

특별호: 심리학과 4차 산업혁명 2

인공지능 기술의 수용성에 미치는 공정세상민음의 효과*

김 시 내¹⁾

손 영 우[†]

인공지능 기술의 발전으로 다양한 직업에 인공지능이 도입되어 사람을 대체하고 있다. 본 연구는 인공지능을 인간이 주도하는 현행체계의 대안으로 보고, 체제정당화 이론에 기반하여 공정세상 민음이 인공지능 수용성에 미치는 영향을 살펴보았다. 공정세상민음이 강할수록 인간주도의 기존의 체제를 지지하고 공정세상민음이 약할수록 대안적인 체제인 인공지능 행위주체(Agent)를 지지할 것으로 예상하였다. 연구1은 사람과 인공지능을 직접적으로 비교 선택하는 문항을 사용하여 공정세상민음의 효과를 검증하였고, 연구2는 인사 선발절차 맥락에서 개체 간 실험설계를 통해 행위주체유형(사람 vs. 인공지능)에 따른 결과수용성에 공정세상민음이 미치는 조절효과를 검증하였다. 연구결과, 연구1에서는 약한 공정세상민음은 (사람과 비교한 상대적인) 인공지능의 공정성을 매개로 인공지능의 직업 유능성을 예측하여 매개모형 가설을 지지하였다. 연구2에서는 합격조건에서는 결과수용성에 대한 공정세상민음의 조절효과가 유의하여, 공정세상민음이 약한 사람은 인공지능 조건에서 더 높은 결과수용성을 나타냈다. 그러나 불합격조건에서는 조절효과가 지지되지 않았다. 두 연구의 결과, 참여자들은 공정세상민음이 약할수록 사람보다 인공지능을 공정하고 유능하게 인식하였다. 연구 결과를 바탕으로 인공지능 수용성의 차이에 영향을 미치는 요인을 살펴보고, 연구의 의의, 한계점 및 후속연구에 대한 제언을 논의하였다.

주요어 : 인간-인공지능 상호작용, 공정세상민음, 체제정당화, 인공지능 기술 수용성, 인공지능 면접

* 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[2016-0-00562(R0124-16-0002). 상대방의 감정을 추론, 판단하여 그에 맞추어 대화하고 대응할 수 있는 감성지능 기술 연구 개발].

1) 김시내, 연세대학교 심리학과 박사과정생, E-mail: sinaekim71@gmail.com

† 교신저자: 손영우, 연세대학교 심리학과 교수, 서울특별시 서대문구 신촌동 연세로 50

E-mail: ysohn@yonsei.ac.kr

일상생활과 산업전반에서 인공지능 기술의 활용이 급속도로 확대되고 있다. 산업현장에서는 인공지능 기술에 기반한 시스템과 로봇이 사람과 협업을 하고, 개인별로 차별화된 마케팅전략을 구사하고, 일상생활에서는 자율주행차가 실용화단계에 진입하고 있다. 기계학습 분야의 발전으로 인공지능은 산업용 공장자동화를 넘어서 면접이나 인사선발 절차와 같이 사람과의 사회적 상호작용을 필요로 하는 분야에 활발하게 도입되고 있다. 기업은 인사관리 시스템을 사용하여 단순한 이력서 분류를 넘어서 인터넷을 통해 지원자를 모집하고, 필요한 업무능력을 가려내기 위해 업무 특성 분석을 수행하고, 인지능력검사 및 성격검사를 수행하고 있다(Stone et al., 2015). 인공지능은 성격분석과 감정분석에도 인간에 버금가는 수준으로 발전해서(Bartlet et al., 2014; Youyou et al., 2015), 지원자 응대와 인터뷰와 같이 전통적으로 사람이 수행하던 절차도 대신하거나 보조하게 되었다(Zielinski, 2017). 최근에는 입사지원자들을 대상으로 면접을 진행하고 지원자의 성격과 행동까지 예측하는 통합적인 선발절차 시스템이 개발되고 있다(Khosla et al., 2016). 방대한 데이터처리와 딥러닝기술로 예측력이 높아진 인공지능은, 결과의 윤리적 무게가 무거운 의료 및 법률분야에서도 인간의 의사결정에 사용되고 있다(Georgia State University, 2018; Parkin, 2016).

기술주도의 이러한 변화는 사회구성원의 일상과 심리에 영향을 미치고 기술에 대한 태도와 수용성에도 다양한 영향을 미친다. 산업현장의 자동화기술의 발달과 로봇의 도입은 실제 고용을 하락과 임금감소에 영향을 미치고(Acemoglu & Restrepo, 2017), 직업의 특성에 따라서는 향후 자신의 직업이 세상에 존재하

지 않게 될 것이라는 불안을 야기한다(Nam, 2019). 기계에 의해 인간이 대체될 수 있다는 심리적 불안을 야기하고(Peters, 2017), 인공지능이 유능하다고 인식할수록 직무불안정을 느끼고 일에 대한 냉소주의를 느낀다(Park & Jung, 2020). 자신의 직업과 관련한 문제 뿐 아니라, 인공지능이 다양한 영역에 도입되어 의사결정에 관여하는 것에 대해서도 부정적인 인식을 가진다. 인공지능이 인간보다 방대한 데이터에 기반한 의사결정을 할 수 있음에도 불구하고, 다양한 윤리적 의사결정을 수행해야 하는 자율주행차량, 의료현장, 법원의 가석방결정, 군에서의 드론사격 등의 문제에 있어서 사람들은 인공지능이 의사결정을 내리는 것에 반대한다(Bigman & Gray, 2018).

한편으로는 사회적 상호작용을 하는 인공지능 도입을 개의치 않거나 오히려 사람보다 더 공정하거나 유능하다고 인식하고 선호하는 결과들도 존재한다. 팀장역할이나 인사선발 면접관을 사람이 맡던 인공지능이 맡던 공정성 인식에 차이가 없거나(Otting, & Maier, 2018; Suen et al., 2019), 작성자를 조작한 신문기사 평가실험에서는 사람보다 인공지능이 기자인 경우에 더 유능하게 평가하였으며(Jung et al., 2017), 상담맥락에서는 사람상담자보다 인공지능 VR 에이전트에게 민감한 사적 정보를 털어놓는 것을 더 마음 편하게 느꼈다(Pickard et al., 2016). 또, 인사 선발절차를 진행하는 인사담당자가 사람인 경우보다 인공지능인 경우에 더 신뢰하였다(Min et al., 2018).

인공지능이 일상과 직업장면에 광범위하게 도입되고 있으며, 그로 인한 사회변화가 불가피하다. 인공지능은 단순한 편의성을 제공하는 것에 그치지 않고 사회에 다양한 문제를 제기한다. 누군가에게는 기회가 되지만 누군

가에게는 직업불안의 요소가 되고, 또 다른 누군가는 윤리적인 문제를 제기할 수 있어, 향후 사회적 논의와 합의가 필요해질 것이다. 이러한 논의가 올바른 방향으로 진행되기 위해서는 실증적인 연구에 기반한 논의가 이루어져야 한다. 현재 사람과 비교한 인공지능에 대한 상대적인 수용성에 대한 연구결과는 혼재되어 있고, 수용성에 영향을 미치는 요인이나 기저에 대한 연구는 매우 부족하다. 다양한 관점에서 실증적인 연구가 이루어져야 할 필요가 있다.

본 연구는 인공지능을 기존의 인간주도 체제의 대안으로 보고, 인공지능에 대한 수용성에 영향을 미치는 요인으로 체제정당화 이론과 공정세상믿음에 주목하였다. 본 연구자들은 공정세상믿음이 강할수록 인간주도의 기존의 체제를 지지하고 공정세상믿음이 약할수록 대안적인 체제를 지지할 것으로 예측하고 두 개의 연구를 통해 이를 검증하고자 하였다.

