인지과학, 제28권 제4호
Korean Journal of Cognitive Science
2017, Vol. 28, No. 4, 329~347.
http://dx.doi.org/10.19066/cogsci.2017.28.4.007

인과적 범주의 속성추론 모델링*

김 신 우† 이 형 철
광운대학교 산업심리학과

범주기반 속성추론에 대한 초기연구들은 전형성, 다양성, 유사성 효과 등 인간 사고에서 나타나는 다양한 현상들을 보고하였다. 이후 연구들은 이러한 추론에서 참가자들의 사전지식이 광범위한 영향을 미친다는 것을 발견하였다. 본 연구에서는 다양한 사전지식들 중 하나인 인과적 지식이 속성추론에 미치는 영향을 검증하고 이를 모델링하였다. 이를 위해 참가자들은 네 개의 속성으로 구성된 범주에서 속성들이 공통원인 혹은 공통효과 인과구조로 연결되었을 때 속성추론과제를 실시하였다. 그 결과 전형성 효과와 더불어 공통원인 구조에서 인과적 마코프 조건(causal Markov condition)에 대한 위배와 공통효과 구조에서 인과적 절감(causal discounting)이 관찰되었다. 이를 모델링하기 위해 참가자들은 표적속성이 존재하는 범주예시와 존재하지 않은 범주예시가 존재할 가능성에 대한 차이값 (즉, $p(E_{F(X)}|Cat)$-$p(E_{F(\sim X)}|Cat)$에 근거하여 속성추론을 수행한다고 가정하였다. 인과모형이론(Rehder, 2003)에 기반하여 범주예시들의 확률값을 계산한 후 각 표적속성에 대한 추론에 적용하였다. 그 결과 모형은 참가자들의 데이터에서 관찰된 전형성 효과뿐만 아니라 인과적 마코프 조건에 대한 위배 및 인과적 절감을 모두 예측한다는 것이 확인되었다.

주제어 : 범주기반 속성추론, 인과추론, 인과모형이론, 전형성

---

* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A5A8017325).
† 교신저자: 김신우, 광운대학교 산업심리학과, (01897) 서울시 노원구 광운로 20
연구분야: 인지심리학
Tel: 02-940-5421, Fax: 02-941-9214, E-mail: shinwoo.kim@kw.ac.kr

귀납추론은 과거의 경험과 지식을 바탕으로 새로운 상황, 대상, 혹은 범주에 대해 잠재적인 판단을 내리는 것을 의미한다. 예를 들어, 독일산 전자제품을 사용했을 때 내구성이 높았던 경험에 비추어 독일산 제품들 전체에 대해 내구성이 높을 것이라는 결론을 내리기도 하고, 어떤 제품이 독일산이라는 것을 알면 그 제품의 내구성이 높을 것이라고 추측하기도 한다. 전자는 과거 사례들을 근거로 독일산 제품이라는 범주에 대한 추론인 반면 후자는 개별 독일산 제품의 속성에 대한 추론이다. 본 연구에서는 후자의 경우, 즉 범주에 대한 기존 지식을 근거로 그 범주에 속하는 특정 대상의 속성에 대한 추론에 대해 검증 및 모델링 하고자 한다.

기존 연구들은 대상의 속성추론에 영향을 미치는 다양한 효과들을 보고하였다. 대표적으로 전형성 효과(typicality effect)는 대상의 범주 전형성이 높을수록 새로운 대상에 대한 속성추론이 강해지는 것을 의미한다. 예를 들어 전형적인 새인 참새나 비둘기가 호르몬 X를 가지고 있다면 부엉이도 호르몬 X를 가지고 있을 것이라고 추측할 개연성이 높지만 비전형적인 펭귄이 호르몬 X를 가지고 있다고 해서 부엉이도 호르몬 X를 가지고 있을 것이라고 짐작하기는 어려울 것이다. 즉, 추론의 기반이 되는 대상의 전형성이 높을수록 속성이 같은 범주의 다른 대상으로 전이될 가능성은 높아진다. 전형성 효과와 더불어 다양성 효과(diversity effect)라는 현상도 발견되었다. 이는 둘 이상의 기존 범주들의 유사성이 낮을수록 (즉, 다양할수록) 새로운 대상에 대한 속성추론이 강해지는 것을 의미한다. 예를 들어, 서로 유사한 참새, 박새, 울새보다 서로 상이한 참새, 타조, 펭귄이 공통적으로 소유한 특징 X를 새로운 어떤 새도 동일하게 소유할 것이라고 생각하는 경향성이 강하다는 것이다.

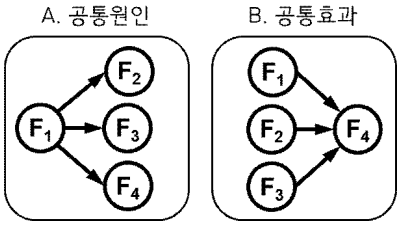
이러한 전형성 효과, 다양성 효과 등은 범주기반 속성추론의 고전인 유사성-범위 모형(similarity-coverage model; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990)과 속성기반 귀납모형(feature-based induction model; Sloman, 1993)의 근간이 되었다. 두 모형들은 다소간의 차이점은 있지만 모두 범주들 간의 유사성을 설명의 핵심적인 요소로 사용한다는 점에서 동일한 접근을 취한다고 볼 수 있다.

범주기반 속성추론의 초기 연구들은 범주 및 범주예시들 간의 유사성 혹은 위계적 관계에 근거하여 이론을 전개했기 때문에 표적속성을 속성 X 혹은 가상의 해부학적 기관명, 뼈 이름 등으로 제시하여 사전지식이 개입할 수 있는 가능성을 차단하였다. 그러나 이후 연구들은 사전지식이 추론에 광범위한 영향을 미친다는 것을 보여주었다. 예를 들어, Heit과 Rubinstein(1994)은 속성과 범주가 어떤 관계를 가지는가에 따라 귀납추론의 강도가 달라진다는 것을 보고하였다. 예를 들어, 행동특성의 경우에는 참치로부터 고래에게 전이되었지만 곰으로부터 고래에게는 일반화되지 않았다. 반면, 해부학적 특성의 경우에는 반대의 현상이 발견되었다. 이 결과는 행동특성은 서식지에 의해 결정되지만 (심해: 참치, 고래) 해부학적인 특성은 분류학적 유사성 (포유류: 곰, 고래)에 근거하기 때문일 것이다. 이 외에도 범주와 속성에 대한 지식이 귀납추론에 미치는 영향을 보여준 연구는 매우 많은데 (Sloman, 1997; Springer & Keil, 1989), 어떤 경우에는 사

전지식이 유사성 효과를 완전히 제거하는 경우도 보고되었다 (Bailenson, Shum, Atran, Medin, & Coley, 2002; Lopez, Atran, Coley, Medin, & Smith, 1997; Proffitt, Coley, & Medin, 2000; Shafto & Coley, 2003). 특히 Rehder (2006, 2009)는 범주속성들이 인과관계로 연결되어 있는 경우의 귀납적 일반화는 유사성에 의해 거의 설명하지 못한다는 것을 보고하였다.

