

범주기반 속성추론: 인과관계 강도의 검증 Category-Based Feature Inference: Testing Causal Strength

조준형¹ · 이형철² · 김신우^{3†}

JunHyoung Jo¹ · Hyung-Chul O. Li² · ShinWoo Kim^{3†}

Abstract

This research investigated category-based feature inference when category features were connected in common cause and common effect causal networks. Previous studies that tested feature inference in causal categories showed unique inference patterns depending on causal direction, number of related features, whether the to-be-inferred feature was cause or effect, etc. However, these prior studies primarily focused on inference patterns that arise from causal relations, and few studies directly explored how the effects of causal relations vary depending on causal strength. We tested feature inference in common cause (Expt. 1) and common effect (Expt. 2) causal categories when causal strengths were either strong or weak. To this end, we had participants learn causal categories where features were causally linked and then perform feature inference task. The results showed that causal strengths as well as causal relations had important impacts on feature inference. When causal strength was strong, inference for common cause feature became weaker but that for the common effect feature became stronger. Moreover, when causal strength was strong and common cause was present, inference for the effect features became stronger, whereas the results were reversed in common effect networks. In particular, in common effect networks, causal discounting was more evident with strong causal strength. These results consistently demonstrate that participants consider not only causal relations but also causal strength in feature inference of causal categories.

Key words: Category-Based Feature Inference, Common Cause, Common Effect, Causal Strength

요약

본 연구는 범주속성들이 공통원인 혹은 공통효과 인과 네트워크로 연결되었을 때 인과강도에 따른 속성추론을 검증했다. 인과범주에서 속성추론을 검증한 기존 연구들은 인과관계의 방향, 연결된 속성의 개수, 원인 혹은 결과의 여부 등에 따라 고유한 추론 패턴이 나타남을 보여주었다. 다만 기존 연구들은 인과관계에 따른 추론편을 주로 탐색했으며 인과관계의 효과가 인과강도에 따라 어떤 변화를 보이는지 확인한 연구는 찾아보기 어렵다. 본 연구에서는 공통원인(실험 1), 공통효과(실험 2) 네트워크에서 인과강도에 따른 속성추론을 검증했다. 이를 위해 참가자들에게 속성들이 인과적 관련성을 가지는 범주를 학습하게 한 다음 속성추론 과제를 실시하도록 했다. 실험 결과 인과관계 뿐만 아니라 인과강도 역시 속성추론에 중요한 영향을 미쳤다. 인과강도가 강할 때 공통원인 속성에 대해서는 추론이 약해진 반면 공통효과 속성에 대해서는 추론이 강해졌다. 또한 인과강도가 강할 때 공통원인이 존재하는 경우 결과속성들에 대한 추론이 강해진 반면 공통효과에서는 반대의 결과가 나타났다. 특히 공통효과에서는 인과강도가 강할 때 인과적 절감이 더 뚜렷하게 나타났다. 이 결과들은 인과적 범주에서의 속성추론에서 참가자들은 인과관계 뿐만 아니라 인과강도를 고려한다는 것을 일관성있게 보여준다.

주제어: 범주기반 속성추론, 공통원인, 공통효과, 인과강도

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구자지원사업의 지원을 받아 수행된 연구임(NRF-2020S1A5A8047804).

¹ 조준형: 광운대학교 산업심리학과 박사과정

² 이형철: 광운대학교 산업심리학과 교수

^{3†} (교신저자) 김신우: 광운대학교 산업심리학과 교수 / E-mail: shinwoo.kim@kw.ac.kr / TEL: 02-940-5421

1. 서론

개념과 범주표상은 고차원적인 인간사고의 기반이 된다. 복잡한 사고와 추론을 수행하기 위해서는 범주적 지식과 개념표상은 필요조건이다. 가령 “모든 인간은 죽는다”라는 전제가 “소크라테스도 죽는다”라는 결론으로 이어지기 위해서는 “소크라테스도 인간이다”라는 범주 위계 혹은 소속성에 대한 범주적 지식을 필요로 한다. 과태료와 범칙금이라는 두 개념의 차이를 아는 것은 불법주차나 속도위반 고지서를 어떤 방식으로 납부할지 판단하는데 중요한 역할을 한다. 새로운 호흡기 바이러스가 출현했을 때 그것이 겨울철에 더 기승을 부릴 것이라 예측하는 것 또한 바이러스의 전형적인 특성에 근거한 범주기반 일반화(category-based generalization)의 결과이다. 환경의 복잡성을 줄이고, 대상의 정체를 확인하는 인지적 도구가 되며, 의사소통의 매개역할을 하고, 특정한 상황에서 (혹은 대상에 대해) 어떻게 행동할지를 알려주며, 지식을 체계화하는 기본단위라는 다양한 측면에서 범주와 개념은 인간 사고에서 중요한 지위를 차지한다(Bruner et al., 1956).

범주는 특히 새로운 대상의 속성에 대한 추론의 근거가 되며 이로 인해 인간 지식의 가용성을 증폭된다. 처음 보는 생물(예를 들어, 펭귄)이 조류라는 것을 알게 되면 그것이 알을 낳을 것이며 부리와 날개를 가질 것이라고 추측할 수 있다. 박쥐가 포유류임을 아는 것은 그것이 알이 아니라 새끼를 낳고 모유를 먹일 것이라는 특성을 추론하는 기반이 된다. 펭귄과 박쥐가 각 상위범주(조류, 포유류)의 전형적인 종은 아니라는 것은 그것들이 여러 측면에서 다양한 특이성을 가질 것이라는 것을 추측케 한다.