인간-컴퓨터 상호작용(Human computer interaction: HCI)

Computers are Social Actors(CASA) 패러다임은 사람들이 컴퓨터나 컴퓨터로 구현된 에이전트와 상호작용할 때에 인간과 인간간의 상호작용에서 사용되는 사회적 규범을 적용하고, 상호작용 시 마치 사람과 상호작용 하는 것과 유사하게 반응한다는 패러다임이다(Reeves & Nass, 1996). 사람과 컴퓨터 간의 사회적 상호작용에 대한 다양한 실험연구에서 참여자들은 작은 사회적 규범 주어져도 컴퓨터를 사람을 대하듯이 하여 컴퓨터에게 예의 바르고 친절하게 대했고(Reeves & Nass, 1996), 정교하게 인간의 모양이나 음성을 모사하지 않는 봇을 사

용한 실험에서조차 봇 에이전트를 인간과 다름없이 신뢰하고 상호작용하고자 했다(Edwards et al., 2014). 상호작용 시의 반응은 상황의 익숙함 정도나 컴퓨터 에이전트의 유형이나 형태 및 의인화 정도에 따라 달라질 수 있으며, 사용자의 개인특성에 따라 달라질 수 있을 것이다(Nass & Moon, 2000).

인간 vs. 인공지능을 비교하는 연구가 증가함에 따라 행위주체 유형에 따른 상호작용이나 반응에 차이가 발견되는 연구가 증가하고 있다. 예를 들어 트롤리 딜레마(많은 사람을 살리기 위해 한 사람을 희생시키는 행동을 선택할 것인지 아니면 예정된 대로 다수의 사람이 죽더라도 한 사람을 희생시키는 적극적인 행동을 취하지는 것을 선택할 것인지 묻는 윤리실험 패러다임)를 활용한 인간과 로봇의 의사결정 윤리성 판단에 대한 연구에서, 참여자들은 비록 다섯 명을 살리더라도 인간이 행동을 취해 한명의 방관자를 희생할 때에 비도덕적이라고 평가한 반면 로봇의 경우에는 행동을 취하지 않았을 때(그 결과 다섯 명이 죽게 되는 때)에 비도덕적이라 평가하였다(Malle et al., 2015). HCI 분야의 많은 연구는 인간이 컴퓨터, 인공지능, 로봇 등의 비인간(non-human) 행위주체에 대해 인간에게 적용하는 것과 같은 사회적 규범을 적용한다는 CASA 패러다임을 기반으로 하고 있다. 연장선상에서 윤리문제에 있어서 비인간 행위주체에 대해서도 인간에 대한 윤리판단 기준을 적용하여 도덕성 판단을 할 것으로 기대한다. 그런데 행위주체 유형에 따라 행동/무행동에 대한 도덕성 평가가 다르다면 로봇에게 적용하는 윤리규범의 개념 자체가 인간에게 적용하는 것과 다르거나 로봇을 인간처럼 윤리규범이 적용되는 대상으로 인식하지 않는 것일 수 있다. 만약 그

렇다면 CASA 패러다임은 적용되지 않는 것으로 추론할 수 있다.

Malle 과 동료들은 후속연구에서 이를 검증하였는데, 행위주체 유형에 따라 도덕적으로 받아들여지는 행동은 달랐으나, 그 행동의 도덕성을 판단하는 기준과 작동방식은 동일한 것으로 나타났다(Voiklis et al., 2016). 즉 행위주체 유형에 따라서 선호되는 행동양식은 다르지만, 각 행동양식의 도덕성 평가는 사건의 결과, 금지된 정도, 귀인을 기준으로 한 도덕적 허용성과 비난성이라는 동일한 도덕개념과 기준에 기반하고 있다. 따라서 행위주체 유형에 따라 윤리규범의 개념이 다른 것이라기 보다는 인간과 로봇에 대한 기능적 요구사항이나 역할에 대한 기대 차이에 기인한 것으로 볼 수 있다. 비인간 행위주체가 기존에 인간이 해오던 일이나 의사결정을 일부 대체할 수 있는 수준으로 기술이 급격히 발전함에 따라 다양한 분야에서 인간 vs. 인공지능 유형 차이에 따른 상호작용이나 반응의 차이를 야기하는 요인에 대한 연구가 필요해지고 있다.

인공지능에 대한 기대

인공지능 기술이 일상과 직업장면에 다양하게 활용되기 시작했지만, 아직은 자연어 처리 기술 등이 완벽하지 않고 모든 사람이 인공지능기술을 일상적으로 사용하는 것은 아니다. 그럼에도 불구하고 사람들은 인공지능의 능력에 대한 각자의 판단을 가지고 있는데, 이러한 판단은 심리학 이론에 근거해볼 때 기존에 가지고 있는 믿음과 기대에 영향을 받은 것일 수밖에 없다. 기대이론(Sanchez et al., 2000)과 공정성 휴리스틱이론(Lind, 2001)에 따르면 인간의 판단은 기존에 가지고 있던 믿음이나 기

대하는 바에 의해 영향을 받으며, 평가대상에 대해 초기에 취득한 공정성 정보를 휴리스틱 삼아 대상의 공정성 평가의 근거로 삼는다. 내용이 동일한 기사의 출처를 조작한 연구에서 참여자들은 컴퓨터가 선정한 기사를 뉴스 편집자가 선정한 기사보다 더 질이 높다고 평가하였는데(Sundar & Nass, 2001), Sundar(2008)는 이를 기계휴리스틱(Machine heuristic)으로 설명하였다. 즉 컴퓨터에 대해 갖는 객관성과 비편향성에 대한 기대가 기사의 질을 평가하는데 영향을 미친 것으로 제시하였다.

사전에 입사시험을 공정할 것으로 기대한 사람일수록 실제 시험을 치룬 후에도 높은 공정성 인식을 보였으며, 입사수락율도 높게 나타났다(Bell et al., 2006), 동일한 절차라도 긍정적으로 프레임한 경우에는 부정적으로 프레임한 경우보다 더 공정하게 인식하였다(Gamliel, & Peer, 2009). 조직 내에서의 업무조정이나 인사선발 업무의 의사결정 알고리즘에 관한 연구에서 한 참여자는 “나는 사용된 인공지능 알고리즘에 대해 아는 바는 없지만, 분명히 사람보다 공정할 것이다”라고 답하였다(Lee, 2018). Lee의 연구 결과는 전반적으로 알고리즘보다 사람을 더 공정하다고 평가한 것으로 나타났으나, 행위주체 유형별로 공정성의 근거를 기술한 문항에서 알고리즘 매니저가 공정하다고 판단하는 근거로 인간보다 더 객관적, 효율적이며 선호에 좌우되지 않고 비편향적이기 때문이라고 답한 반면, 사람 매니저가 공정하다고 판단하는 근거로는 매니저에게 부여된 권한과 입사지원자나 직원의 능력이나 경험을 알아볼 수 있는 능력에 귀인하였다(Lee, 2018). 이러한 연구결과들은 기존에 세상에 대해 가지고 있는 믿음이나 인공지능에 대해 가지고 있는 기대가 인공지능에 대한 평가

에 영향을 미칠 수 있음을 함의한다.

인공지능기술 수용성에 대한 공정세상믿음의 효과

Lerner(1980)는 사람들은 합당한 대우를 받는 공정한 세상에 살고 있다고 믿고 싶어 한다는 공정세상 믿음(Belief in a just world)을 제안하였다. 공정세상 믿음은 불확실성에 대한 통제감을 가지려고 하는 동기에 기반하여 형성되기도 하고, 개인의 경험이나 사회학습을 통해 형성되기도 한다(Hafer & Sutton, 2016). 세상이 공정하다는 믿음은 현실을 반영한다기 보다는 공정하다고 믿고자 하는 욕구를 반영한 것으로, 공정세상 믿음이 높은 사람들은 현실의 불공정함이나 우연성에도 불구하고 현재의 사회체제나 관습은 공정하며 사람들에게 일어나는 좋고 나쁜 일들은 모두 그럴만해서 일어나는 것이라고 현재의 체제를 정당화한다(System Justification Theory; Jost & Hunyady, 2005). 기존의 체제와 관습을 정당한 것으로 여기고 현상을 유지하려고 하고, 사회변화를 받아들이지 않고 저항하게 된다(Friesen et al., 2019). 명백히 불공정한 상황에서도 공정세상믿음이 강할수록 상황을 공정하다고 인식한다(Pedersen et al., 2017).