기존 연구들은 범주학습, 표상, 사용에서 인과적 지식이 중요한 영향을 미친다는 것을 지속적으로 보고하였다. 예를 들어, 범주속성간 인과적 관련성은 범주표상의 기반이 되고 (Ahn & Medin, 1992; Kim, Luhmann, Pierce, & Ryan, 2009; Medin, Wattenmaker, & Hampson, 1987), 범주화에도 영향을 미치며(Ahn, 1998, Ahn, Kim, Lassaline, &Denis, 2000; Rehder, 2003; Sloman, Love, & Ahn, 1998), 새로운 속성에 대한 일반화의 근거가 되기도 한다(Hajichristidis, Sloman, Stevenson, & Over, 2004; Kim, Yopchick, & de Kwaadsteniet, 2008; Medin, Coley, Storms, & Hayes, 2003; Rehder & Hastie, 2004).

본 연구는 범주내 속성들의 인과관계가 속성추론에 미치는 영향을 실험을 통해 확인하고 이를 설명하는 모형을 제안한다. 어떤 새를 보았을 때 그것이 '날 수 있을지'를 추론한다든가 새로운 버섯을 보았을 때 그것이 '먹기에 안전한지'를 판단할 때 추론자는 그 대상의 관찰 가능한 속성과 범주에 대한 지식에 근거하여 판단할 것이다. 어떤 버섯이 화려한 색을 가지고 있는 경우 화려한 색을 가진 버섯은 독을 가지고 있다는 지식에 근거하여 그것이 위험하다는 판단을 할 수 있을 것이다. 만약 추론자가 버섯의 어떤 유전자가 화려한 색, 주름진 모양, 그리고 독성의 공통원인이라는 인과적 지식구조를 가지고 있다면 버섯의 화려한 색 뿐만 아니라 주름을 살펴보고 독성유무를 판단할 것이다. 인과모형(causal model)은 이러한 종류의 추론을 검토하기 위한 효율적인 도구가 될 수 있다. 인과모형에서 각 매듭(node)은 속성을 나타내고 인과관계는 화살표로 표시한다.



(그림 1) 네 개의 속성(F1, F2, F3, F4)으로 이루어진 인과적 범주의 예

그림 1은 F1, F2, F3, F4의 네 가지 속성으로 구성된 범주와 범주속성들 간의 인과관계를 보여준다. 그림 1A에서 F1은 원인이고 F2, F3, F4는 F1의 결과이며, 그림 1B에서 F4는 결과이고 F1, F2, F3는 원인이다. 본 연구에서는 그림 1A의 공통원인(common cause) 구조와 1B의 공통효과(common effect)의 구조에서의 속성추론을 검증하고 모델링하고자 한다. 본 연구와 유사한 인과

구조를 사용하여 속성추론 혹은 일반화를 검증한 기존의 여러 연구들이 존재한다(Rehder, 2006; Rehder, 2009; Rehder & Burnett, 2005; Rehder & Hastie, 2004). 이 연구들이 활용한 범주들과 인과관계적 지식은 유사한 것들이었지만 구체적인 가설과 검증한 내용에서는 뚜렷한 차이가 있었다. 예를 들어 Rehder와 Hastie(2004)는 인과관계적 지식에 의한 범주 응집성(coherence)이 추론에 미치는 영향을 확인하였으며, Rehder(2006)는 인과관계와 유사성이 모두 존재할 때 추론은 하나의 방식으로만 이루어진다는 것을 보여준 연구였다. Rehder와 Burnett(2005)은 다양한 인과적 범주에서의 속성추론을 검증하였으나 이에 대한 계산적 모델링은 시도하지 않았다. Rehder(2009)는 인과구조에 따른 속성의 범주 전체에 대한 일반화를 검증한 반면, 본 연구에서는 속성이 (전체 범주가 아닌) 특정 범주예시에 존재하는지 아닌지에 대한 추론을 검증하고 이에 대한 모델링 결과를 보고할 것이다.

실 험

네 개의 속성이 공통원인 혹은 공통효과 구조(그림 1)를 구성하는 범주에서의 속성추론을 검증하였다. 통제조건의 참가자들은 속성간 인과관계가 없는 범주에서 속성추론을 실시하였다.

방 법

실험자극 및 설계

실험을 위해 구성한 생물 (키호개미), 자연물 (미야 별), 인공물 (넵튠 컴퓨터)의 세 가지 범주를 사용하였다 (Rehder & Kim, 2006; Rehder & Kim, 2009). 각 범주들은 네 개의 속성차원을 가지며 각 차원은 두 가지 속성값을 가지는데, 해당 범주의 전형적인 값이며 다른 하나는 비전형적인 값이었다. 전형적인 속성값의 기저율은 '대부분'으로 제시하였고 비전형적인 속성값은 '일부'로 제시하였다. 전형적인 속성은 F로 표시하고 비전형적인 속성은 ~F로 표시할 것이다. 따라서 F1은 첫 번째 차원의 속성이 전형적이라는 뜻이며 ~F2는 두 번째 차원의 속성이 비전형적이라는 것을 나타낸다.

표 1은 키호 개미를 예시로 제시한 속성차원과 속성값이다. 전형적인 키호 개미는 혈중 황산철 농도가 높고, 면역기능이 활성화 되었으며, 혈액이 진하고, 개미집을 빠른 속도로 건설한다. 각 범주들은 네 가지 차원의 속성뿐만 아니라 조건에 따라 속성간 인과관계를 가지고 있었다. 통제조건에서는 속성간 관련성이 없었지만, 공통원인의 인과구조에서는 높은 혈중 황산철 농도

〈표 1〉 범주차원, 속성, 및 속성값의 예 (키호 개미)

속성 차원	전형적인 값 (대부분)	비전형적인 값 (일부)
혈중 황산철 농도	높음	낮음
면역기능 활성화	활성화됨	억제됨
혈액농도	진함	묽음
개미집 건설 속도	빠름	느림

가 다른 세 가지 전형적인 속성들의 원인이 되는 반면 공통효과의 인과구조에서는 빠른 개미집 건설속도는 다른 세 가지 전형적인 속성들에 의한 결과가 된다.