범주기반 추론은 다양한 형태로 발생할 수 있다. 가장 기본적인 것은 상위범주의 특성이 하위범주 혹은 단일 개체로 전이되는 방식으로 삼단논법 연역법의 기본이 된다. 모든 사람은 죽으며 소크라테스도 사람이라는 것은 그도 죽음을 피할 수 없다는 결론을 내리게 한다. 반면 일반적으로 범주기반 추론은 귀납사고의 형태에 대해 연구되어 왔다(Osherson et al., 1990; Rips, 1975). 가령 모든 참새가 호르몬 X를 가진다고 할 때 다른 새들도 호르몬 X를 가지는지 혹은 박새도 호르몬 X를 가지는지를 추론하는 것은 비확정적인 귀납추론을 요구

한다. 이때 참새의 특성은 다른 새들의 특성을 추론하는데 도움이 되지만 반대로 펭귄 혹은 타조의 특성은 다른 새들로 전이되기 어렵다. 이는 참새가 다양한 새들에 대한 대표성을 가지기 때문인데 이를 전형성 효과(typicality effect)라 부른다. 참새와 박새가 호르몬 X를 가진다고 해서 모든 새가 호르몬 X를 가진다는 확신을 가지기는 어려울 것이다. 반면, 참새와 펭귄이 호르몬 X를 가진다면 다른 모든 새들도 호르몬 X를 가진다는 결론에 좀 더 높은 확신을 가질 수 있을 것이다. 이는 {참새, 박새}보다 {참새, 펭귄}의 조합이 더 포괄적이기 때문으로 이를 다양성 효과(diversity effect)라 부른다. 만약 {참새, 타조, 펭귄}이 모두 호르몬 X를 가진다면 모든 다른 새들도 호르몬 X를 가질 것이라는 결론에 더욱 강한 확신을 가질 수 있을 것이다.

본 연구에서는 여러 유형의 범주기반 속성추론 중 특정 범주에 속하는 개체들의 속성에 대한 추론을 검증하고자 한다. 예를 들어, 고등학생 범주에서 서로 다른 학생들의 학업열의를 추론한다고 가정해보자. 학생 A가 아침일찍 등교를 하고, 머리가 단정하며, 교복을 입는 등의 특성을 보인다면 아마도 평균이상의 학업열의를 가질 것이라 추론할 수 있다. 반면, 학생 B가 아침 10시에 기상을 하고, 머리를 핑크색으로 염색했으며, 찢어진 청바지를 입는 모습을 보인다면 그 학생의 학업열의를 의심하게 될 것이다. 이는 부분적으로 A가 고등학생 범주에서 높은 전형성을 보이는 반면 B는 비전형적이기 때문이다. 즉 사람들은 특정 개체의 속성을 추론할 때 그것이 속하는 범주의 전형성에 근거해 추론하는 경향을 보인다. 어떤 새로운 새를 보고 그것이 날수 있을지를 추론할 때도 그 새가 참새나 박새와 유사한 전형적인 모습을 보일 때와 달리 펭귄이나 타조와 같이 비전형적인 외양을 보이는 경우 날지 못할 것이라 판단할 가능성이 높다.

범주 전형성은 특정 범주에 속하는 개체들에 대한 속성추론을 돕는 강력한 휴리스틱 즉 직관적 추론수단이다. 전형성이 속성추론의 근거가 되는 이유는 명확하지 않지만 범주의 원형이론(Rosch & Mervis, 1975)에 근거한다면 범주내 개체들이 보이는 가족 유사성(family resemblance)때문인 것으로 추측해 볼 수 있다. 다만 이러한 Roschian 관점은 범주를 규정하는 것이 서로 다른 범주 간 속성의 상관관계를 의미하는 것이기 때문에

범주내 속성들이 실제로 서로 통계적 상관을 가지는지는 명확하지 않다(Murphy, 2002; Rehder & Burnett, 2005). 이보다는 아마도 속성추론에서의 전형성 효과는 전형적인 개체가 해당 범주의 일반적인 특성을 가질 것이라는 암묵적 가정에 근거한 것으로 해석하는 편이 더 타당할 것으로 보인다. 즉 어떤 범주의 전형적인 개체는 그 범주의 다른 전형적인 속성들도 가질 것이라는 것이다. 이는 특히 범주 응집성(category coherence)이 높은 본질화된 범주(essentialized category)의 특성이기도 하다(Gelman, 1988; Lee et al., 2020; Lassaline & Murphy, 1996; Neufeld, 2022; Rehder & Hastie, 2004).

그런데 앞선 예시들(고등학생의 학업열의, 새의 비행 가능성)에서 범주 전형성 뿐만 아니라 범주와 그 속성에 대한 다양한 사전 지식이 추론에 개입할 수 있다는 것을 쉽게 알 수 있다. 즉 추론자는 대상의 전형성 뿐만 아니라 추론 대상이 되는 속성 및 여러 관련 속성들에 대한 지식을 속성추론에 활용하게 된다. 가령 고등학생의 경우 아침일찍 등교하고, 머리가 단정하며, 교복을 입는다는 것은 학업열의라는 변인의 결과이기 때문에 역방향 인과추론을 통해 학업열의를 가늠할 수 있다. 또한 목소리, 키, 몸무게와 같은 변인들은 학업열의와 별로 관련이 없기 때문에 추론에 활용하지 않을 가능성이 높다. 펭귄이나 타조의 경우 날개의 크기에 비해 몸이 무겁기 때문에 충분한 양력을 얻지 못해 날지 못할 것이라는 추론을 할 수 있다. 즉 어떤 대상의 속성추론에서는 전형성 뿐만 아니라 속성들 간의 관련성과 이들에 대한 지식이 추론에 영향을 미치게 된다(Doh & Lee, 2020; Kim et al., 2008; Medin et al., 2003; Proffitt et al., 2000; Rehder, 2009).

본 연구에서는 범주 기반 속성추론에서 범주내 속성들이 다양한 인과적 관련성을 가질 때 인과강도(즉, $Pr(\text{Effect}|\text{Cause})$)에 따른 추론패턴을 검증하고자 한다. 기존 연구들은 범주내 속성들의 인과관계는 속성추론에 강한 영향을 미치며 인과 네트워크의 유형에 따라 추론의 패턴이 크게 달라짐을 보고했다(Choi et al., 2021; Kim & Li, 2017; Rehder & Burnett, 2005; Rehder & Hastie, 2004). 다만 기존 연구들에서는 인과네트워크의 각 유형이 추론에 미치는 고유한 효과를 탐색했을 뿐 이러한 효과가 인과강도에 따라 어떻게 달라지는지는 확인하지 않았다. 범주관련 다양한 판단 과제

(특히, 범주화, 전형성 판단 등)에서 속성간 인과관계의 강도가 중요한 영향을 미친다는 것은 여러차례 보고되었으나(Rehder, 2003; Rehder & Kim, 2009), 범주 기반 속성추론 과제에서 인과 강도에 따른 추론패턴을 검증한 연구는 찾아보기 어렵다. 본 연구에서는 범주내 속성들이 공통원인(common cause) 혹은 공통효과(common effect) 네트워크의 형태로 관련성을 가질 때(Fig. 1) 인과강도에 따른 속성추론 패턴을 검증할 것이다.