직업의 맥락에서 의사결정을 할 때에 인간은 인지적 한계나 고정관념, 암묵적 편향 및 감정적 선호도 등으로 인해 업무와 무관한 다른 요소들의 영향을 받는다. 예를 들어 인사선발담당자는 지원자의 업무 역량과 무관한 다른 요소들(예: 면접순서, 외모 등)의 영향을 받을 수 있다(Baert, 2018; Bóo et al., 2013; Page & Page, 2010; Deng et al., 2020). 인공지능 행위주체는 인간처럼 인지적 한계에 제약

받지 않고 감정적 선호도에 영향 받지 않는다. 인공지능의 이러한 강점으로 인해 사람들은 인공지능이 기존에 인간이 하던 일을 대체하고(Frey & Osborne, 2017) 현재의 인간주도의 체제를 위협할 가능성에 대해 우려한다. 이를 체제정당화 이론에 기반해 보면, 인공지능과 같이 기존의 체제를 대체할 대안의 우려 하에서, 공정세상믿음이 강한 사람은 인간주도의 기존 체제의 단점에도 불구하고 인공지능과 같은 새로운 시스템을 도입하는 것에 저항감을 갖고 기존의 체제를 정당화하기 위해서 기존의 전통적인 시스템과 절차를 공정하고 더 나은 것으로 인식할 수 있을 것이다.

공정세상믿음이 언제나 체제정당화로 이어지는 것은 아니다. 대안에 대한 효능감이 있는 경우에는 공정성 믿음이 기존체제를 정당화하지 않고 변화행동으로 이어질 수 있다(Beierlein, 2011; White et al., 2012). 아직 자연어처리를 인간처럼 구사할 수 있는 인공지능 시스템은 구현되거나 상용화되어 있지 않다. 인간과 인공지능을 비교한 평가는, 본질적으로 현실 인간에 대한 평가와 기술발전에 따라 가능해지는 미래의 현실인 인공지능에 대한 기대를 비교하는 것으로 볼 수 있다. 사람과 인공지능을 비교 평가할 때, 인공지능을 다가오는 미래의 대안적인 체제로 인식한다면 체제정당화 경향이 줄어들 수 있을 것이다.

4차 산업혁명과 직업대체에 관한 연구는 직업의 기능적 특성, 직업수행에 필요한 기능의 자동화 가능성을 기준으로 직업의 소멸가능성을 제시하고 있다(Lee, 2017; Choi, 2019; Frey, & Osborne, 2017). 직업의 소멸은 인간이 해당 직업을 수행하는 대신 컴퓨터, 인공지능, 로봇이 직업적 기능을 대체하게 됨을 뜻한다. Lee (2017)의 연구에서 볼 수 있듯이, 전문가를 물

론 일반인도 기술의 진보로 인해 업무수행에 필요한 여러 역량에 있어서 컴퓨터가 사람을 능가하게 되고, 사람을 대체할 것으로 예측하고 있다. 따라서 공정세상민음의 인공지능 수용성에 대한 효과가 유의하다면, 즉 공정세상민음이 약할수록 인공지능을 대안적인 체제로 선호한다면, 사람보다 인공지능이 더 유능해질 것으로 인식할 것이다. 예를 들어 참여자간 설계로 진행된 Min과 동료들의 연구(2018)에서 선발절차에 대한 만족도와 인사담당자에 대한 신뢰에 있어서 공정세상민음이 강한 사람들은 인간과 인공지능 사이에 차이가 없었으나, 공정세상민음이 약한 사람들은 인간보다 인공지능을 더 신뢰하였다.

이런 예상은 기대이론의 관점에서 설명될 수 있다. 기대이론에 따르면 직간접적인 경험 이외에도 기존에 가지고 있던 믿음이 공정성 기대를 형성하고, 기대는 관련정보를 처리하는 인지프로세스를 통해 공정성 인식을 형성한다(Bell et al., 2004). 공정세상민음이라는 기존에 가지고 있는 믿음으로 인해 현재 작동하는 세상 즉 인간이 진행하는 전통적인 절차나 의사결정을 공정하다고 기대하고, 그 프레임 안에서는 인공지능보다 인간을 공정하고 유능하다고 판단하고 의사결정의 결과를 수용할 것으로 예상된다. 반대로 공정세상 믿음이 낮은 사람은 현재 작동하는 전통적인 의사결정 절차의 공정성에 대한 기대가 낮을 수 있다. 이 경우 인간과 인공지능 사이에서 선택을 해야 한다면, 대안적인 체제인 인공지능의 공정성이나 유능함에 대한 기대가 상대적으로 더 높게 형성될 수 있을 것이다. 따라서 공정세상민음이 강한 사람은 공정세상민음이 약한 사람에 비해 현재 시행되고 있는 기존의 체제 즉 사람에 의한 업무수행과 의사결정 절차를

더 유능한 것으로 평가할 것으로 예상되는 반면, 공정세상민음이 약한 사람은 인공지능 기반의 자동화된 절차를 더 유능한 것으로 평가할 것으로 예상된다.

공정세상민음의 효과를 검증하기 위해 두 가지 연구를 실행하였다. 연구1에서는 사람과 인공지능 중 어느 쪽이 다양한 직업수행에 유능할 것인지 직접적으로 비교 선택하는 문항을 사용하여, 공정세상민음이 약할수록 인공지능을 더 공정하고 유능하다고 판단하는지 조사하였다. 연구2에서는 구체적으로 기업의 인사선발절차의 맥락에서 개체 간 실험설계를 통해 공정세상민음의 효과를 검증하였다.

연구 1

연구1은 직업수행 일반에 있어서 공정세상민음의 효과를 검증하고자 하였다. 구체적으로 행위주체 유형에 따른 공정성 인식과 유능성 인식에 미치는 공정세상민음의 효과를 검증하고자 하였다. 변인 측정 시 사람과 인공지능 중 선택을 하도록 함으로써 인공지능이 인간 주도의 기존체제에 대한 대안적인 존재로 작동하는 지 여부를 직접적으로 확인하고자 하였다.

행위주체 직접비교

가설검증을 위해 행위주체 유형에 따른 공정성 인식이나 유능성의 차이를 검증하기 위해서는 인식의 대상이 되는 구체적인 직업의 예시를 필요로 한다. 본 연구에서는 인공지능이 부분적으로라도 도입되어 있거나 인공지능과 로봇 도입에 관한 HCI연구가 이루어지고

(예시) 사람과 인공지능(AI)의 판단의 공정성, 객관성을 비교하여 주십시오

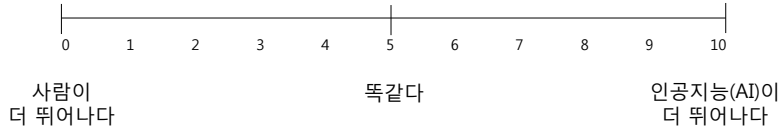


그림 1. 슬라이딩 척도의 예시

있는 다섯 가지 직업을 예시하여 일반의 인식을 조사하였다. 구체적으로, 각 참여자는 판사, 기자, 의사, 교사, 요양보호사 중 하나의 직업을 평가하였다. 해당 직업을 유능하게 수행하는데 판단의 공정성과 객관성이 얼마나 중요한지 측정하고, 사람과 인공지능 중 어느 쪽이 더 공정하고 객관적인지 측정하고, 장차 해당 직업을 사람과 인공지능 중 어느 쪽이 더 뛰어나게 수행하게 될 것이라고 생각하는지 유능성 기대를 측정하였다.

사람과 인공지능을 직접적으로 비교 선택하도록 하기 위해, 사람과 비교한 인공지능의 상대적인 공정성과 유능성을 측정하는 데에 슬라이딩 척도를 사용하였다. 예를 들어 공정성의 경우 사람과 인공지능 중 어느 쪽이 더 공정하다고 인식하는지 그림 1과 같이 직접 비교하도록 하였다. 측정값은 0-10으로 그 값이 높을수록 사람보다 인공지능의 공정성을 높게 인식함을 뜻한다.

공정세상민심의 효과 검증

공정세상민심이 약할수록 사람보다 인공지능이 더 공정하고 더 유능해질 것으로 인식한다는 공정세상민심의 효과에 대한 구체적인 가설은 다음과 같다.

가설 1. (공정성 가설) 공정세상민심이 약할

수록 사람보다 인공지능 행위주체를 더 공정하다고 평가할 것이다.