각 참가자는 하나의 범주만 학습한 후 속성추론 과제를 실시하였다. 따라서 범주는 피험자간 요인이었다. 또한 속성들의 인과관계에 따라 공통원인 구조, 공통효과 구조, 통제조건을 피험자간 요인으로 구성하였다. 그 결과 실험은 3 (인과구조: 공통원인, 공통효과, 통제조건) x 3 (범주 유형: 개미, 별, 컴퓨터)의 피험자간 요인설계로 구성되었다.

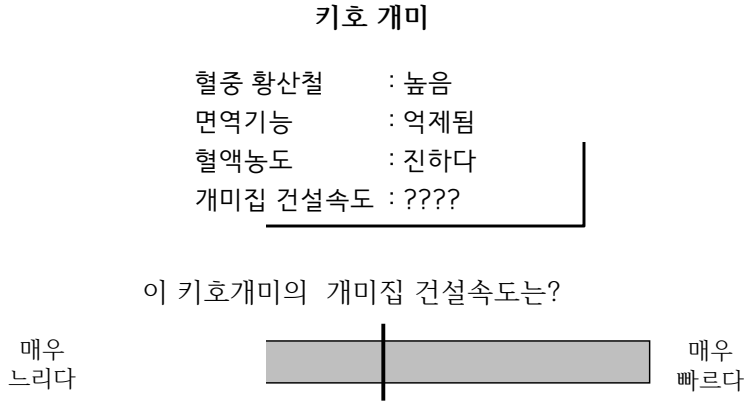
참가자

총 90명의 학부생들이 실험에 참여하였다. 각 조건에 동일한 인원을 무선적으로 할당하여, 각 범주 및 각 인과구조에 30명씩 배정하였다.

절차

참가자들은 해당 조건의 범주 관련 지식을 학습한 후 속성추론 과제를 수행하였다. 범주학습 단계에서 참가자들은 컴퓨터에 파워포인트 화면으로 제시된 범주지식을 앞뒤로 넘기며 자율적으로 학습하였다. 첫 화면에는 범주의 이름과 그 범주를 소개하는 간략한 두, 세 문장의 글을 제시하였다. 두 번째 화면에서는 범주의 속성차원과 전형적인 그리고 비전형적인 속성값과 이들의 기저율을 제시하였으며, 세 번째 화면에서는 범주 속성간의 인과관계를 서술하였다. 마지막 화면에서는 인과관계를 요약하여 그림 1과 유사한 도표를 제시하였다. 학습 후 참가자들은 서술형 시험을 수행하였는데, 이때 범주의 각 속성, 속성값, 기저율, 인과관계에 대해 정확하게 기술한 경우 속성추론을 수행하도록 하였다.

속성추론에서는 참가자들에게 범주예시를 하나씩 제시한 후 네 개의 속성값 중 제시하지 않은 하나의 속성값에 대해 추론하도록 지시하였다. 네 속성차원 중 하나의 값을 추론하는 과제에서 나머지 세 차원은 두 가지 값을 가질 수 있기 때문에 참가자들은 총 32개($4 \times 2 \times 2 \times 2 = 32$)의 대상에 대한 속성추론을 실시하였으며, 각 시행은 무선적으로 제시하였다. 가령, 1-3



(그림 2) 속성추론 시행의 예

번째 속성이 전형적일 때 4번째 속성값을 추론하는 시행의 범주예시는 111X로 나타낼 수 있다. 따라서 X110는 2-3번째 속성이 전형적이고 4번째 속성이 비전형적일 때 첫 번째 속성의 값에 대해 추론하는 시행이 된다.

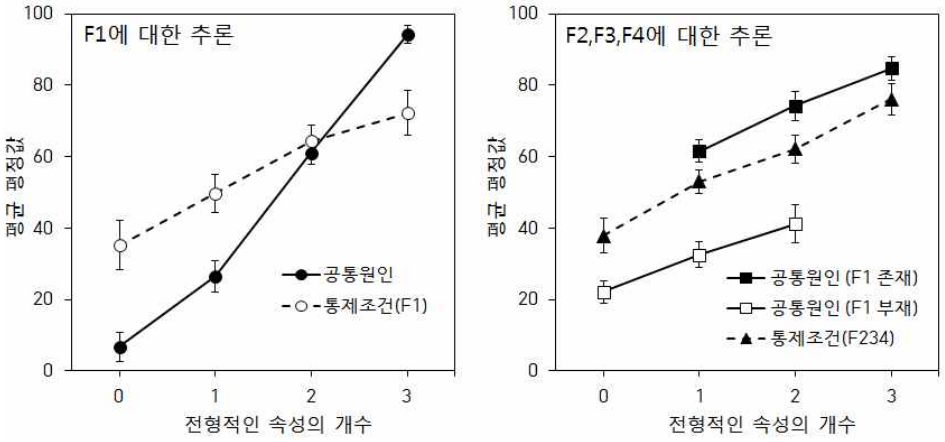
그림 2는 속성추론의 한 시행을 보여준다. 각 시행의 화면 상단에는 범주 이름을 제시하였고 범주 예시는 그 아래 네 줄의 텍스트로 제시하였다. 하나의 행에 속성명칭과 속성값을 제시하였으며 추론해야할 속성차원에서는 속성값을 "????"로 제시하였다. 참가자들은 좌우 방향키를 사용하여 해당 속성값의 존재 여부에 대해 세로막대를 움직인 후 엔터키를 눌러 반응하였다. 세로막대는 10점 단위로 이동하였으며 반응은 0-100의 범위로 저장되었다.

결과 및 논의

모든 참가자들은 범주지식 습득 후 실시한 서술형 시험을 통과하였다. 피험자간 요인인 범주유형(개미, 별, 컴퓨터)에 따른 속성추론 결과에 차이가 없었기 때문에 결과를 통합하였다.

공통원인: 결과 및 논의

그림 3은 공통원인에서의 속성추론 결과를 보여준다. 좌측은 공통원인이 되는 F1에 대한 추론결과와 이에 대응하는 통제조건의 결과이다. 통제조건의 결과는 일반적인 전형성 효과를 보여준다. 공통원인에서의 속성추론 결과는 속성간 인과관계가 강한 영향을 미쳤다는 것을 보여준다. F1에 대한 추론에서 통제조건과 유사하게 참가자들은 전형성 효과를 보였지만 속성의 개수가 증가함에 따라 가파르게 증가하는 평정값은 인과관계의 효과를 보여준다.