범주 기반 속성추론에서 인과관계의 효과를 면밀히 검토하기 위해 인과모형(causal model)을 활용하는 것은 매우 유용하다(Sloman, 2005). Fig. 1은 범주속성과 이들간의 인과관계를 노드와 링크로 표현한 예시를 보여준다. 먼저, Fig. 1A는 속성 F1이 F2, F3, F4의 공통원인(common cause)인 경우를 보여주며, Fig. 1B는 속성 F1이 F2, F3, F4의 공통효과(common effect)임을 보여준다. 범주 A, B에서 속성간 네트워크의 형태는 동일하지만 인과관계의 방향으로 인해 고유한 속성추론 패턴이 나타난다. 예를 들어, 범주 A에서 속성 F4를 추론할 때 원인속성인 F1의 존재여부(즉, 속성값)를 확인했다면 F2, F3의 속성값을 알 필요도 없고 알더라도 인과추론과 무관하다. 왜냐하면 오로지 F1만 F4와 인과적 관련성을 가지며 F2, F3는 인과적 연결관계가 부재하기 때문이다. F1의 존재여부를 모르는 경우라면 F2, F3를 통해 F1을 추론하고 이에 따라 F4에 대해 판단하게 된다. 다만 F1의 값이 확정적인 경우 F2, F3는 추론과 무관해진다. 다시말해 F1은 F2, F3의 영향을 차단(screen-off)한다(Hausman & Woodward, 1999). 이는 공통원인의 원리(principle of the common cause)로 불리기도 하고(Reichenbach, 1956) 인과적 마코프 조건(causal Markov condition)이라 칭하기도 한다(Pearl, 2000).

그런데 기존 연구들은 Fig. 1A와 같은 공통원인 네트

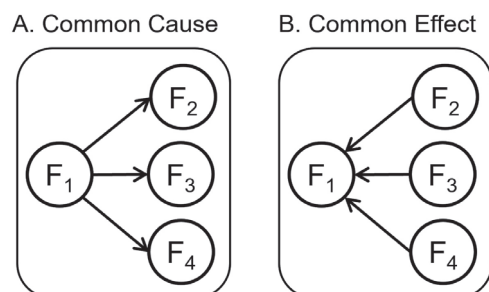


Fig. 1. Common cause (A) and common effect (B) causal networks

워크에서 결과속성들의 개수가 늘어남에 따라 다른 결과속성들의 존재에 대한 확신이 올라간다는 결과를 보고했다(Kim & Li, 2017; Rehder & Burnett, 2005). 즉, F4에 대한 추론에서 F2, F3의 속성이 0, 1, 2개로 늘어날수록 F4에 대한 확신이 높아지는 결과가 나타났다. 본 연구에서는 이러한 인과적 마코프 조건에 대한 위배가 속성들 간의 인과강도에 따라 어떻게 달라지는지를 검증하고 한다(e.g., O'Neill et al., 2022).

다음으로 범주B(Fig. 1B)의 공통효과에서 결과속성인 F1이 존재하는 경우 원인속성들(F2, F3, F4)에 대한 추론에서 인과적 절감(causal discounting)이 발생한다. 예를 들어, F4에 대한 추론에서 F1이 존재하고 F2, F3이 부재한다면 추론자는 F4가 존재할 것이라는 강한 확신을 가지게 된다. 왜냐하면 F1의 존재를 설명할 수 있는 유일한 원인이 F4이기 때문이다. 그런데 속성 F2, F3의 개수가 0, 1, 2개로 늘어날수록 F4에 대한 확신이 오히려 줄어들게 된다. 이는 어떤 인과적 결과를 설명하는 다른 원인들이 이미 존재하는 경우 추가적인 원인에 대한 확신이 줄어드는 인과적 절감때문이다(Laux et al., 2010; Morris & Larrick, 1995; Spellman et al., 2001).

기존 연구들(Kim & Li, 2017; Rehder & Burnett, 2005)은 공통효과 네트워크에서 인과적 절감이 발생함을 보고한 바 있다. 본 연구에서는 이러한 인과적 절감이 인과강도에 따라 어떻게 달라지는지 검증할 것이다. 직관적으로 볼 때 인과강도가 강할 때 인과적 절감은 더 뚜렷해질 것으로 추측해 볼 수 있다.

2. 실험 1: 공통원인 네트워크

범주내 속성들이 공통원인(Fig. 1A)의 형태로 인과관계를 가질 때 인과강도에 따른 속성추론 패턴을 검증했다. 이를 위해 가상의 범주와 범주속성 및 인과관계를 구성하였으며, 참가자들은 서로 다른 속성들의 존재여부에 대해 판단하는 과제를 수행했다.

2.1. 방법

2.1.1. 실험자극

실험을 위해 빅토리아 새우(생물), 로고스 자동차(인공물), 탄산 나트륨(자연 무생물)의 서로 다른 세 가지 범주

를 사용했다(Rehder & Kim, 2006; Rehder & Kim, 2009). 각 범주를 다른 종류의 것들로 구성함으로써 실험 결과의 외적 타당도를 높이고자 했다. 각 범주들은 두 개의 속성값을 가질 수 있는 네 개의 속성(차원)으로 구성했다. 두 속성값 중 하나는 전형적인 값으로 참가자들에게 ‘많은(many)’으로 제시했다. 다른 하나의 값은 비전형적이었으며 ‘일부(some)’로 표현했다. Table 1은 빅토리아 새우를 예시로 속성 차원과 속성값을 보여준다. 전형적인 빅토리아 새우는 많은 양의 ACh 신경전달물질을 가지며, 회피반응이 길고, 수면주기가 가속화되며, 무게가 무겁다.