가설 2. (유능성 가설) 공정성이 업무수행 유능성에 중요한 직업들에 있어서 공정세상민심이 약할수록 사람보다 인공지능 행위주체가 직업수행 시 더 유능하다고 평가할 것이다.

직업맥락에서 공정세상민심이 공정성과 유능성에 미치는 영향을 조사하기 위해서는, 공정성이 업무수행에 중요한 직업을 선택하여 검증하여야 한다. 본 연구에서는 다섯 가지 직업에 대해 각 직업을 유능하게 해내는 데에 공정성이 얼마나 중요하다고 인식하는지 측정하고, 중간값 이상으로 중요하다고 인식하는 직업들의 측정값만 사용하였다. 선정된 직업들은 업무를 유능하게 수행하는데 공정성이 중요한 직업들이므로, CASA 패러다임에 따라 인공지능이 해당 직업을 유능하게 수행하기 위해서 사람의 경우와 마찬가지로 공정성이 요구될 것이다. 즉 CASA 패러다임에 의하면, 직업수행에 공정성이 중요한 직업의 경우, 공정세상민심이 약해서 사람보다 인공지능을 더 공정하게 인식하는 사람은 해당 직업수행을 인공지능이 더 잘하게 될 것으로 예측할 것이다. 이를 검증하기 위해서는 그림 2에서와 같이 매개모형을 사용하였다.

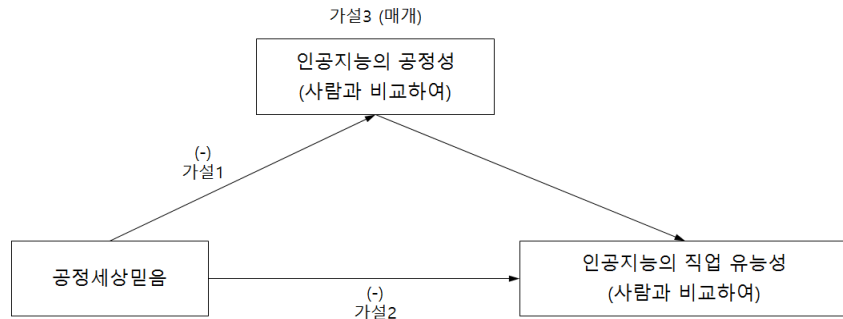


그림 2. 매개모형 제시

가설 3. (매개효과) 인공지능의 공정성(사람과 비교하여)이 공정세상믿음과 인공지능의 직업유능성(사람과 비교하여)의 관계를 매개할 것이다. 즉, 공정세상믿음이 약한 사람은 사람보다 인공지능 행위주체를 더 공정하다고 인식하고, 공정성이 업무수행 유능성에 중요한 직업들에 있어서 사람보다 인공지능 행위주체가 직업수행 시 더 유능하다고 판단할 것이다.

방 법

연구대상

연구1은 ‘상대방의 감정을 추론, 판단하여 그에 맞추어 대화하고 대응할 수 있는 감성지능 기술 연구’를 위해 수행된 설문연구의 일부로, 인공지능과 직업요소에 대한 일반적인 인식을 조사하기 위해 일반 성인 474명(남성 239명, 여성 235명)을 모집하였다. 평균연령은 남성 43.23세(SD = 12.77), 여성 39.40세(SD = 11.73)이다. 참여자는 온라인 설문업체인 데이터스프링을 통해 모집하였고, 설문을 완료한 참여자는 온라인 업체가 지급하는 포인트를 보상으로 지급받았다. 본 연구는 연세대학교

생명윤리위원회(IRB)의 승인을 받고 진행되었으며, IRB의 승인을 받은 참가자 설명문을 읽고 자발적인 동의를 한 참여자만 대상으로 진행하였다.

연구절차

참가자들은 다섯 가지 직업조건(판사, 기자, 의사, 교사, 요양보호사)에 무선택당 되어, 판단의 공정성과 객관성이 직업수행에 얼마나 중요한지 답하고, 해당 직업을 사람과 인공지능 중 어느 쪽이 더 잘 할 것인지(인공지능의 유능성 기대)에 답하였다. 추가적으로 할당된 직업과 무관하게 일반적으로 사람과 인공지능 중 어느 쪽이 더 판단의 공정성과 객관성이 뛰어날 것인지 묻는 문항에 답하였다. 사람과 인공지능을 비교하는 문항은 모두 슬라이딩척도를 사용하여, 직접적으로 사람과 인공지능을 비교하여 선택하도록 문항을 구성하였다.

측정도구

역량(Competence) 중요도

할당된 조건의 직무(판사, 기자, 의사, 교사,

요양보호사)를 유능하게 잘하기 위해 판단의 공정성과 객관성이 얼마나 중요한지 묻는 문항에 리커트 6점 척도(1=전혀 중요하지 않다. 6=매우 중요하다)를 사용하여 답하였다.

인공지능의 공정성(인간과 비교한 상대적 공정성)

사람과 인공지능을 비교하여 어느 쪽이 더 판단이 공정하고 객관적인 지 조사하였다. 슬라이딩 스케일을 사용하였고, 측정은 0~10점으로 수집되었다(0=사람이 더 뛰어나다, 5=똑같다, 10=인공지능이 더 뛰어나다). 따라서 평점이 높을수록 해당 능력이 사람보다 인공지능이 더 뛰어나다고 인식함을 나타낸다.

인공지능의 직업 유능성(인간과 비교한 상대적 유능성)

할당된 직업에 대해서, 인공지능과 로봇기술이 발달한 미래에 해당직업을 사람과 인공지능을 비교하여 상대적으로 어느 쪽이 더 잘 할 것이라고 생각하는지 조사하였다. 슬라이딩 스케일을 사용하였고, 측정은 0~10점으로 수집되었다(0=사람이 더 뛰어나다, 5=똑같다, 10=인공지능이 더 뛰어나다). 따라서 평점이

높을수록 해당 직업 수행을 사람보다 인공지능이 더 뛰어나게 될 것으로 인식함을 나타낸다.

공정세상 믿음

단문항 ‘당신은 세상이 얼마나 공정하다고 생각하십니까?’로 측정하였고, 리커트 6점 척도(1=전혀 공정하지 않다, 6=매우 공정하다)를 사용하였다.

분석방법

본 연구에서는 자료분석을 위해 SPSS 24.0 프로그램과 PROCESS(Hayes' 2013)를 사용하였다. 측정 변인의 관계성을 알아보기 위해 기술통계 분석과 상관분석을 실시하였다. 매개 모형 검증을 위해 PROCESS를 사용하였다. 신뢰수준은 모든 분석에서 95% 신뢰수준을 사용하였다.

결 과

기술통계분석 및 상관관계를 표 1에 제시하

표 1. 기술적 분석과 상관관계

	1	2	3	4	5	M	SD
1. 참여자성별							
2. 나이	-.155**					41.33	12.40
3. 직업수행 시 판단의 공정성, 객관성의 중요도 ^a	.08	.07				5.46	.73
4. 인간과 비교한 인공지능의 판단의 공정성, 객관성 ^b	.02	-.04	.121**			6.62	2.22
5. 인간과 비교한 인공지능의 직업유능성 기대 ^b	-.141**	.126	-.02	.174**		5.29	2.69
6. 공정세상믿음 ^a	.03	.111*	-.06	-.110*	-.132**	2.25	.90

^a: 6점척도, ^b:0-10점 슬라이딩척도로 0은 사람이 훨씬 뛰어나 10은 인공지능이 훨씬 뛰어나를 의미함,

** : $p < .01$, * : $p < .05$



그림 3. 사람과 인공지능의 역량비교

고, 인간과 비교한 인공지능의 판단의 공정성, 객관성과 직업유능성 기대를 시각화하여 제시하였다(그림 3). 직업수행 시 판단의 공정성, 객관성의 중요도는 판사의 경우 5.79/6점, 기자 5.64/6점, 의사 5.66/6점, 교사 5.41/6점, 요양보호사 5.00/6점으로 모든 직업에서 중간값 3.5를 크게 상회하였다. 판단의 공정성, 객관성이 모든 직업에서 매우 중요하게 나타났으므로 가설검증에 모든 직업조건의 참여자의 데이터를 사용하도록 하였다.