(그림 3) 공통원인 인과구조에서의 속성추론 결과

통계적 검증을 위해 좌측 F1 추론결과에 대한 2(인과관계: 공통원인, 통제조건) x 4(속성의 개수: 0, 1, 2, 3) 혼합변량분석을 실시하였다. 그 결과 인과관계의 주효과는 유의미 하지 않았으나, $F(1, 58) = 1.92$, $p = .172$, 속성 개수의 주효과는 유의미하여 전형성 효과를 보여주었다, $F(3, 174) = 85.39$, $p < .001$. 특히 인과관계와 속성개수간의 상호작용이 유의미하였는데, $F(3, 174) = 14.45$, $p < .001$, 이는 참가자들이 추론에서 인과관계를 활용하였음을 보여준다.

그림 3의 우측은 공통원인 F1의 존재여부에 따른 결과속성들 (F2, F3, F4)에 대한 추론결과와 이에 대응하는 통제조건의 결과를 보여준다. 먼저 통제조건에서는 좌측과 동일하게 일반적인 전형성 효과를 확인할 수 있었다. 그러나 공통원인에서는 공통원인인 F1의 존재여부에 따라 속성추론의 결과가 크게 달라졌다. 즉, 공통원인(F1)이 존재여부와 상관없이 속성의 개수가 증가할수록 평정값이 높아졌지만 (통제조건과 비교하여) F1이 존재할 때는 평정값이 높아졌고 F1이 없을 때는 평정값이 낮아졌다.

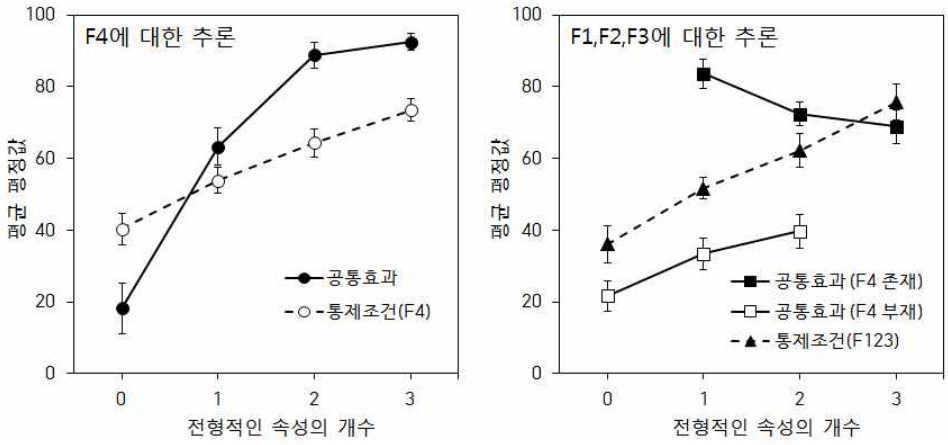
통계적 검증을 위해 F1이 존재하지 않을 때의 평정값에 대한 2(인과관계: 공통원인, 통제조건) x 3(속성의 개수: 0, 1, 2)의 혼합변량분석을 실시하였다. 그 결과 속성개수의 주효과는 유의미 하였으며, $F(2, 116) = 10.93$, $p < .001$, 인과관계의 주효과도 유의미하여, $F(1, 58) = 39.44$, $p < .001$, 공통원인인 F1이 존재하지 않을 때 그 결과가 되는 F2, F3, F4에 대한 추론이 약해진 것을 확인할 수 있었다. 두 요인간의 상호작용은 유의미하지 않았다, $F(2, 116) = .059$, $p = .943$. 공통원인인 F1이 존재할 때의 평정값에 대해서도 2(인과관계: 공통원인, 통제조건) x 3(속성의 개수: 1, 2, 3)의 혼합변량분석을 실시하였다. 그 결과 속성개수의 주효과는 유의미 하였으며, $F(2, 116) = 18.88$, $p < .001$, 인과관계의 주효과도 유의미하여, $F(1, 58) = 13.78$, $p < .001$, 공통원인인 F1이 존재할 때 그 결과속성들에 대한 추론이 강해진 것을 확인할 수 있었다. 두 요인들의 상호작용은 유의미하지 않았다, $F(2, 116) = .033$, $p = .968$.

이 결과들은 속성추론에서 인과관계적인 지식이 강한 영향력을 미친다는 것을 시사한다. 그림 3 좌측의 상호작용이 보여주는 것처럼 참가자들은 공통원인인 F1의 추론에서 결과속성들의 존재여부에 민감하게 반응하였다. 그림 3 우측의 결과는 공통원인인 F1 존재여부에 따른 결과속성에 대한 추론강도가 크게 달라지는 것을 보여준다. 이와 더불어 F1이 존재할 때와 존재하지 않을 때 모두 참가자들은 인과적 마코프 조건(causal Markov condition)을 위배하였다는 것을 알 수 있다(Hausman & Woodward, 1999; Pearl, 2000; Reichenbach, 1956). 인과적 마코프 조건에 따르면 공통원인인 F1의 존재여부를 확인할 수 있는 경우에는 결과속성들(F2, F3, F4)의 존재여부가 다른 결과속성의 추론에 영향을 미치지 않아야 한다. 따라서 우측의 결과에서 공통원인 조건에서는 (F1이 존재할 때와 존재하지 않을 때 모두) 평정값은 속성의 개수와 독립적으로 평평하게 나타나야 한다. 이 결과는 속성추론이 온전히 규범적 인과추론에 의해서만 결정되는 것이 아니라 전형성에 의해 영향을 받는다는 것을 보여준다.

공통효과: 결과 및 논의

그림 4는 공통효과에서의 속성추론 결과를 보여준다. 좌측은 공통효과가 되는 F4에 대한 추론결과와 이에 대응하는 통제조건의 결과이다. 공통효과의 결과는 속성추론에서 인과관계가 강한 영향을 미쳤다는 것을 보여준다. F4에 대한 추론에서 전형적인 속성의 개수가 증가할수록 평정값은 높아졌지만 평정값의 증가량은 오히려 줄었다는 것을 알 수 있다. 이는 F1, F2, F3가 모두 F4의 원인이기 때문에 하나 이상만 존재한다면 F4의 존재를 충분히 설명할 수 있기 때문일 것이다.