범주 속성들은 공통원인 네트워크의 인과관계를 가졌다(Fig. 1A). 예를 들어, 빅토리아 새우의 경우 공통원인에서는 ‘ACh 신경전달물질’이 다른 세 가지 전형적인 속성들의 원인이었으며($F1 \rightarrow F2$, $F1 \rightarrow F3$, $F1 \rightarrow F3$), 공통효과에서는 ACh 신경전달물질이 다른 세 가지 전형적 속성들의 인과적 결과였다($F2 \rightarrow F1$, $F3 \rightarrow F1$, $F4 \rightarrow F1$). 속성간 인과관계는 원인 \rightarrow 결과를 한 문장으로 서술하고 그 이유를 한 두 문장으로 추가 서술하는 방식으로 제시했다(e.g., Stephan & Waldmann, 2022). 이때 원인 \rightarrow 결과 문장에서 ‘거의 항상(almost always)’ 혹은 ‘때때로(sometimes)’라는 부사를 사용해 인과강도를 조작했다. 예를 들어, ‘많은 양의 ACh 신경전달물질은 (거의 항상/때때로) 도피반응을 길게 만든다’의 방식으로 인과강도를 조작했다.

2.1.2. 참가자 및 실험설계

총 36명의 학부생이 실험에 참여했으며 세 종류의 범주(새우, 자동차, 나트륨)와 두 가지 인과강도 조건(강함, 약함)에 동일한 인원을 무선적으로 할당했다. 각 참가자는 하나의 범주와 하나의 인과강도 조건에서만 실험을 수행했다. 따라서 실험은 3(범주유형: 새우, 자동차, 나트륨) \times 2(인과강도: 강함, 약함)의 참가자간 요인설계로 구성되었다. 본 실험은 IRB의 승인을 획득한 후 실시했다.

Table 1. Features and their values (Victoria Shrimp)

Feature dimension	Typical value (many)	Atypical value (some)
ACh neurotransmitter	High quantity	Low quantity
Flight response	Long-lasting	Short
Sleep cycle	Accelerated	Decelerated
Body weight	High	Low

2.1.3. 절차

참가자들은 범주학습과 속성추론의 순서로 실험을 수행했다. 범주학습에서는 컴퓨터 화면에 제시된 범주 지식을 학습했다. 첫 화면에는 범주명칭과 해당 범주를 소개하는 한두 문장을 제시했다. 두 번째 화면에는 해당 범주의 속성차원과 각 차원의 전형적, 비전형적 속성값과 기저율을 제시했으며, 세 번째 화면에는 속성간 인과관계를 문장으로 제시했다. 마지막 화면에는 Fig. 1과 같은 형태로 속성간 인과관계를 요약하여 제시했다. 참가자들이 학습이 완료되었음을 실험자에게 알리면 서술형 시험을 수행하도록 했다. 각 범주의 속성값, 기저율, 인과관계, 인과강도를 묻는 질문에 모두 정확하게 답한 경우 속성추론을 실시하도록 했다.

속성추론에서 참가자들은 제시된 범주예시에서 “????”로 표시된 속성차원의 값을 추론하는 과제를 수행했다. 화면의 최상단 중앙에 범주명을 제시하고 그 아래에 F1, F2, F3, F4를 한줄에 하나씩 줄을 바꾸어 제시했다. 총 4개의 속성차원 각각에 대한 추론에서 나머지 3개 속성들이 두가지 값을 가질수 있으므로 참가자들은 총 32회(4 × 2 × 2 × 2)의 속성추론 시행을 실시했다. 예를 들어, 첫 번째 속성차원에 대한 추론은 {x000, x001, x010, x100, x011, x101, x110, x111}의 총 8회의 시행으로 구성되었다. 화면의 하단에는 “????”에 대한 추론을 요청하는 질문과 함께 추론 평정을 위한 가로형태의 평정막대를 제시했다. 평정막대의 정중앙에 세로선을 제시했으며 참가자들은 좌우 방향 키로 세로선을 움직여 적절한 위치로 이동한 다음 엔

터키를 눌러 반응했다. 참가자들의 반응은 [0 100]의 점수로 저장되었다.

2.2. 결과 및 논의

참가자간 요인인 범주유형(새우, 자동차, 나트륨)에 따른 평정값의 유의미한 차이가 나타나지 않아 결과를 통합해 분석했다. 즉 Fig. 2A와 2B의 평정값에 대한 인과강도, 속성의 개수, 범주유형을 요인으로 하는 삼원 분산분석에서 범주의 효과는 모두 유의미하지 않았다, all $p_s > .5$. Fig. 2는 범주속성들이 공통원인 네트워크(Fig. 1A)로 연결되었을 때의 속성추론 결과이다. 먼저 공통원인인 F1에 대한 추론결과(Fig. 2A)를 보면 결과속성의 개수가 늘어날수록 F1에 대한 추론이 강해짐을 알 수 있다. 이는 부분적으로는 전형성 효과로 해석할 수 있지만 인과강도에 따라 결과패턴이 크게 달랐다. 특히 속성의 개수가 1, 2개인 경우 인과강도가 강할 때 F1의 존재에 대한 확신이 현저하게 약했다. 이는 원인속성 F1이 존재하고 인과강도가 강하다면 결과속성들(F2, F3, F4)이 일부만 존재할 가능성이 낮다는 인과추론을 반영하는 것으로 해석할 수 있다.

통계적 검증을 위해 F1에 대한 추론평정값에 대해 2(인과강도: 약함, 약함) × 4(속성의 개수: 0, 1, 2, 3) 참가자간 분산분석을 실시했다. 그 결과 속성개수의 주효과는 유의미했으며, $F(3, 102) = 197.90, p < .001, \eta_p^2 = .85$, 인과강도의 주효과 역시 유의미했다, $F(1, 34) = 11.61, p < .01, \eta_p^2 = .26$. 특히 속성개수와 인과강도