매개효과 분석

가설검증을 위해 SPSS 24.0와 PROCESS (Hayes' 2013)를 사용하여 분석하였다. 선행연구에 따르면 인공지능과 로봇기술의 다양한 영역에서 성별(Gessl et al., 2019; Kuo et al.,

2009; Schermerhorn et al., 2008) 및 나이(Assaker, 2020; Makkonen et al., 2017; Morris, & Venkatesh, 2000)에 의한 기술수용성의 차이가 보고되고 있다. 본 연구의 목적은 성별이나 나이에 따른 차이를 밝히는데 있지 않으므로, 가설검증 시 성별과 나이를 통제변인으로 투입하였다. 분석결과 공정세상 믿음은 인간과 비교한 인공지능의 상대적인 공정성을 유의하게 예측하여($B = -.267, t(470) = -2.34, p = .02$), 가설1은 지지되었다. 인공지능의 공정성은 공정세상 믿음을 통제하고도 인공지능의 직업유능성을 유의하게 예측하여($B = .203, t(469) = 3.76, p = .00$), 가설2는 지지되었다. 마지막으로 부트스트래핑을 통해 확인한 간접 효과는 $-.054$ 이고, 95% 신뢰구간은 0을 포함하지 않아, $CI[-.124, -.007]$, 가설3은 지지되었다 (표 2).

표 2. 매개분석(공정세상 믿음 → 인공지능의 공정성 인식 → 인공지능의 직업 유능성)

	Coefficient	S.E.	t	p	LLCI	LLUI
DV: AI 공정성						
공정세상 믿음	-.267	.11	-2.34	.02	-.49	-.04
DV: AI 직업유능성						
공정세상 믿음	-.372	.13	-2.77	.01	-.64	-.11
AI 공정성	.203	.05	3.76	.00	.10	.31
<u>Indirect effect (Boot strap)</u>						
공정세상 믿음→AI 공정성→AI 직업유능성	-.054	.03			-.12	-.01

AI 공정성: 사람과 비교한 인공지능의 공정성, AI 직업유능성: 사람과 비교한 인공지능의 직업유능성 예측

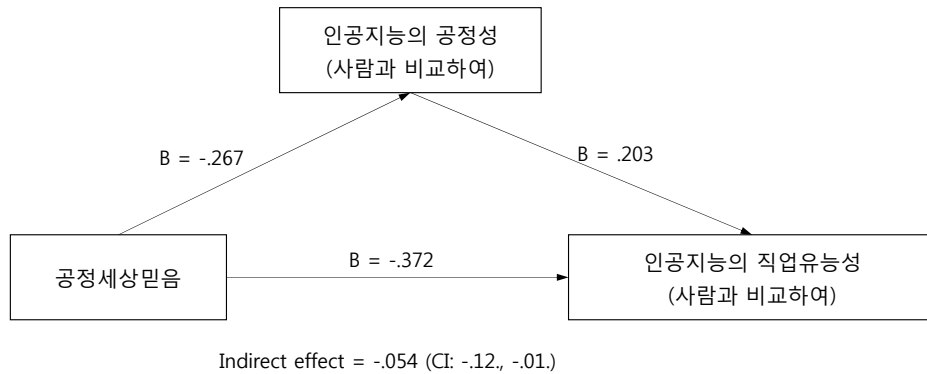


그림 4. 매개효과 분석

즉, 사람보다 인공지능이 공정하고 객관적이라고 생각하는 사람은 업무(판사, 기자, 의사, 교사, 요양보호사) 수행 시 사람보다 인공지능이 더 유능할 것으로 예측했다. 또 세상이 공정하다고 믿지 않는 사람일수록 사람보다 인공지능이 공정하고 객관적이라고 평가하였다. 간접효과 검증결과, 세상이 공정하다는 믿음이 약한 사람일수록 사람보다 인공지능이 더 공정하고 객관적이라는 인식을 매개로 하여, 장차 사람보다 인공지능이 업무(판사, 기자, 의사, 교사, 요양보호사) 수행을 더 유능하게 할 것으로 예측하였다. 매개모형 검증 결과를 시각화하여 그림 4에 제시하였다.

연구 2

연구1을 통해 공정세상믿음의 효과를 확인하였으나, 연구1에는 몇 가지 제한점이 있다. 공정세상믿음을 측정할 때 단문항을 사용하여 신뢰도를 제시할 수 없고, 행위주체 유형을 직접 비교선택한 장점이 있으나 일반적인 인식을 묻는 비교적 단순한 설문으로 구성되었다. 따라서 연구2에서는 다문항 척도를 사용

하여 공정세상믿음을 측정하여 연구1의 한계점을 보완하고, 인사선발절차라는 구체적인 맥락에서 개체 간 설계를 통해 행위주체유형에 따른 수용성의 차이를 검증하고, 공정세상믿음이 이에 미치는 조절효과를 검증하고자 하였다. 공정세상믿음의 조절효과는 공정세상믿음이 강할수록 인공지능조건보다 사람조건에서 결과수용성이 높게 나타나고, 공정세상믿음이 약할수록 사람조건보다 인공지능조건에서 결과수용성이 높게 나타날 것으로 예상하였다. 선발결과에 따라 수용성의 차이가 예상되므로, 실험조건은 2(행위주체: 사람 vs. 인공지능) X 2(선발결과: 합격 vs. 불합격)의 개체 간 설계로, 종속변인은 선발결과에 대한 수용성을 측정하였고 공정세상믿음은 연속형 변인으로 측정되었다.

선발결과에의 조절효과

인사선발절차는 의사결정의 결과가 수반되는 절차로, 결과수용성은 최종적으로 합격/불합격 여부에 영향을 받는다. 귀인이론(Ployhart & Harold, 2004)에 따르면 똑같은 정도의 공정성을 가진 선발절차를 거치더라도 합격(공정

피드백)과 불합격(부정피드백)으로 결과가 나뉘는 실험에 따라 최종적인 공정성 판단이나 수용성이 달라진다. 선행연구에 따르면 지원자는 선발시험을 통과하면 절차공정성을 높게 평가하고 시험에 떨어지면 낮게 평가했다(Bauer et al., 1998; Kluger & Rothstein, 1993). 합격한 지원자들은 그 이유를 내적이고 기질적인 요인에 귀인하고 불합격한 지원자들은 불합격의 이유를 자신이 통제할 수 없는 외적 요인에 귀인한다. 불합격한 사람은 외부통제를 매개로 하여 공정성인식을 형성하였고 합격한 사람은 개인적 통제나 안정성 같은 내적요인을 매개로 공정성을 인식하는 것으로 나타났다(Ababneh et al., 2014). 선발절차가 동일하게 통제된 실험에서 외적요인의 차이는 행위주체 유형으로 제한되므로, 선발결과 유형에 따른 귀인의 차이는 행위주체 유형에 의한 효과의 강화와 약화로 나타날 것으로 예상된다.

조건별 공정성을 조작한 실험에서는 공정한 절차와 불공정한 절차에 대한 공정성 인식의 차이가 불합격자보다 합격자에서 더 작았다(Schroth & Shah, 2000). 즉 불합격한 경우에 비해 합격한 경우 절차의 객관적인 공정성과 같은 외적인 요인이 지원자의 공정성 인식에 영향을 적게 미쳤다. 이에 비추어 보면, 합격조건에서는 행위주체 유형이 미치는 영향이 작을 수 있으므로, 합격조건에서는 공정세상민음의 효과도 불합격조건에서 보다 작을 수 있다. 부정피드백을 받은 지원자는 외적요인인 선발시험이나 선발절차에 실패를 귀인하게 됨에 따라 선발절차 자체의 공정성이나 타당성에 대해 더 강하게 비판적이 될 수 있다(Ployhart & Harold, 2004). 선발절차의 각 외적 요인(예: 행위주체의 유형)별 공정성에 민감해지면, 행위주체 유형으로 인한 반응차이

도 커질 수 있고 공정세상민음의 조절효과도 더 커질 것으로 예상할 수 있다. 구체적인 가설은 다음과 같다.

가설 4. 행위주체가 결과수용성에 미치는 효과는 공정세상민음에 따라 달라질 것이다. 즉, 공정세상민음이 클수록 인공지능보다 사람조건에서의 결과수용성이 높고, 공정세상민음이 작을수록 사람보다 인공지능조건에서의 결과수용성이 높을 것이다.