통계적 검증을 위해 좌측 F4 추론결과에 대한 2(인과관계: 공통효과, 통제조건) x 4(속성의 개



(그림 4) 공통효과 인과구조에서의 속성추론 결과

수: 0, 1, 2, 3) 혼합변량분석을 실시하였다. 그 결과 인과관계의 주효과는 유의미하였으며, $F(1, 58) = 5.11$, $p < .05$, 이는 공통효과에서의 더 높은 평정값에 따른 결과이다. 속성개수의 주효과는 유의미하여 전형성 효과를 보여주었다, $F(3, 174) = 82.82$, $p < .001$. 특히 인과관계와 속성개수간의 상호작용이 유의미하였는데, $F(3, 174) = 11.21$, $p < .001$, 이는 추론에서 참가자들이 인과관계를 활용한 결과로 해석할 수 있다.

그림 4의 우측은 공통효과 F4의 존재여부에 따른 원인속성들 (F1, F2, F3)에 대한 추론결과와 이에 대응하는 통제조건의 결과를 보여준다. 특히 공통효과 구조에서 F4의 존재여부에 따라 속성추론의 결과가 크게 달라진 것을 알 수 있다. 즉, 공통효과(F4)가 존재하지 않을 때는 F4의 인과적 원인이 되는 속성들(F1, F2, F3)에 대한 추론평정값은 통제조건보다 낮았지만 전형성 효과가 나타난 것을 확인할 수 있다. 그런데 공통효과(F4)가 존재할 때는 다른 원인속성들의 개수가 증가할수록 평정값이 오히려 낮아지는 것을 확인할 수 있다. 이는 절감(discounting)효과로써, 인과추론에서 여러 원인들이 인과적 결과(여기서는, F4)를 이미 충분히 설명하는 경우 추가적인 이유나 원인에 대한 확신이 오히려 줄어드는 현상이다(Morris & Larrick, 1995; Spellman, Price, & Logan, 2001). 예를 들어, 높은 시험점수는 실력, 낮은 난이도, 부정행위 등 다양한 원인에 의해 발생할 수 있다. 이때, 부정행위가 있었다는 것을 안다면 난이도가 낮아 점수가 높게 나왔다고 생각하기는 어려울 것이다. 혹은, 자동차 사고는 브레이크 결함, 미끄러운 도로, 운전자 부주의 등 다양한 원인에 의해 발생할 수 있지만 만약 빙판길에서 사고가 났다면 브레이크 결함을 의심하기는 매우 어려울 것이다.

통계적 검증을 위해 F4가 부재할 때의 평정값에 대한 2(인과관계: 공통효과, 통제조건) x 3(속성의 개수: 0, 1, 2)의 혼합변량분석을 실시하였다. 그 결과 속성개수의 주효과는 유의미하였다, $F(2, 116) = 15.76$, $p < .001$. 인과관계의 주효과도 유의미하였으며, $F(1, 58) = 31.60$, $p < .001$, 이는 공통효과 조건에서 F4가 존재하지 않을 때 그 원인속성에 대한 추론이 약화된 결과로 해석할 수 있다. 두 요인간의 상호작용은 유의미하지 않았다, $F(2, 116) < 1.0$. 공통효과인 F4가 존재할 때의 평정값에 대해서도 2(인과관계: 공통원인, 통제조건) x 3(속성의 개수: 1, 2, 3)의 혼합변량분석을 실시하였다. 그 결과 속성개수의 주효과는 유의미하지 않았는데, $F(2, 116) < 1.0$, 이는 인관관계와 속성개수의 유의미한 상호작용에 의한 것으로 해석할 수 있다, $F(2, 116) = 11.75$, $p < .001$. 즉, 통제조건에서는 속성개수가 증가함에 따라 평정값이 증가하지만 공통효과에서는 절감에 의해 평정값이 오히려 감소하여 교차상호작용이 나타났다. 인과관계의 주효과도 유의미하였는데, $F(1, 58) = 10.99$, $p < .01$, 이는 공통효과에서 F4가 존재할 때 원인속성의 존재에 대한 전반적으로 높은 확신에 의한 것으로 해석할 수 있다.

이 결과들은 앞선 공통원인에 대한 결과에서와 마찬가지로 속성추론에서 인과관계적인 지식이 강한 영향력을 미친다는 것을 보여준다. 특히 그림 3과 4의 결과는 인과관계의 유형에 따라 추론의 질적인 특성도 크게 달라진다는 것을 보여준다. 즉 참가자들은 속성들의 연합(association)

뿐만 아니라 인과관계의 방향을 고려하여 추론을 수행한다는 것을 확인할 수 있다. 공통효과의 결과에서 주목할 것은 F4에 대한 추론(그림 4의 좌측)과 F4가 존재할 때 F1,2,3에 대한 추론(그림 4의 우측)에서 절감효과가 뚜렷하게 나타난다는 점이다. 이후 검증할 모형은 인과추론에서의 이러한 질적인 특성들을 설명할 수 있어야 할 것으로 보인다.

속성추론의 이론적 모델링

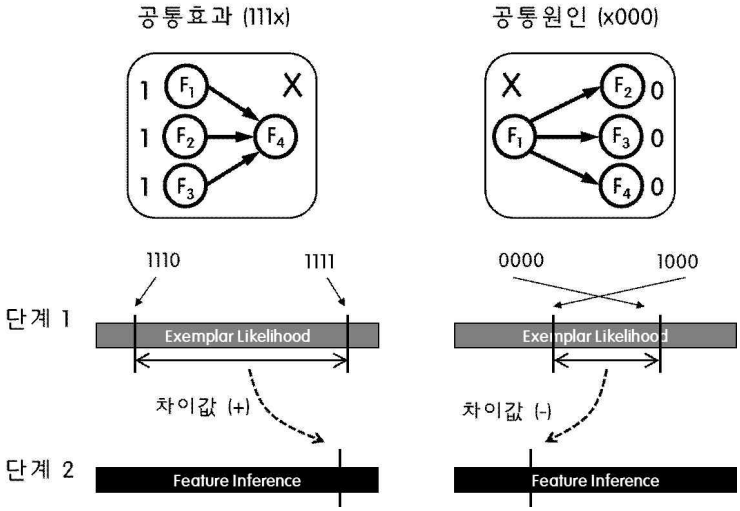
먼저 속성추론 모형의 작동방식과 계산절차에 대해 서술하고, 모델링 결과를 보고하고자 한다.