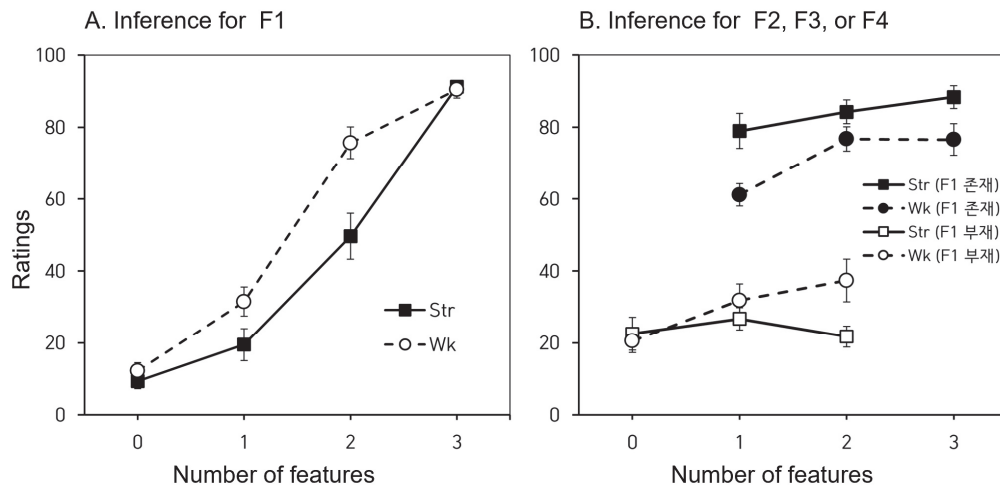


Fig. 2. Experiment 1: Inference for F1 (A) and F2, F3, or F4 (B)

의 상호작용이 유의미했는데, $F(3, 102) = 5.37, p < .01, \eta_p^2 = .14$, 이는 참가자들이 인과강도에 따라 속성추론을 달리했음을 보여준다.

Fig. 2B는 F1의 존재여부에 따른 결과속성들(F2, F3, F4)에 대한 추론결과를 보여준다. 원인속성 F1의 존재할 때(우측 위)와 부재할 때(좌측 아래) 결과속성에 대한 추론 평정값은 큰 차이를 보였다. 특히 이러한 차이(즉, F1의 존재여부에 따른 차이)는 인과강도가 약할 때 보다 강할 때 더 크게 나타났다, $t(34) = 3.37, p = .002$.

통계적 검증을 위해 먼저 F1이 부재할 때의 평정값에 대한 2(인과강도: 약함, 강함) \times 3(속성의 개수: 0, 1, 2) 참가자간 분산분석을 실시했다. 그 결과 속성개수에 따른 주효과는 유의미한 수준에 미치지 못했으며, $F(2, 68) = 3.0, p = .056, \eta_p^2 = .08$, 인과강도의 주효과는 또한 유의미하지 않았다, $F(1, 34) = 2.29, p = .140, \eta_p^2 = .06$. 속성개수와 인과강도의 상호작용 또한 유의미하지 않았다, $F(2, 68) = 2.86, p = .064, \eta_p^2 = .08$. 이 결과들은 원인속성 F1이 부재하여 인과추론이 가능하지 않을 때는 인과강도에 따른 조건간 차이가 발생하지 않음을 보여준다.

다음으로, F1이 존재할 때의 평정값에 대한 2(인과강도: 약함, 강함) \times 3(속성의 개수: 1, 2, 3) 참가자간 분산분석을 실시했다. 그 결과 속성의 개수에 따른 주효과는 유의미했으며, $F(2, 68) = 6.16, p = .003, \eta_p^2 = .15$, 인과강도의 주효과 역시 유의미했다, $F(1, 34) = 16.70, p < .001, \eta_p^2 = .33$. 반면, 속성개수와 인과강도의 상호작용은 유의미하지 않았다, $F(2, 68) = .93, p = .401, \eta_p^2 = .03$. 이 결과들은 원인속성인 F1이 존재하는 경우 인과강도가 약할 때 보다 강할 때 결과속성들에 대한 추론이 강해짐을 보여준다.

결과들을 종합하면 참가자들은 추론과정에서 속성들의 인과패턴뿐만 아니라 인과강도를 고려한다는 것을 보여준다. 인과강도가 강할 때 결과속성들(F2, F3, F4)이 부분적으로만 존재하는 경우 공통원인 F1의 존재가능성을 낮게 평가하는 것으로 나타났다 (Fig. 2A). 또한 공통원인인 F1이 존재하는 경우 인과강도가 강할 때 결과속성들의 존재에 대한 확신이 더욱 강하다는 것을 알 수 있다(Fig. 2B). 공통원인 F1이 부재하여 속성간 인과적 관련성이 없는 경우에는 인과강도에 따른 차이가 나타나지 않았다(Fig. 2B).

3. 실험 2: 공통효과 네트워크

범주내 속성들이 공통원인(Fig. 1B)의 형태로 인과관계를 가질 때 인과강도에 따른 속성추론 패턴을 검증했다. 이를 위해 가상의 범주와 범주속성 및 인과관계를 구성하였으며, 참가자들은 서로 다른 속성들의 존재여부에 대해 판단하는 과제를 수행했다.

3.1. 방법

실험자극, 참가자, 실험설계, 절차는 실험 1과 동일했다. 다만, 각 범주속성들은 공통원인이 아닌 공통효과 네트워크의 인과관계를 가졌으며(Fig. 1B), 참가자들은 이에 상응하는 인과관계적 지식을 습득했다(F2 \rightarrow F1, F3 \rightarrow F1, F4 \rightarrow F1). 본 실험은 IRB의 승인 후 실시했다.

3.2. 결과 및 논의

참가자간 요인인 범주유형(새우, 자동차, 나트륨)에 따른 평정값의 유의미한 차이가 나타나지 않아 결과를 통합해 분석했다. 즉 Fig. 3A와 3B의 결과에 대한 인과강도, 속성의 개수, 범주유형을 요인으로 하는 삼원 분산분석에서 범주의 효과는 모두 유의미하지 않았다, all $ps > .05$. Fig. 3는 범주속성들이 공통효과 네트워크(Fig. 1B)로 연결되었을 때의 속성추론 결과이다. 먼저 공통효과인 F1에 대한 추론결과(Fig. 3A)를 보면 원인속성의 개수가 늘어날수록 결과속성 F1에 대한 추론이 강해짐을 알 수 있다. 특히 인과강도가 약한 조건과 비교하여 강한 조건에서 원인속성이 0에서 1로 단 한 개 늘어날 때 결과속성 F1에 대한 평정값의 변화가 더 컸으며, $t(34) = 2.76, p = .009$, 이는 참가자들이 인과강도를 추론에 반영했음을 보여주는 직관적인 증거가 된다.

