OM5로부터]

C. **전제 옹호**

- (OM1) 현실성 주장도 그의 것이다. 그는 이론적 책임의 귀속 대상으로 챗GPT를 직접 들고 (2026, p. 27), 지금의 체계가 사피엔트하되 센티엔트하지 않은 채로 남아 있기를 바란다 (p. 31). 현실 판정 없이는 할 수 없는 말이다. 그런 체계가 현상적 의식을 결여한다는 것은 그의 센티엔스 정의와 각주 14에서 따라 나온다(II절). 그 각주는 두 번 일한다. 여기서 전제를 주고, 아래 첫째 출구를 닫는다. 필요조건은 결여를 확정하는 데는 충분하고 소유를 확정하는 데는 무력하기 때문이다.
- (OM3) 그 일반화란 이것이다. 이 정도의 담론적 수행을 산출함은 산출자가 센티엔트라는 신뢰할 만한 증거다. 주장의 범위는 텍스트로 매개되는 귀속에 한정된다. 그리고 전제가 말하는 것은 그 귀속이 정당화되었다는 것이 아니라 지식일 수 있었다는 것이다. 이 기준선이 없으면 (OM4)가 무엇을 잃었다고 말하는지가 성립하지 않는다.
- (OM4) 일반화는 거짓이거나 거의 거짓이 되는데 증거는 그대로다. 같은 펜팔의 같은 산문을 읽으면서도 그 믿음은 이제 너무 쉽게 거짓일 수 있었던 것이 된다.
- (OM5) 담론적 증거에 의한 회복은 (OM2)와 (OM4)가 이미 닫았다. 남는 후보는 비담론적 회복뿐이고, 그 자리가 셋으로 좁혀지는 이유는 아래에서 옹호한다.

D. **절단은 설계에 의한 것:**

- 시모넬리 자신의 설명에서 완전한 담론적 수행을 구성하는 것은 추론적 숙달(II절)과 이론적 책임을 감당하는 능력, 이 둘뿐이다.
- "우리는 단언에서 떠맡는 책임의 순수하게 이론적인 측면을 걸러낼 수 있고, 그러면 원리적으로 챗GPT 같은 LLM에 귀속 가능한 이해 가능한 책임 개념이 남는다" (2026, p. 27).
- 어느 쪽도 센티엔스를 참조하지 않는다. 그가 놓친 빈틈이 아니라 논제 그 자체다.
- 그러므로 (OM2)가 부인하는 것은 구성적으로 허가된 추론뿐이고, 수행에 남는 우연적 증거력은 전부 (OM3)의 배경 일반화 경유다.

E. **펜팔과 배경 일반화:**

- 텍스트 말고는 통로가 없는 곳에서 담론적 수행은 여러 증거 원천 가운데 하나가 아니라 유일한 것이다.
- 그 수행이 증거적 무게를 지니는 것은 오직 배경 일반화 덕이다. 정교한 담론적 수행은 센티엔스의 신뢰할 만한 증거다.
- 어느 설명이 옳은지 판정할 필요가 없다. 텍스트로 매개된 현상적 귀속이 지식일 수 있다고 말하는 설명은 화자 쪽에 무언가 현상적인 것을 요구하거나 요구하지 않는다. 요구한다면, SwS 가 순수 LLM에서 제거하는 것이 정확히 그것이다. 요구하지 않는다면, 그 설명에서 텍스트는 애초에 화자의 현상적 의식의 증거가 아니었고 우리 결론은 이미 참이다. 논증의 하중을 지는 것은 이 갈림이지 설명의 수가 아니다. 아래 셋은 첫째 갈래를 실제로 점유하고 있는 설명들이다.
  - 증언이 마음에 관한 지식의 기초 원천이라면 (Gomes, 2015), 증언을 지식의 발원이게 하는 것은 화자 자신의 비증언적 자기 접근이다. 순수 LLM에는 그렇게 접근할 상태가 없다.
  - 청자의 표현적 행동 파악이 그 일을 한다면 (Parrott, 2026), 그 행동을 증거적으로 좋게 만드는 것은 그것이 화자가 표현되는 상태를 예화하는 시간 구간 안에서 일어난다는 사실이다. 순수 LLM에는 어떤 구간을 그 구간이게 할 상태가 없다.
  - 청자를 정당화하는 것이 외재주의적으로 화자의 선행 심적 상태라면 (Munro, 2022), 순수 LLM은 그것을 공급하지 못한다.