속성추론모형의 작동방식과 계산절차

속성추론모형의 작동방식

그림 5는 추론의 대상속성이 존재할 가능성이 높은 경우 (공통효과: 111X)와 존재할 가능성이 낮은 경우 (공통원인: X000) 각각에 대해 모형이 추론하는 과정을 도식적으로 보여준다. 먼저, 제안모형은 속성추론이 두 단계를 거쳐 이루어진다고 가정한다. 단계 1에서는 추론대상이 되는 속성이 존재하는 경우와 존재하지 않는 두 가지 각 경우의 범주예시가 존재할 가능성에 대한 확률계산이 발생한다. 그림 5의 좌측을 예로 들면, F4가 존재하는 경우의 확률은 p(1111)로 표현할 수 있고 존재하지 않는 경우는 p(1110)으로 표현할 수 있다. 단계 2에서는 두 확률간의 차이에 대한 계산이 이루어지고, 이에 근거하여 해당 속성이 존재할 가능성에 대한 판단이 발생한다고 가정한다. 그림 5의 좌측을 예로 들면 F4가 존재할 가능성에 대한 판단은 p(1111) - p(1110)에 근거한다. 실제로 111x에서는 모든 원인속성이 존재하므로 F4의 존재에 대한 확신이 높아야 한다. 그런데, 공통효과에서는 p(1111)과 p(1110)의 차이값이 매우 클 것이라는 것은 쉽게 예측할 수 있다.

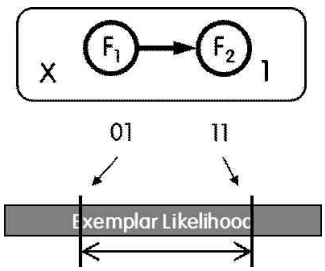
반면, 그림 5의 우측은 속성의 존재가능성이 매우 낮은 경우이다. 공통원인 구조에서 결과속성들이 전혀 존재하지 않기 때문에 F1에 대한 확신은 매우 낮아야 한다. 실제로 공통원인 구조에서는 p(1000) < p(0000)이며 따라서 차이값은 음수가 되어 F1에 대한 모형의 추론은 낮게 도출될 것이다.



(그림 5) 공통효과와 공통원인에서의 추론방식

속성추론모형의 계산절차

모형의 계산과정을 간결하게 기술하기 위해 두 개의 속성으로 이루어진 범주를 가정해보자 (그림 6). 이 범주에서 X1이 주어지면 앞서 기술한 바와 같이 속성추론은 다음에 근거하여 발생한다.



(그림 6) 두 개의 속성으로 이루어진 인과범주에서의 추론

$$\text{Rating}(F_1) = f[p(11) - p(01)] \quad (1)$$

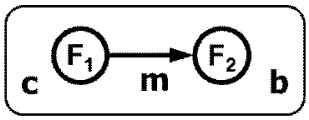
수식 (1)을 일반화하면 아래와 같은 식을 도출할 수 있다.

$$\text{Rating}(F_X) = f[p(E_{F(X)}|\text{Cat}) - p(E_{F(\sim X)}|\text{Cat})] \quad (2)$$

즉, 어떤 속성 F_X에 대한 추론은 그 속성이 존재하는 범주예시의 확률 $p(E_{F(X)}|Cat)$에서 존재하지 않는 범주예시의 확률 $p(E_{F(\sim X)}|Cat)$을 뺀 값이다. 실험결과와의 비교는 다음 식 (3)을 통해 이루어질 수 있다.

$$Rating(F_X) = K_1*[p(E_{F(X)}|Cat) - p(E_{F(\sim X)}|Cat)] + K_2 \quad (3)$$

(3)번 식에서 K_1과 K_2는 모형의 계산결과를 [0 100]의 반응척도로 변환하는 역할을 한다. 이때, 위의 식 (3)에서 $p(E_{F(X)}|Cat)$ 및 $p(E_{F(\sim X)}|Cat)$에 대한 계산은 인과적 범주화를 성공적으로 설명해온 인과모형이론(Causal Model Theory: CMT)에 근거하여 계산할 수 있다(Rehder, 2003; Rehder & Kim, 2006; Rehder & Kim, 2010).



(그림 7) 두 개의 속성으로 이루어진 인과범주에서의 계산과정

그림 7은 인과모형이론의 $p(E_{F(X)}|Cat)$ 혹은 $p(E_{F(\sim X)}|Cat)$의 계산을 설명하기 위해 제시한 단순한 범주이다. 이때 'c'는 속성 F1이 존재할 확률이며 'm'은 속성 F1이 존재할 때 F2가 발생할 확률이며, 'b'는 F1의 존재와 독립적으로 F2가 존재할 확률이다. 'c', 'm', 'b'를 이용하여 간단한 계산을 하면 그림 7의 범주에서 구성가능한 모든 예시들(00, 01, 10, 11)에 대한 확률계산식을 도출할 수 있다. 표 2은 각 예시들에 대한 계산공식을 보여준다. 표 2의 공식을 활용하면 최종적으로 0X, X0, X1, 1X의 추론에 대한 계산식을 구성할 수 있다. 표 3는 각 추론에 대한 계산식이다. 가령 0X를 예로 들면, p(01) - p(00)으로 계산할 수 있는데 표 2의 식을 공식 (3)을 적용하면 $K_1*[(1-c)*b-(1-c)*(1-b)] + K_2$ 이라는 계산식이 도출된다. 그림 7과 같이 두 개의 속성만으로 이루어진 인과범주의 계산만을 예시로 서술하였다. 그러나 속성의 개수가 늘어나거나 인과구조가 복잡해지더라도 간단한 확률계산을 통해 추론공식을 쉽게 도출할 수 있다.

〈표 2〉 그림 7의 범주예시들의 확률계산

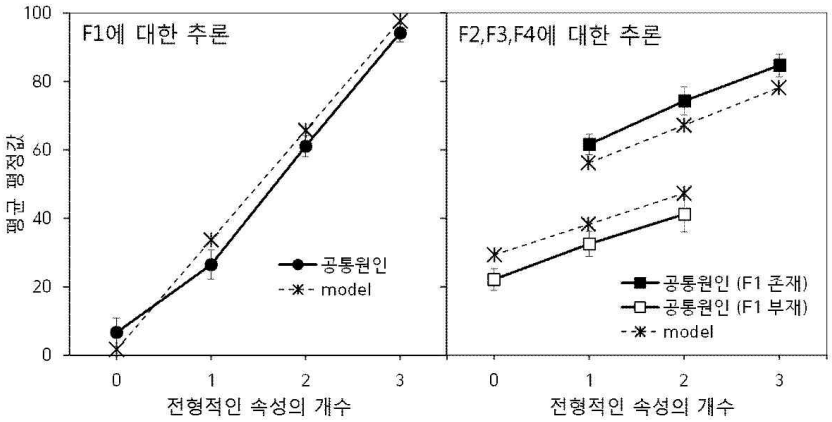
범주예시	L(E; c, m, b)
00	(1-c)*(1-b)
01	(1-c)*b
10	c*(1-m)*(1-b)
11	c*(m+b-m*b)