통계적 검증을 위해 결과속성 F1에 대한 평정값에 대해 2(인과강도: 약함, 약함) \times 4(속성의 개수: 0, 1, 2, 3) 참가자간 분산분석을 실시했다. 그 결과 속성개수의 주효과는 유의미했으며, $F(3, 102) = 150.02, p < .001, \eta_p^2 = .82$, 인과강도의 주효과는 유의미하지 않았다, $F(1, 34) = .80, p = .377, \eta_p^2 = .02$. 다만, 속성개수와 인과강도의 상호작용이 유의미했는데, $F(3, 102) =$

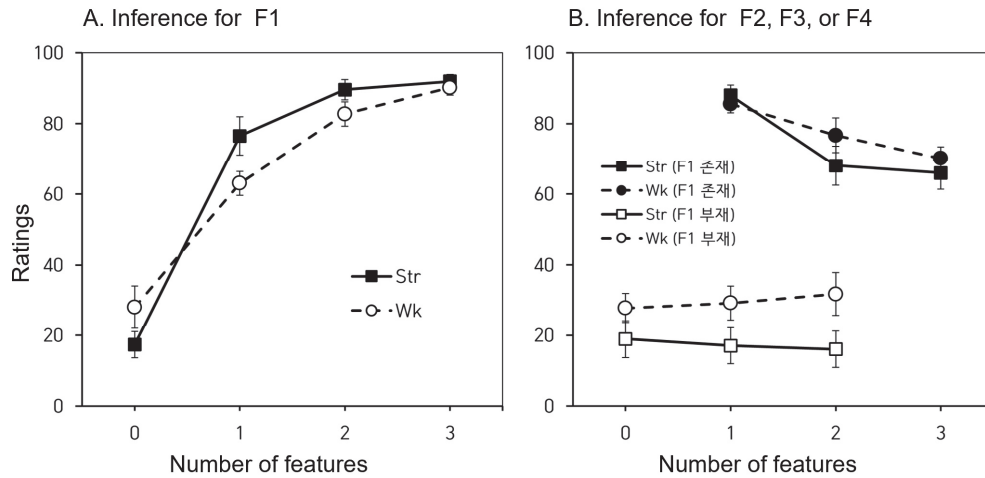


Fig. 3. Experiment 2: Inference for F1 (A) and F2, F3, or F4 (B)

4.04, $p < .01$, $\eta_p^2 = .11$, 이는 인과강도에 따라 참가자들의 추론패턴이 달랐음을 보여준다.

Fig. 3B는 F1의 존재여부에 따른 원인속성들(F2, F3, F4)에 대한 추론결과를 보여준다. 결과속성 F1이 부재할 때(좌측 아래)와 존재할 때(우측 위) 원인속성에 대한 추론 평정값은 큰 차이를 보였다. 특히 결과속성 F1이 부재하는 경우 인과강도가 강할 때 참가자들은 원인속성의 존재가능성을 낮게 평가했다. 결과속성 F1이 존재하는 경우 인과적 절감(causal discounting)은 인과강도가 강할 때와 약할 때 모두 나타났다.

통계적 검증을 위해 먼저 결과속성 F1이 부재할 때의 평정값에 대한 2(인과강도: 약함, 강함) \times 3(속성의 개수: 0, 1, 2) 참가자간 분산분석을 실시했다. 그 결과 속성개수에 따른 주효과는 유의미하지 않았으며, $F(2, 68) = .01$, $p = .989$, $\eta_p^2 < .01$, 속성개수와 인과강도의 상호작용 역시 유의미하지 않았다, $F(2, 68) = .21$, $p = .813$, $\eta_p^2 = .01$. 다만, 인과강도의 주효과는 유의미했다, $F(1, 34) = 10.35$, $p = .003$, $\eta_p^2 = .23$. 이 결과는 인과강도가 강하고 결과속성 F1이 부재할 때 참가자들은 원인속성들(F2, F3, F4)도 존재하지 않을 것이라 판단했음을 보여준다.

다음으로, F1이 존재할 때의 평정값에 대한 2(인과강도: 약함, 강함) \times 3(속성의 개수: 1, 2, 3) 참가자간 분산분석을 실시했다. 그 결과, 속성개수의 주효과는 유의미했으며, $F(2, 68) = 11.67$, $p < .001$, $\eta_p^2 = .26$, 이는 인과적 절감(causal discounting)이 발생했음을 보여준다. 즉, 결과속성 F1이 존재할 때 이를 설명하는 원인속

성이 최소 1개는 존재해야하기 때문에 속성개수가 1개인 경우 (즉, F1만 존재하는 경우) 평정값은 매우 높았다. 반면 원인속성이 이미 존재하는 경우 추가적인 설명을 제공하는 원인속성은 불필요해지므로 추론강도가 점차 약해지는 현상이 발생한다. 다만, 인과강도의 주효과는 유의미하지 않았으며, $F(1, 34) = 1.00$, $p = .324$, $\eta_p^2 = .03$, 속성개수와 인과강도의 상호작용 역시 유의미하지 않았다, $F(2, 68) = .94$, $p = .397$, $\eta_p^2 = .03$.

공통효과에서의 이 결과들은 참가자들이 범주속성들의 인과패턴뿐만 아니라 인과강도를 추론에 반영한다는 것을 보여준다. Fig. 3A에서 확인할 수 있듯이 인과강도가 강할 때는 (약할 때와 비교하여) 원인속성이 0에서 1로 증가할 때 결과속성 F1에 대한 확신이 17.3에서 76.5점으로 크게 증가했던 반면 추가적인 원인속성들이 추론에 미치는 영향은 비교적 작았다. 반면 인과강도가 약할 때는 이러한 경향성이 완화되었는데, 이 때문에 인과강도와 속성개수의 상호작용이 유의미했던 것으로 보인다. Fig. 3B에서는 인과추론의 전형적인 패턴인 인과적 절감이 나타났다. 또한 결과속성 F1이 부재한 경우 인과강도가 강할 때 원인속성에 대한 존재를 더욱 부정적으로 평가하는 것을 확인할 수 있었다.