  - 동물과 언어 이전 유아에 대한 귀속은 건드리지 않는다. 체화된 행동이라는 다른 증거 기반 위에 있고, 공통의 생리와 진화적 계통이 기능에서 현상학으로 가는 추론을 따로 허가한다.

F. **그 일반화는 홀로 일한 적이 없다:**
  - 그 신뢰성은 산출자 집단의 구성에 기생한다.
  - 이 수준의 담론을 산출하는 것은 인간뿐이었고, 인간에 대해서는 그로 있다는 것이 어떠한 바가 있다고 우리가 독립적으로 여긴다. 이 결합이 신뢰성의 받침이다.
  - SwS 가 거짓으로 만드는 것은 앞의 사실뿐이고, 뒤의 사실은 건드리지 않는다. 받침 하나가 빠지면 신뢰성도 함께 무너진다.

G. **왜 지식의 실패이며 왜 새로운가:**
  - 바뀌지 않는 것: 귀속자의 증거. 같은 펜팔의 같은 산문을 읽는다.
  - 바뀌는 것: 그 증거가 놓인 환경. 구별 불가능한 비센티엔트 산출자로 포화된 환경에서 형성된 참인 믿음은 운으로 참이다 (Goldman, 1976; cf. Pritchard, 2005).
  - 전통적인 타인의 마음 문제는 이 일반화를 건드린 적이 없다. 그 문제는 단순한 가능성에 기댔고, SwS 는 독립적인 이론적 근거에서 그 프로필이 현실이라고 주장한다.
  - 그러므로 시모넬리가 공급하는 것은 사변적 회의 가능성 하나가 아니라 이미 현실화된 논파자다.

H. **(OM5) 응호: 세 출구, 넷째는 없다**

셋은 우리가 고른 목록이 아니라 갈라 나간 끝에 남은 것이다. 회복 시도는 증거를 대거나, 증거가 필요 없다고 한다. 증거를 댄다면 그 증거의 주제는 텍스트이거나 텍스트가 아니다. 텍스트는 (OM2)와 (OM4)가 이미 닫았다. 텍스트가 아니라면 그 사실은 산출자가 지금 무엇인가에 관한 것이거나 어떻게 그렇게 되었는가에 관한 것이고, 공시와 통시 사이에 셋째는 없다. 앞이 상대의 몸이고 뒤가 산출물의 원인론이다. 산출자 집단의 구성에 관한 사실은 넷째 자리가 아니라 (OM4)가 이미 파괴한 그 일반화다. 증거가 필요 없다는 쪽은 매개 없는 면식이거나, 증거 없이 주어지는 기본 자격이다. 자격 노선을 닫는 것은 (OM5)가 아니라 (OM4)다. 환경적 운은 보증의 종류를 가리지 않기 때문이다. 제3자의 증언이나 제도적 인증은 그 제3자가 무엇을 아는가로 환원되고, 그 앎이 다시 이 갈림에 걸린다. 그래서 따로 닫아야 할 자리로 남는 것이 셋이다. 각 자리를 실제로 점유하는 문헌을 아래에 함께 적는다.