〈표 3〉 그림 7에서 구성 가능한 속성추론의 계산식

속성추론	계산식
0X	f[p(01)-p(00)] = K1*[(1-c)*b-(1-c)*(1-b)] + K2
X0	f[p(10)-p(00)] = K1*[c*(1-m)*(1-b)-(1-c)*(1-b)] + K2
X1	f[p(11)-p(01)] = K1*[c*(m+b-m*b)-(1-c)*b] + K2
1X	f[p(11)-p(10)] = K1*[c*(m+b-m*b)-c*(1-m)*(1-b)] + K2

속성추론모형의 타당성 검증

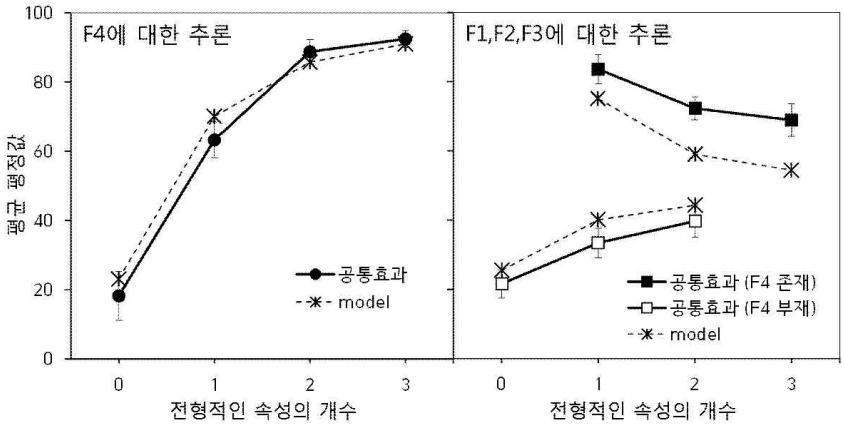
앞서 서술한 계산방식을 적용한 모형이 공통원인 및 공통효과 구조에서의 추론을 설명하는지를 검증하였다. 이를 위해 각 참가자의 평정값에 대해 편차제곱의 합을 최소화하는 모형의 예측값을 계산하였으며, 이때 평정값을 arcsine 함수로 변환하여 사용하였다. 그림 8은 공통원인의 모형검증 결과를 보여준다(avg. R^2 = .78). 좌측 공통원인이 되는 F1에 대한 추론결과에서 결과속성들의 개수가 늘어남에 따라 평정값이 증가하는 경향성이 참가자들의 반응과 모형의 예측에서 매우 유사하게 나타났다. 그림 8의 우측은 F1의 존재여부에 따른 결과속성들(F2, F3, F4)에 대한 추론결과와 모형의 예측을 보여준다. 모형은 F1이 존재할 때와 존재하지 않을 때 각각 과소추정 및 과대추정 하는 경향을 보여주었다. 하지만 전반적으로 볼 때 원인이 되는 F1의 존재여부와 상관없이 다른 속성들의 개수가 증가함에 따라 속성추론에 대한 확신이 선형적으로 증가하는 경향성을 모형이 유사하게 예측하였다. 이는 공통원인 구조에서 참가자들이 인과적 마코프 조건(causal Markov condition)을 위배하는 것을 모형이 그대로 예측한다는 것을 보여준다.



(그림 8) 공통원인 인과구조에서의 모형검증결과

즉, F1의 존재여부가 알려진 경우에는 다른 속성의 존재여부가 추론에 영향을 미치지 않아야 하기 때문에 (F1이 존재할 때와 존재하지 않을 때 모두) 규범적 모형은 평정값이 속성의 개수와 독립적으로 평평하게 나타나야 한다고 예측한다. 반면, 이 결과는 제안모형이 인간의 추론경향성을 유사하게 예측한다는 것을 보여준다.

그림 9는 공통효과의 모형검증 결과를 보여준다(avg. R^2 = .67). 좌측 공통효과인 F4에 대한 추론결과에서 원인속성들의 개수가 늘어남에 따라 평정값이 증가하는 경향성이 참가자들의 반응과 모형의 예측에서 매우 유사하게 나타났다. 그림 9의 우측은 F4의 존재여부에 따른 원인속성들(F1, F2, F3)에 대한 추론결과와 모형의 예측을 보여준다. 전반적인 경향성은 참가자의 평정과 모형의 예측이 유사하게 나타났으나 모형은 F4가 존재할 때와 존재하지 않을 때 각각 평정값을 과소추정 및 과대추정 하는 경향성을 보여주었다. 먼저 공통효과(F4)가 부재할 때 다른 속성의 개수가 증가함에 따라 속성추론에 대한 확신이 증가하는 경향성을 모형이 유사하게 예측하였다. 특히 공통효과(F4)가 존재할 때 다른 원인속성들의 개수가 증가할수록 속성평정값이 오히려 낮아지는 절감(discounting)을 모형이 유사하게 예측하였다. 즉, 다른 원인들이 F4를 이미 충분히 설명하는 경우 추가적인 이유나 원인에 대한 확신이 줄어드는 경향성을 보였으며, 이 결과는 제안모형이 인간의 추론경향성을 유사하게 예측한다는 것을 보여준다.



(그림 9) 공통효과 인과구조에서의 모형검증결과

결 론

본 논문에서는 속성들이 서로 인과적으로 연결되어 있는 범주에서의 속성추론결과를 검증하고 모형의 타당성을 탐색하였다. 이를 위해 인공적으로 구성한 세 개의 범주를 사용하였으며

각 범주들은 네 개의 속성차원으로 구성하였다. 조건에 따라 참가자들은 각 범주의 속성들과 인과관계를 학습하였는데 인과적 범주에서는 공통원인 혹은 공통효과 구조를 학습하였다. 학습이 완료되면 참가자들은 32개의 속성추론 시행을 실시하였다. 그 결과 모든 조건에서 전형성 효과가 발견되었으며 공통원인 구조에서는 인과적 마코프 조건에 대한 위배가 발생하였고 공통효과 구조에서는 절감이 발생한다는 것을 확인하였다. 검증을 위해 인과모형이론에 기반한 속성추론모형을 구성하여 모형이 참가자들의 반응패턴을 유사하게 예측하는지를 검증하였다. 그 결과 모든 조건에서 나타난 속성추론의 주요한 특징들을 모형이 그대로 재현한다는 것이 확인되었다.