4. 종합논의

본 연구는 범주내 속성들이 다양한 인과적 관련성을 가질 때 인과강도에 따른 범주기반 속성추론을 검증했

다. 기존 연구들에서는 범주속성들이 특정한 인과관계로 연결되었을 때 인과관계의 방향, 연결된 속성의 개수, 원인속성인지 결과속성인지의 여부 등에 따라 고유한 추론패턴이 나타난다는 것을 보여주었다(Choi et al., 2021; Kim & Li, 2017; Rehder & Burnett, 2005; Rehder & Hastie, 2004). 다만 기존 연구들은 주로 인과관계에 따른 추론패턴을 탐색했기 때문에 인과관계의 효과가 인과강도에 따라 어떤 변화를 보이는지는 확인한 연구는 찾아보기 어렵다. 범주속성들이 공통원인 혹은 공통효과 네트워크로 연결되었을 때의 인과강도에 따른 추론패턴을 검증하기 위한 실험을 실시했다. 이를 위해 참가자들에게 속성들이 인과적으로 연결된 범주들을 학습하게 한 다음 속성추론 과제를 실시하도록 했다. 실험 1과 2 각각에서 범주속성들은 공통원인 혹은 공통효과의 인과관계를 가졌으며 참가자들은 할당된 조건에서 속성간 인과관계를 습득했다. 속성간 인과관계를 서술하면서 “거의 항상(almost always)” 혹은 “때때로(sometimes)”라는 부사를 통해 인과강도를 조작했다.

실험 1과 2에서 모두 인과적 범주에서 관찰되는 전형적인 추론패턴이 나타났으며 특히 인과강도에 따라 평정값이 달라졌다. 실험 1(공통원인)에서는 원인속성인 F1에 대한 추론이 인과강도에 따라 통계적으로 유의미한 차이가 나타났다. 인과강도가 강한 경우 결과속성이 일부 존재하더라도 그 원인이 되는 F1에 대한 추론이 비교적 약하게 발생했다(Fig. 2A). 이러한 결과는 범주에 대한 심리적 본질주의(psychological essentialism)와 맞닿아 있다. 심리적 본질주의에 따르면 사람들은 특정 범주에 속하는 것들은 잘 드러나지 않는 어떤 본질속성(essential feature)을 공유하며 이러한 본질속성이 다른 관찰가능한 속성들의 인과적 원인이 된다는 믿음을 가진다(Gelman, 2003; Medin & Ortony, 1989). 이러한 믿음은 서로 다른 범주간 지각적 경계를 강화하는 결과를 낳게 되는데 이는 기존 연구들(Hirschfeld, 1996; Kim et al., 2021; Rehder & Kim, 2009)에서 다양한 방법론을 통해 밝혀진 사실이다. 심리적 본질주의에 따른 범주 경계가 강화되는 현상과 본 연구에서 원인속성에 대한 추론패턴은 모두 공통원인 네트워크에서 발생한다는 특성을 공유한다. 즉 공통원인 네트워크에서 사람들은 특정 범주를 다른 것들과 구별되는 것으로 인식하며 범주 구성원들이 동질적일 것이라는 기대를 가진다고 해석할 수 있다.

본 연구의 또 다른 흥미로운 결과는 Fig. 2A와 3A에

서 F1에 대한 추론패턴이 반대였으며 인과강도가 강할 때 더 뚜렷한 대비가 나타났다는 점이다. 이는 참가자들이 추론과정에서 인과관계의 방향과 인과강도를 모두 고려했음을 보여준다. 즉 반대되는 추론 패턴은 인과관계의 방향을 고려했음을 보여주며 대비의 크기는 인과강도를 반영했음을 나타내는 증거가 된다. 인과추론에서 순행추론과 역행추론이 다르다는 것은 잘 알려진 사실이다(Waldmann, 2001). 본 결과는 참가자들이 범주속성간 단순 연합이 아닌 방향성을 고려한 추론을 실시했다는 것을 보여준다.

Fig. 2B와 3B에서 F1이 존재하는 경우에서의 추론은 또 다른 대비를 보여준다. 인과강도가 약할 때 보다 강할 때, 공통원인에서는 추론강도가 강해졌고 공통효과에는 추론강도가 약해졌음을 알 수 있다. 이는 앞서 언급한 바와 마찬가지로 참가자들이 인과강도와 인과관계의 방향을 모두 고려해 속성추론을 실시했음을 보여주는 증거가 된다.

Fig. 2B와 3B에서 F1이 부재한 경우의 추론은 비교적 유사한 패턴을 나타냈다. 두 경우 모두 인과강도가 강할 때 속성추론이 약화되었다. 이는 인과강도가 강함에도 불구하고 원인 혹은 결과로써의 F1이 부재하다면 인과적으로 연결된 다른 속성들도 존재하지 않을 것이라는 추론을 반영하는 것으로 보인다. 반면 인과강도가 약할 때는 세부적으로는 약간의 추론패턴 차이가 발생했다. 공통원인(Fig. 2B)에서는 다른 결과속성의 개수가 늘어날수록 속성추론이 강해지는 결과가 나타난 반면 공통효과(Fig. 3B)에서는 패턴이 나타나지 않았다. 이 차이는 아마도 인과관계의 방향에 따른 것으로 보인다. 즉, 공통원인에서는 원인속성 F1이 없기 때문에 인과관계를 배제한 상태에서 추론평정을 실시할 수 있고 이는 일반적인 전형성 효과로 이어지게 된다. 반면 공통효과에서는 결과속성 F1이 없다는 것은 이에 대한 원인속성들도 없다는 것을 의미하므로 다른 원인속성의 개수와 상관없이 속성추론의 강도가 일관되게 낮게 나온 것으로 추측할 수 있다.