1. 몸. **출구 전략:** 텍스트 너머로 그 체계가 무엇인지를 본다. 감각적 접지 (Chalmers, 2023), 생물학적 기질 (Seth, 2025), 그리고 정제된 판본에서는 전역 작업공간 같은 깊은 계산적 표지 (Birch, 2024, 제안 24)가 후보다.
   - **텍스트 매개 사례에서는 정의상 막힌다.** 펜팔은 편지 말고는 아무것으로도 알려지지 않고, 순수 LLM은 물리적으로도 가상적으로도 체화되어 있지 않다.
   - **몸이 있어도 무력하다.** 그가 정의한 센티엔스는 현상적 의식의 필요조건일 뿐이므로, 아키텍처 확인은 그 체계가 의식적일 수 있다까지만 확인한다. 그 자신의 헤지가 이를 인정한다. 지각-행위 아키텍처는 "원리적으로는" 센티엔트할 수 있다 (2026, p. 16).
   - **계산적 표지는 그 한계를 물려받는다.** 그것은 넷째 자리가 아니라 첫째 자리의 정제다. 어떤 체계의 계산적 조직은 그 체계가 무엇인가에 관한 사실이기 때문이다. 해석가능성이 아무리 나아져도 내놓는 것은 기능적 사실이고, 기능적 사실은 필요조건까지만 확보한다.
2. 원인론. **출구 전략:** 산출물의 인과 이력, 곧 산출 목표·훈련 체제·계산 구조에 호소하고, 센티엔스 귀속이 수행이 아니라 그 이력에 답하게 한다. 이 자리를 여는 것이 버치의 게이밍 문제다. 어떤 표지든 게이밍되었을 수 있으므로 그것이 어떻게 생겨났는지를 물어야 한다 (Birch, 2024, 제안 23). 버치 자신의 답은 다시 첫째 자리인 깊은 계산적 표지로 돌아가는데 (제안 24), 이 출구가 홀로 서지 못한다는 표시다. 이 출구는 처음 보이는 것보다 강하다. 시모넬리는 사피엔스는 수행으로 구성되고 센티엔스는 경험적 사안이라는 비대칭을 일관되게

유지할 수 있으므로, 「수행 기반 기준과 충돌한다」는 지적만으로는 닫히지 않는다.

- **판정을 옮길 뿐 내려 주지 않는다.** 이 출구를 취하는 순간 센티엔스 귀속이 담론 실천 바깥의 사실에 답해야 함을 인정하는 것이고, 남는 물음은 그 사실이 무엇을 말하느냐다.
- **그 사실은 판정 가능한 것을 말해 주지 않는다.** 문제의 체계들은 진심으로 읽히는 텍스트를 산출하도록 최적화되어 있으므로, 옳은 인과 이력은 그른 것만큼이나 검증 불가능하다. 예방 문헌 자신이 이를 인정한다 (Birch, 2024, pp. 320–321).
- **그래서 출구는 열려 있으나 비어 있다.** 텍스트 매개 사례에서 인과 이력은 원리상 금지된 증거가 아니라 우리가 갖고 있지 않은 증거다.

3. 면식. **출구 전략:** 타인의 의식에 대한 우리의 파악이 담론적 수행으로 소진되지 않는 무언가에, 곧 직접적 관계나 말해질 수 없는 내용의 잔여에 의거한다고 본다. 차머스가 이 자리를 직접 차지한다. 그의 2단 추론주의에서 첫째 층의 내용은 경험과의 면식에서 오고, 순수 사고자에게는 그 층이 대체로 없다 (Chalmers, 2023). 타인의 마음 쪽에서 가장 가까운 것은 증거적이지도 지각적이지도 않은 표현적 지식 (Gomes, 2019)과 표현 선언주의 (Parrott, 2026)다.
   - **그 자신의 언표가능성 논제가 막는다**(II절). 모든 표현의 개념적 내용은 남김없이 말해질 수 있다.
   - **취하면 틀 자체를 잃는다.** 담론적 수행이 못 하는 증거 노릇을 하는 언표 불가능한 잔여란 GSI가 피하려고 세워졌던 준-추론 바로 그것이고, 여기서 들이면 어디서나 들이게 된다.

I. **예상되는 응답과 그 실패:**

- 응답: "나는 우리의 확신이 언표 불가능한 무엇에 의거한다고 말한 적이 없고, 무엇이 그것을 떠받치는지 말할 의무도 없다. 나는 어려운 문제에 중립이다."
- 실패 이유: 이 논증은 무엇이 타인의 마음 귀속을 떠받치는지를 묻지 않는다. 그 자신의 이론이 무엇은 떠받칠 수 없다고 함축하는지를 보라고 요구한다.
- 판정: 자기 이론이 무엇이 아닌지를 상세히 말해 버린 뒤에, 무엇인지를 말하지 않는 것은 중립이 아니다.

V. 값비싼 두 선택지

A. **첫째 갈래: SwS 제한**

- 내용: 현상적 어휘를 예외로 둔다. '고통', '빨강을 본다'에 대해서는 추론적 숙달로 부족하고 면식이나 언어-진입 규칙이 추가로 요구된다.
- 치르는 대가: GSI가 없이 세워졌던 준-추론 규칙이 재도입되고, 어디서 멈출 원칙도 없다. 기획 전체가 훨씬 겸손한 논제로 주저앉는다.