제안한 모형은 몇 가지 핵심적인 요소를 제안한다. 먼저 어떤 속성의 존재여부를 추론할 때 사람들은 그것이 존재하는지 혹은 아닌지의 두 가지 가능성을 내적으로 추정하고 그 차이에 근거하여 속성추론을 한다는 것이다. 이는 식 (2)에서 명시적으로 제시되어 있다. 사실 어떤 속성의 존재'여부'를 추론하는 과제는 본질적으로 이러한 두 가지 가능성에 대한 판단을 요청하고 있는 것이기 때문에 이는 개연성이 높은 가정이다. 만약 두 가지 가능성 (즉, 존재하거나 혹은 하지 않거나)이 유사해서 어떤 판단을 내리기 어려운 경우 그 차이값은 0이 될 것이고, 존재할 가능성이 높으면 차이값은 (+)값이 되고, 존재하지 않을 가능성이 높으면 차이값은 (-)값이 될 것이다. 본 실험에서는 참가자들의 이러한 판단과정의 결과를 반응척도에 반영할 수 있도록 그림 2와 같은 척도를 제시하여 그 결과를 모형의 예측과 비교하였다.

각 범주예시의 존재가능성에 대한 추정은 기존의 인과모형이론(CMT; Causal Model Theory)에 근거하였다. 인과모형이론은 특히 속성들이 서로 인과적으로 연결된 범주에서 개별 범주예시들의 전형성(대표성) 및 복수의 범주가 존재할 때의 범주화를 성공적으로 설명하였다. 본 연구는 인과모형이론에 기반하여 속성추론에서 참가자들이 어떤 속성이 존재하거나 혹은 존재하지 않을 때의 가능성에 대한 차이에 근거하여 모형의 추론을 예측하였다. 그런데 사람들의 인과적 지식은 명시적으로만 존재하는 것은 아니다. 인과모형은 표상수준에서는 모든 종류의 인과관계를 담아낼 수 있지만 인과관계에 대한 학습경로에 따라 추론의 결과가 달라질 가능성은 얼마든지 존재한다(e.g., Holyoak & Cheng, 2011; Waldmann & Hagmayer, 2005). 따라서 인과적 표상의 종류에 따른 범주속성에 대한 추론양상을 본 연구에서 제안한 모형이 설명해내는 지를 검증하는 것도 필요할 것이다.

본 연구에서는 인과관계의 특수성을 대표할 수 있는 공통원인과 공통효과 구조를 사용하여 속성추론을 검증하였다. 그런데 기존 연구에서는 범주내 속성들의 인과관계가 연쇄적으로 발생할 때의 범주화(예를 들어, X → Y → Z)에 대해 이론적 논쟁이 있어왔다(Ahn & Marsh, 2006; Rehder & Kim, 2010). 본 추론모형은 범주화 모형에 기반하기 때문에 속성간 인과관계가 사슬구조일 때에도 속성추론을 성공적으로 예측하는지를 검증할 필요가 있다.

마지막으로 향후 연구에서는 실제로 참가자들에게 어떤 속성 X가 존재하는 범주예시와 존재

하지 않은 범주예시의 전형성에 대해 평정하게 하고 이 두 값의 차이가 해당 속성 X에 대한 추론과 얼마나 일치하는지를 피험자내 설계를 통해 검증하는 것이 필요할 것이다. 이를 통해 본 모형의 핵심 기제인 존재와 부재의 가능성의 차이에 근거하여 속성추론이 발생한다는 것을 간접적인 모델링이 아닌 직접적인 데이터로 보여줄 수 있을 것이다.

참고문헌

Ahn, W., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81-121.

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69*, 135-178.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361-416.

Bailenson, J. N., Shum, M. S., Atran, D. L., Medin, J. D., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84*(1), 1-53.

Hadjichristidis, C., Sloman, S. A., Stevenson, R., & Over, D. (2004). Feature centrality and property induction. *Cognitive Science, 28*, 45-74.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the Causal Markov Condition. *British Journal for the Philosophy of Science, 50*, 521-583.

Heit, E., & Rubinstein, J. (1994). Similarity and Property Effects in Inductive Reasoning. *Journal of Experimental Psychology, 20*, 411-422.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology, 62*, 135-163.

Kim, N. S., Luhmann, C. C., Pierce, M. E., & Ryan, M. M. (2009). The conceptual centrality of causal cycles. *Memory & Cognition, 37*, 744-758.

Kim, N. S., Yopchick, J. E., & de Kwaadsteniet, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review, 15*(1), 81-88.

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*, 331-355.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251-295.

Marsh, J., & Ahn, W. (2006). The role of causal status versus inter-feature links in feature weighting. In

R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 561-566). Mahwah, NJ: Erlbaum.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517-532.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185-200.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: 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*, 811-828.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141-1159.

Rehder, B. (2006). When causality and similarity compete in category-based property induction. *Memory & Cognition, 34*, 3-16.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science, 33*, 301-343.

Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264-314.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91*, 113-153.

Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 659-683.

Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition, 37*, 715-729.

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1171-1206.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 641-648.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231-280.

Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking and Reasoning, 3*, 81-110.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive*

Science, 22, 189-228.

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*, 193-208.

Springer, K. and Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development, 62*, 767-781.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 216-227.

1차 원고 접수: 2017. 11. 12
1차 심사 완료: 2017. 12. 11
2차 원고 접수: 2017. 12. 15
2차 심사 완료: 2017. 12. 16
최종 게재확정: 2017. 12. 18

(*Abstract*)

Modeling feature inference in causal categories

ShinWoo Kim Hyung-Chul O. Li

Kwangwoon University

Early research into category-based feature inference reported various phenomena in human thinking including typicality, diversity, similarity effects, etc. Later research discovered that participants' prior knowledge has an extensive influence on these sorts of reasoning. The current research tested the effects of causal knowledge on feature inference and conducted modeling on the results. Participants performed feature inference for categories consisted of four features where the features were connected either in common cause or common effect structure. The results showed typicality effects along with violations of causal Markov condition in common cause structure and causal discounting in common effect structure. To model the results, it was assumed that participants perform feature inference based on the difference between the probabilities of an exemplar with the target feature and an exemplar without the target feature (that is, $p(E_{F(X)}|Cat)$-$p(E_{F(\sim X)}|Cat)$). Exemplar probabilities were computed based on causal model theory (Rehder, 2003) and applied to inference for target features. The results showed that the model predicts not only typicality effects but also violations of causal Markov condition and causal discounting observed in participants' data.

Key words : Category-based feature inference, causal reasoning, causal model theory, typicality