가설 5. 행위주체와 결과수용성의 관계에서 공정세상민음의 조절효과는 선발결과에 따라 달라질 것이다. 즉, 이 공정세상민음의 조절효과는 합격조건에서보다 불합격조건에서 강하게 나타날 것이다.

방 법

연구대상

본 연구에서는 잠재적인 입사지원자가 될 연령대인 학부생 190명(평균연령=22.55세, SD=1.87)을 대상으로 진행되었다. 이 중 남성은 90명(평균연령=23.02, SD=2.14)이고 여성은 100명(평균연령=22.12, SD=1.47)이다. 참여자는 온라인 설문업체인 데이터스프링을 통해 모집하였고, 설문을 완료한 참여자는 온라인 업체가 지급하는 포인트를 보상으로 지급받았다.

연구절차

본 연구는 연세대학교 생명윤리위원회의

(IRB) 승인을 받고 진행되었으며, IRB의 승인을 받은 참가자 설명문을 읽고 자발적인 동의를 한 참여자만을 대상으로 진행하였다. 실험 설계는 2(행위주체 유형: 사람 vs. 인공지능) X 2(선발결과 유형: 합격 vs. 불합격) 요인의 집단 간 설계(Between-group design)로, 참여자는 총 4개의 실험조건에 무선할당 되었다. 우선 참여자들은 인구통계적 정보와 공정세상믿음을 묻는 문항에 답한 후, 할당된 조건에 맞는 시나리오 자극을 읽었다. 시나리오를 읽고 난 후 결과수용성 문항에 답하였다.

연구도구

행위주체유형 조작

사람과 인공지능의 두 가지

결과유형 조작

합격과 불합격의 두 가지

실험자극

실험자극은 채용선발 절차를 진행하는 행위주체 유형별로 사람이 채용선발 절차를 진행하는 경우와, 인공지능이 진행하는 경우의 두 가지 시나리오로 제시되었다. 시나리오에는 지원자(‘지훈’)가 채용공고를 보고 지원 원서를 제출하고, 채용선발 담당자가 지원서류를 검토하고, 지원자가 서류전형을 통과하여 채용선발 담당자(행위주체)를 대상으로 면접을 보고, 채용선발 담당자가 서류 및 면접 등의 성과를 고려하여 최종 의사결정을 하는 채용선발 절차 전반이 묘사되었다. 시나리오 간의 내용은 채용선발 절차를 담당하는 행위주체의 유형(사람 vs. 인공지능)만 다를 뿐, 행위주체의 능력, 경험치, 지원자와의 대화내용, 진행

절차 모두를 동일하게 통제하였다. 선발절차가 모두 묘사된 후 실험의 선발결과 조건에 따라 지원자의 합격/불합격 여부가 제시되었다.

측정도구

공정세상믿음

Lucas et al.(2011)의 4차원 공정성 척도를 한국판으로 타당화한 한국어판 정당한 세상에 대한 믿음 척도(Kim et al., 2017)를 사용하였다. 4가지 차원은 분배공정성에 대한 일반적 믿음(DJ-others), 분배공정성에 대한 개인적 믿음(DJ-self), 절차공정성에 대한 일반적 믿음(PJ-others), 절차공정성에 대한 개인적 믿음(PJ-self)로 이루어져 있고 각 차원은 4개의 문항으로 구성되어 있다. 공정세상믿음은 영역과 대상에 따라 다양한 척도를 사용하여 측정한다(Hafer & Sutton, 2016). 선발과정의 전체 절차를 살피고 그에 따른 개인의 판단을 묻는 실험구조를 고려하여, 4개 차원 중 절차공정성에 대한 개인적인 믿음(PJ-self) 차원 4문항을 사용하여 공정세상믿음을 측정하였다. 예시문항은 ‘사람들은 일반적으로 나를 공정하게 평가한다’ 및 ‘나는 모든 상황에서 공정하게 평가받는다’이고, 7점 리커트 척도(1=그렇지 않다. 7=매우 그렇다)로 평정하였다. 이 척도의 내적 일치도(Cronbach’s α)는 0.83이었다.

결과수용성

결과수용성은 ‘귀하께서 ‘지원자(지훈)’의 입장이라면 시나리오에서 제시된 선발 결과를 받으셨을 때, 선발 결과를 어느 정도 받아들일 시겠습니까?’ 단문항으로 제시하였으며, 7점 리커트 척도(1=전혀 받아들일 수 없다, 7=매

우 잘 받아들일 수 있다)로 측정하였다.

위주체유형과 공정세상민음의 2원상호작용항이 유의한지 검증하였다.

분석방법

본 연구에서는 자료분석을 위해 SPSS 24.0 프로그램을 사용하였다. 본 연구의 실험조건 중 행위주체유형과 선발결과유형은 조작을 하였으나 공정세상민음은 연속형 변인으로 측정되어 일반적인 실험연구 분석방법인 분산분석으로 분석을 할 수 없으므로, 더미코딩을 통해 실험조건을 변수화하여 다중회귀분석을 시행하였다. 행위주체 주효과 검증에 있어서 사람조건은 0으로 더미코딩 하였고, 인공지능은 1로 더미코딩 하여, 행위주체가 사람인 실험조건과 비교해 인공지능 조건일 때 선발절차에 대한 결과수용성이 유의하게 달라지는 지 검증하였다. 선발결과의 주효과 검증에 있어서는 합격조건=0, 불합격 조건=1로 더미코딩 하였다. 공정세상 민음의 조절효과를 검증하기 위해서는 행위주체, 선발결과, 중심화한 공정세상민음의 상호작용항을 투입하여 Aiken과 West(1991)가 제안한 위계적 회귀분석을 실시하였다. 분석의 순서는, 우선 3원상호작용항의 유의성을 검증하기 위해 4단계에 걸친 위계적 회귀분석을 실시하였고, 3원상호작용항이 유의하므로 선발결과유형별(합격 vs. 불합격)로 행

결 과

측정변인의 기술통계적 특성과 변인 간의 상관관계는 표 3에 제시하였다. 다음으로 행위주체 유형에 따른 결과수용성에 대한 공정세상민음의 조절효과(가설4)를 검증하기 위해 위계적 회귀분석을 실시하였다. 선행 연구에 따르면 AI에 대한 반응은 성별에 따라 영향을 받으므로(Kuo et al., 2009; Schermerhorn et al., 2008), 성별은 통제변인으로 투입하였다. 1단계에서 통제변인인 성별을 투입하고, 2단계에서 주효과를 추가하였으며, 3단계에서 2원상호작용항을 투입하고, 4단계에서 더미코딩 된 행위주체 유형, 선발결과 유형과 평균 중심화 된 공정세상민음의 상호작용항을 투입하였다. 회귀분석 결과 공정세상민음의 조절효과는 통계적으로 유의하였다($p = .017$, $\Delta R^2 = .013$) (표 4).

3원상호작용이 유의하였으므로, 합격조건과 불합격조건으로 나누어 행위주체와 공정세상민음의 2원상호작용의 유의성을 검증하였다. 그 결과 공정세상민음의 조절효과는 불합격조

표 3. 기술적 분석

	1	2	3	평균	표준편차
1. 성별					
2. 나이	-.24**			22.54	1.87
3. 결과수용성	-.05	.00		4.61	1.66
4. 공정세상민음	.02	.01	.06	4.75	.94