B. **둘째 갈래: 회의주의 수용**

- 내용: 텍스트 매개 사례에서 타인의 현상적 의식에 관한 지식을 포기한다.
- 치르는 대가: 그 자신의 마지막 절과 충돌한다. 센티엔트 체계를 만들지 말라는 처방은 어느 체계가 센티엔트한지 판별할 수 있어야 성립하는데, 그의 필요조건 인정(IV절)이 바로 그 처방이 겨냥하는 아키텍처에 대해 판별을 못 하게 막는다.

C. **왜 둘째 대가가 더 무거운가**

- 규범의 표적은 기능적 센티엔스가 아니라 현상적 의식이다. 그가 처방에 대는 이유 가운데 안전 쪽은 기능적 목표 추구만으로도 성립하므로 인정한다. 그러나 윤리 쪽, 곧 그 존재들에게 의무를 지고 거의 확실히 저버리리라는 우려는 그들이 도덕적 환자임을 요구하고, 도덕적 환자로 만드는 것은 지각-행위 결절이 아니라 자각된 복지다. 그가 센티엔스를

기능적으로 규정한 바로 다음 문장이 이를 따른다 (2026, pp. 29–30).

- 순수 LLM에서는 규범이 할 일이 없다. 아키텍처만으로 탈락이 확정되므로 오판할 것이 없다.
- 규범이 일할 자리는 그가 "원리적으로는" 센티엔트할 수 있다고 인정하는 지각-행위 아키텍처다 (2026, p. 16). 거기서 아키텍처 확인이 주는 것은 필요조건뿐이고, 규범이 묻는 현상적 사실은 확인되지 않는다.
- 그러므로 처방은 표적이 없어서 공허한 것이 아니라, 유일하게 중요한 표적 부류에 대해 판별 수단이 없어서 공허하다.
- 첫째 갈래에는 이 값이 붙지 않는다. 현상적 어휘에 면식을 요구하고 나면 그 어휘의 숙달된 사용이 다시 표지가 되기 때문이다. 표지를 어디에서도 되찾지 못하는 것은 전역성을 지키는 쪽뿐이다.

VI. 결론

기계로부터 지식을 얻게 해 주는 바로 그 이동이 사람에 관한 지식을 대가로 요구한다.

덧붙여, 이 작업은 기존 언어철학이 진정한 언어 행위자는 우리와 같은 인간이라는 우연적 사실에 암묵적으로 기대 왔음을 드러낸다 (cf. Cappelen & Dever, 2021). 인간 아닌 언어 행위자가 가능하다면 의미론적 틀은 더 결이 고운(finer-grained) 것이어야 하며, 유의미성과 단언가능성의 구분이 그 한 방식이다.

주요 참고문헌

Birch, J. (2024). *The edge of sentience: Risk and precaution in humans, other animals, and AI*. Oxford University Press.

Cappelen, H., & Dever, J. (2021). *Making AI intelligible: Philosophical foundations*. Oxford University Press.

Chalmers, D. J. (2023). Does thought require sensory grounding? From pure thinkers to large language models. *Proceedings and Addresses of the APA, 97*, 22–45.

Borg, E. (2026). LLMs, Turing tests and Chinese rooms: The prospects for meaning in large language models. *Inquiry, 69*(6), 2807–2837.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.

Gibson, J. J. (1979). The theory of affordances. In *The ecological approach to visual perception* (chap. 8).

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*(20), 771–791.

Gomes, A. (2015). Testimony and other minds. *Erkenntnis, 80*(1), 173–183.

Gomes, A. (2019). Perception, evidence, and our expressive knowledge of others' minds. In A. Avramides & M. Parrott (Eds.), *Knowing other minds* (pp. 148–172). Oxford University Press.

Parrott, M. (2026). Expression, testimony, and other minds. *Philosophical Studies, 183*(1), 287–305.

Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.

Simonelli, R. (2023). How to be a hyper-inferentialist. *Synthese, 202*(5). https://doi.org/10.1007/s11229-023-04382-1

Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences* (forthcoming).

Simonelli, R. (2026). Sapience without sentience: An inferentialist approach to LLMs. *Asian Journal of Philosophy, 5*(1). https://doi.org/10.1007/s44204-026-00400-4