본 연구는 범주 속성들이 공통원인 혹은 공통효과 인과네트워크로 연결되었을 때 인과강도에 따른 추론패턴을 검증했다. 인과적 범주에서의 속성추론에 대한 기존 연구들은 대부분 인과패턴의 효과를 검증하는데 집중되어 왔다. 본 연구에서는 인과패턴에 더하여 인과강도가 추론에 미치는 영향을 검증했다. 실험 결과 인과패턴 뿐만 아니라 인과강도도 속성추론에 중요한 영

향을 미친다는 것이 나타났다. 인과강도가 강할 때 공통원인 F1에 대해서는 추론이 약해진 반면 공통효과 F1에 대해서는 추론이 강해졌다. 또한 인과강도가 강하고 F1이 존재할 때 공통원인에서는 F1의 결과속성들에 대한 추론이 강해진 반면 공통효과에서는 F1의 원인속성들에 대한 추론은 약해졌다. 반대로 인과강도가 강하고 F1이 부재할 때는 공통원인에서 F1의 결과속성들에 대한 추론이 약해진 반면 공통효과에서는 F1의 원인속성들에 대한 추론은 강해졌다. 또한 인과강도가 강할 때 공통효과에서 인과적 절감이 더 뚜렷하게 나타났다. 이 결과들은 참가자들이 인과적 범주의 속성추론에서 인과관계뿐만 아니라 인과강도를 반영한다는 것을 명확하고 일관성있게 보여준다.

연구결과에 따른 여러 함의점에도 불구하고 본 연구는 몇 가지 한계점이 있다. 먼저 범주속성의 기저율이 속성추론 과정에 명확하게 반영되지 않았다는 점은 참가자들이 추론과정에서 인과관계에만 과도하게 집중했으며 전형성에 따른 기저율은 간과했음을 보여준다. 또한 실험 조건에서 인과적 사슬구조가 존재하지 않아 인과강도에 따른 마코프 조건에 대한 검증을 하지 못했다. 마지막으로 인과강도가 약하거나 강한 두 조건을 비교했을 뿐 인과적 연결이 존재하지 않는 조건이 포함되지 않아 중립적인 비교를 하지 못했다는 한계를 가진다. 이러한 취약점들을 보완한 후속 연구가 필요할 것이다.

REFERENCES

- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. John Wiley and Sons.
- Choi, I, Li, H-C. O., & Kim, S. (2021). Category-based feature inference in causal chain (인과적 사슬구조에서의 범주기반 속성추론). *Science of Emotion & Sensibility*, 24(1), 59-72. DOI : 10.14695/KJSOS.2021.24.1.59
- Doh, E. Y., & Lee, G.-H. (2020). Effect of interaction between category coherence and base rate on presumption of reasons for preference (범주 응집성과 기저율의 상호작용이 선호의 이유 추정에 미치는 효과). *Korean Journal of Cognitive Science*, 31(3), 77-102. DOI: 10.19066/cogsci.2020.31.3.002
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20(1), 65-95. DOI: 10.1016/0010-0285(88)90025-4
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780195154061.001.0001
- Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the Causal Markov Condition. *British Journal for the Philosophy of Science*, 50(4), 521-583. DOI: 10.1093/bjps/50.4.521
- Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. The MIT Press.
- Kim, N. S., Yopchick, J. E., & de Kwaadsteniet, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review*, 15(1), 81-88. DOI: 10.3758/pbr.15.1.81
- Kim, S., Jo, J.-H., & Li, H.-C. (2021). Psychological essentialism and category representation (심리적 본질주의와 범주표상). *Korean Journal of Cognitive Science*, 32(2), 55-73. DOI: 10.19066/cogsci.2021.32. 2.001
- Kim, S., & Li, H.-C. O. (2017). Modeling feature inference in causal categories (인과적 범주의 속성추론 모델링). *Korean Journal of Cognitive Science*, 28(4), 329-347. DOI: 10.19066/cogsci.2017.28.4.007
- Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review*, 3(1), 95-99. DOI: 10.3758/BF03210747
- Laux, J. P., Goedert, K. M., & Markman, A. B. (2010). Causal discounting in the presence of a stronger cue is due to bias. *Psychonomic Bulletin & Review*, 17(2), 213-218. DOI: 10.3758/PBR.17.2.213
- Lee, G.-H, Li, H.-C., & Kim, S. (2020). Effect of interaction between the base-rate and category coherence on property generalization (범주 응집성과 기저율의 상호작용이 속성 일반화에 미치는 효과). *Korean Journal of Cognitive and Biological Psychology*, 32(1), 1-19. DOI: 10.22172/cogbio.2020. 32.1.001
- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, 10(3), 517-532.

- DOI: 10.3758/BF03196515
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge University Press. DOI: 10.1017/CBO9780511529863.009
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, 102(2), 331-355. DOI: 10.1037/0033-295X.102.2.331
- Murphy, G. L. (2002). *The big book of concepts*. MIT Press.
- Neufeld, E. (2022). Psychological essentialism and the structure of concepts. *Philosophy Compass*, 17(5), e12823. DOI: 10.1111/phc3.12823
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97(2), 185-200. DOI: 10.1037/0033-295X.97.2.185
- O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, 223, 105036. DOI: 10.1016/j.cognition.2022.105036
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 811-828. DOI: 10.1037/0278-7393.26.4.811
- Rehder, B. (2003). A Causal-Model Theory of Conceptual Representation and Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1141-1159. DOI: 10.1037/0278-7393.29.6.1141
- Rehder, B. (2009). Causal-based property generalization. *Cognitive Science*, 33(3), 301-344. DOI: 10.1111/j.1551-6709.2009.01015.x
- Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, 50(3), 264-314. DOI: 10.1016/j.cogpsych.2004.09.002
- Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, 91(2), 113-153. DOI: 10.1016/S0010-0277(03)00167-7
- Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 659-683. DOI: 10.1037/0278-7393.32.4.659
- Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition*, 37(6), 715-729. DOI: 10.3758/MC.37.6.715
- Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press. DOI: 10.2307/2216858
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, 14(6), 665-681. DOI: 10.1016/S0022-5371(75)80055-7
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573-605. DOI: 10.1016/0010-0285(75)90024-9
- Sloman, S. A. (2005). *Causal models: How we think about the world and its alternatives*. New York: Oxford University Press. DOI: DOI:10.1093/acprof:oso/9780195183115.001.0001
- Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory & Cognition*, 29(2), 193-208. DOI: 10.3758/BF03194913
- Stephan, S., & Waldmann, M. R. (2022). The role of mechanism knowledge in singular causation judgments. *Cognition*, 218, 104924. DOI: 10.1016/j.cognition.2021.104924
- Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, 8(3), 600-608. DOI: 10.3758/BF03196196
- 원고접수: 2022.07.22
수정접수: 2022.09.09
게재확정: 2022.09.19.