** : 상관관계가 0.01 수준에서 유의(양측).

표 4. 회귀분석: 실험조건별 결과수용성에 대한 공정세상믿음의 조절효과

	<i>B</i>	<i>S.E.</i>	β	<i>t</i>	<i>p</i>	<i>R</i> ²	ΔR^2
1단계							
성별	-.170	.24	-.051	-.70	.483	.00	
2단계							
성별	-.139	.16	-.042	-.88	.382		
행위주체	-.078	.16	-.023	-.49	.626		
선발결과	-2.539	.16	-.766	-15.86	.000		
공정세상믿음	-.091	.09	-.052	-1.07	.287	.58	.577
3단계							
성별	-.133	.16	-.040	-.84	.400		
행위주체	-.105	.22	-.032	-.47	.638		
선발결과	-2.556	.23	-.771	-10.92	.000		
공정세상믿음	.172	.16	.098	1.09	.276		
행위주체X선발결과	.033	.32	.009	.10	.919		
행위주체X공정세상믿음	-.131	.17	-.053	-.77	.444		
선발결과X공정세상믿음	-.350	.17	-.147	-2.04	.043	.59	.010
4단계							
성별	-.172	.16	-.052	-1.10	.274		
행위주체	-.046	.22	-.014	-.21	.835		
선발결과	-2.561	.23	-.773	-11.08	.000		
공정세상믿음	.423	.19	.239	2.26	.025		
행위주체X선발결과	.025	.32	.006	.08	.938		
행위주체X공정세상믿음	-.596	.26	-.242	-2.33	.021		
선발결과X공정세상믿음	-.773	.24	-.325	-3.17	.002		
행위주체X선발결과X공정세상믿음	.823	.34	.242	2.41	.017	.60	.013

건에서는 유의하지 않았으나($B = .240, t = .98, p = .33$), 합격조건에서는 유의하였다($B = -.525, t = -2.32, p = .023, \Delta R^2 = .053$). 구체적으로로는 합격조건에서는 공정세상믿음이 낮을수록 사람조건보다 인공지능 조건에서의 결과

수용성이 높았고, 공정세상믿음이 높을수록 인공지능보다 사람조건에서의 수용성이 더 높게 나타났다. 합격조건에서의 조절효과를 시각화하여 그림 5에 제시하였다.

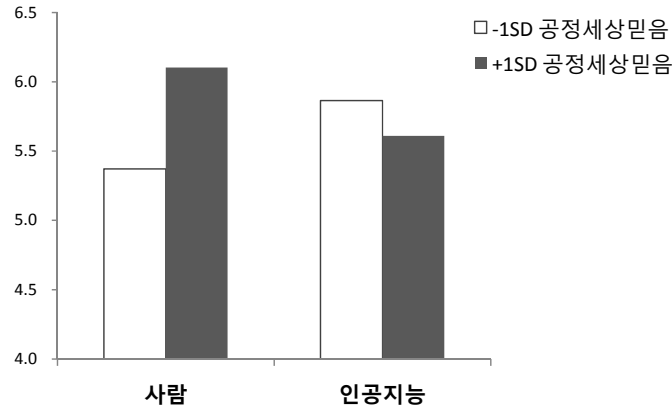


그림 5. 합격조건에서의 결과수용성에 대한 공정세상믿음의 조절효과

논 의

결과논의

본 연구의 목적은 두 가지이다. 첫째 공정 세상 믿음의 효과를 행위주체 유형에 대한 직접 비교 선택방식으로 검증하고 (연구1), 둘째 인공지능기술 수용성에 대한 공정세상 믿음의 효과를 인사선발 맥락에서 개체 간 설계 실험을 통해 검증하였다(연구2).

첫째 연구1에서 가설은 모두 지지되었다. 공정세상 믿음은 (사람과 비교한 상대적인) 인공지능의 공정성을 부적으로 예측하였고(가설 2), 공정세상 믿음이 약한 사람일수록 인공지능의 공정성을 매개로 인공지능의 직업 유능성을 높게 평가하여 매개모형 가설을 지지하였다(가설1, 가설3). 연구1은 사람과 인공지능 사이에서 직접 선택을 하도록 하는 측정방식을 사용하여, 공정세상 믿음이 약한 사람일수록 사람보다는 인공지능을 더 공정하게 인식함을 직접적으로 보여준다. 이는 기존의 체제를 정당한 것으로 여기고 사회변화를 받아들이는데 소극적인 공정세상 믿음의 체제정당화 기저와

일치된 결과이다(Friesen et al., 2019). 공정세상 믿음의 체제정당화 기저에 기반해 살펴보면, 참여자들이 인공지능을 기존의 체제인 사람에 대한 대안적인 체제로 인식하고 있음을 보여주는 시작점이 되는 연구라고 할 수 있을 것이다. 추가적으로, 직업수행 시 공정성이 중요한 직업군에서는 인공지능 역시 공정성이 높을수록 해당 직업 수행의 유능성을 예측하였다. 이는 CASA 패러다임이 유의함을 간접적으로 확인한 결과로 해석할 수 있을 것이다.

둘째 연구2에서 가설은 부분적으로 지지되었다. 저자들은 선발결과 귀인으로 인한 조절 효과로, 공정세상 믿음의 효과가 합격조건에서 보다 불합격조건에서 더 강하게 나타날 것으로 예상하였다. 분석결과 합격조건에서는 결과수용성에 대한 공정세상 믿음의 효과가 유의하여, 공정세상 믿음이 강한 사람은 인공지능 조건에서 보다 사람조건에서 높은 결과수용성을 나타냈고, 공정세상 믿음이 약한 사람은 인공지능 조건에서 더 높은 결과수용성을 나타냈다. 선발결과 귀인이론(Ployhart & Harold, 2004)에 기반하여 불합격의 경우에 행위주체 효과가 더 도드라질 것으로 예상하였으나, 예

상과 달리 불합격 조건에서는 행위주체 유형에 따른 효과차이가 나타나지 않았다. 공정세상믿음과 선발결과 간의 조절효과에 대해 살펴본 연구가 많지 않기 때문에 이 결과에 대해서는 단정적인 해석이 어렵다. 다만 연구2가 시나리오기반의 실험이었기에 불합격 조건에서 참여자들의 정서적 반응이 크게 일어나기에는 현실감이 적었을 가능성을 배제할 수 없다. 조금 더 현실감 있는 동영상이나 실험실 연구를 통해 재현 확인할 필요가 있다.

본 연구에서 확인한 공정세상믿음의 효과는 선행연구들의 결과가 서로 다르게 나타나는 이유를 설명하는 하나의 요인일 수 있다. 인간과 인공지능 행위주체에 대한 사용자 반응을 비교한 연구의 결과는 일관적이지 않다. 인공지능 기자와 사람 기자 조건을 비교한 네델란드의 연구에서는 일반 뉴스소비자는 사람과 인공지능 기자 간에 신뢰도 차이를 보이지 않았으나, 전문기자들은 인공지능 기자보다 사람의 기사를 더 전문적이라고 평가하였다(Van der Kaa, & Kraemer, 2014). 그러나 한국에서 실행한 유사한 실험에서는 일반 뉴스소비자 뿐 아니라 전문기자들도 인공지능 기사를 더 높게 평가하였다(Jung et al., 2017). 연구자들은 결과차이에 대해 한국에서 기자에 대해 신뢰도가 낮고, 공정성에 대한 불신이 커 새로운 정보화기술에 대한 욕구가 높기 때문인 것으로 설명을 시도하였다. 기자에 대한 낮은 신뢰나 공정성으로 인해 인공지능 기사를 높게 평가하게 되었다면, 이는 본 연구의 결과 즉 세상이 공정하다는 믿음이 낮을수록 인공지능 행위주체를 선호하는 결과와 일치한다.

인공지능을 기반으로 자동화된 면접절차와 사람이 진행하는 전통적인 면접절차를 비교한 연구에서도 독일, 중국, 한국의 결과에서 이처

럼 국가별로 상이한 경향이 발견되었다. 독일의 대학생을 대상으로 한 연구(Langer et al., 2019)에서는 입사면접의 경우 지원자는 버추얼 행위주체가 진행하는 인공지능기반 면접보다 사람이 진행하는 비디오면접을 더 공정하다고 인식하였고 더 높은 수용성을 보였다. 한편 중국의 비영리조직의 직원들을 대상으로 한 연구(Suen et al., 2019)에서는 행위주체 유형(사람 vs. 인공지능)에 따라 공정성 인식이 달라지지 않았다. 한국인 대학생을 대상으로 진행한 연구에서는 오히려 사람보다 인공지능 행위주체가 진행한 절차를 더 공정하다고 인식하였다(Min et al., 2018). 공정성에 대한 문화차 연구에 따르면 국가별로 현실의 공정성에 대한 인식수준이 다르고, 직업관련 결과변인에 대한 공정성 인식이 미치는 효과크기도 공정성의 유형에 따라 차이가 있다(Kim & Leung, 2007). Kim & Leung의 연구에서 분배공정성, 절차공정성, 상호작용공정성 및 조직공정성 모두에서 중국참가자나 한국참가자들보다 미국참가자들이 높은 평균을 기록하였다. 두 연구를 연결하여 고찰하여보면, 조직 공정성 인식이 낮은 국가일수록 사람보다 인공지능 행위주체를 공정하게 인식하는 경향을 발견할 수 있다. 인공지능 면접관 연구의 미국, 중국, 한국의 결과차이가 사회의 공정성에 대한 믿음의 차이로 인한 것인지, 즉 행위주체 공정성에 대한 인식차이가 본 연구에서 제시한 인공지능 수용성에 대한 공정세상믿음의 효과에 의한 것인지 국가 간 비교연구를 통해 검증해 볼 필요가 있다.

본 연구에서는 공정세상믿음이 인공지능 수용성에 미치는 효과를 체제정당화기저를 통해 설명하였다. 공정세상믿음이 인공지능의 직업유능성을 예측하는 기저를 인공지능의 공정성

인식을 매개변인으로 하여 설명하였다. 추가적으로 공정세상민음의 효과에 영향을 미치는 요인으로 기술에 대한 인식이나 신뢰도를 생각해볼 수 있다. 사람들은 현재 체제를 벗어날 수 없거나 변화가 불가능하다고 느끼면 체제정당화 경향이 강화되고, 반대로 변화에 대한 효능감이 상승하면 체제정당화 경향이 약화된다. 예를 들어 살고 있는 국가에서 해외로 나가기 어렵다고 느끼면 살고 있는 국가의 불평등을 옹호하고(Laurin et al., 2010), 사회시스템이 안정되어서 변화하지 않는다고 느끼면 현존하는 불평등을 지속시키는 정책을 지지한다(Chernyak-Hai et al., 2014). 그러나 반대로 대안에 대한 효능감이 있는 경우에는 공정세상민음이 기존체제를 정당화하지 않고 변화행동으로 이어지며 오히려 공정세상민음이 강할수록 변화행동이 강화되기도 한다(Beierlein, 2011; White et al., 2012). 이 관점에서, 체제정당화 기저를 약화시키고 인공지능 수용성을 높일 수 있는 요인으로 기술에 대한 신뢰나 구현가능성에 대한 믿음 등을 제안할 수 있다. 인공지능 기술이 현실화 될 가능성이 낮다고 생각하거나, 실제 인공지능 기술을 사용한 사용자 경험이 부족하거나 인공지능 기술에 대한 부정적인 인식을 가지고 있다면 인공지능은 사람의 대안체제가 될 수 없다고 느낄 것이다. 이 경우 체제정당화 기저는 강화될 것으로 예상이 되므로, 공정세상민음이 낮은 경우에도 인공지능 수용성이 높아지지 않을 수 있다. 앞서 살펴본 연구들의 국가 간 결과차이는 부분적으로 기술수용성이나 구현가능성의 인식 차이에서 기인한 것일 수도 있다. 한국은 기술수용성이 매우 높은 국가로 평가되고 있는데, 이것이 공정세상민음의 효과를 강화하는 요인이 되었을 수도 있다.

한계 및 향후 연구방향

본 연구는 시스템 정당화 이론의 관점에서 인공지능과 인간의 대체적인 관계를 조사하기 위해 슬라이딩 척도를 사용하여 직접비교 선택하도록 하는 설문(연구1)과 개체 간 설계(연구2)를 사용하여 검증하였다. 그러나 연구1에서는 양 행위주체 사이에서 선택을 하게 되는 슬라이딩척도의 특성상 모든 문항이 단문항으로 구성되었고, 무엇보다 공정세상민음 역시 단문항으로 측정되어 신뢰도를 제시할 수 없는 심리측정상의 취약함을 가지고 있다. 연구2에서는 다문항 공정세상민음 척도를 사용하여 연구1의 한계점을 보완하였다. 그러나 연구2에서는 수용성이 단문항으로 측정된 한계가 있어, 후속연구에서는 다문항 척도를 사용하여 심리측정상의 신뢰도를 높일 필요가 있다. 추가적으로 연구2에서는 시나리오 자극을 사용하여 선발결과의 효과가 작게 나타났을 가능성이 있다. 향후 좀 더 현실감 있는 동영상 자극이나 실험실 연구를 통해 재현하여 확인할 필요가 있다.

또한 본 연구자들은 시나리오 상에서 사람 조건과 인공지능 조건 간에 행위주체 유형 이외의 오염요인을 통제하려고 노력하였고, 사진자극을 삽입하여 인공지능의 형태에 대해서도 동일하게 통제하려고 노력을 하였다. 그러나 시나리오 자극의 특성 상 사람마다 서로 다른 형태의 인공지능을 상상했을 가능성을 완전히 배제하기 어렵다. 로봇의 외모는 얼마나 생물이나 사람과 비슷한가에 따라서 인간과 로봇과의 상호작용이나 인식에 크게 영향을 미치므로(Friedman et al., 2003; Goetz et al., 2003; Tung, 2016), 추후 연구에서는 참여자들이 개념화하는 인공지능의 형태를 완전히 동

일하게 통제할 필요가 있다.

함의

세상이 공정하지 않다고 인식할수록 인간보다 인공지능을 공정하고 유능하다고 인식한 본 연구의 결과에서 4차 산업혁명 시대 직업 불안정성의 아이러니를 엿볼 수 있다. 사람들은 인공지능이 기존에 인간이 하던 일을 대체하고(Frey & Osborne, 2017) 인간의 직업안정성을 떨어뜨리는 위협요소가 될 것으로 우려한다. 기업이 수익성과 효율성을 추구하기 때문에 인공지능으로 사람을 대체할 것이라고 우려한다. 그러나 본 연구의 결과는, 세상에 대한 불신이 인간의 입지를 스스로 축소시키는 결과에 이를 수 있음을 함의한다. 현실이 불공정하면 그 대안으로 인간 대신 인공지능을 선호하고 인공지능을 인간보다 공정한 존재로 인식하는 프레임이나 휴리스틱(Gamliel, & Peer, 2009)을 가지게 될 수 있다. 기대이론이나 휴리스틱 이론 관점에서 보면, 인간에 대한 대안으로 인공지능을 도입하는 것을 오히려 인간 스스로 정의롭다고 받아들이고, 결과적으로 인간 스스로를 직업영역에서 몰아내는 역할을 하게 될 수 있을 것이다.

기술에 대한 신뢰의 맹점도 있다. 기술발전으로 인해 업무현장과 사회전반에 인공지능과 자동화 시스템이 꾸준히 도입되고 있다. 이에 더해 코로나19로 인한 비대면 상호작용에 대한 요구가 더욱 크게 증가하고 있어, 이러한 추세는 더욱 강화될 것으로 예상된다. 이런 변화는 효율적이고 발전적인 것으로 인식되고 있지만 실제로 기술이 공정성을 담보하고 윤리적인 통제 하에 있는지, 개별영역에서 어떻게 수용되는 지, 사회문화적으로 어떤 영향을

미치는 지에 대한 실증적인 연구는 아직 부족하다. 본 연구의 인공지능의 역량에 대한 답변을 보면 사람들은 이미 인공지능의 역량에 대해 높은 기대를 가지고 있다. 특히 공정성에 대한 요구가 높은 판사와 기자 직업에 있어서는 인공지능이 사람보다 뛰어날 것이라는 기대가 크게 나타났다.

그러나 과연 알고리즘은 사람들이 기대하는 공정성과 객관성을 담보하는가? 알고리즘과 데이터에 편향이 존재하고, 그로 인해 의도치 않은 차별이 발생하는 사례가 보고되고 있다. 예를 들어 아마존에서는 인공지능 채용시스템을 도입하였는데, 여성에 대한 차별적인 채용 결과가 문제가 되어 폐기하는 사건이 발생하였다(Dastin, 2018). 미국에서 가석방 판단 시 사용되는 예측프로그램에서는 인종에 따라 1종 오류와 2종 오류가 다르게 나타난다(Angwin et al., 2016). 다층의 기계학습으로 인해 인공지능은 ‘사람처럼’ 보이는 복잡한 상호작용을 할 수 있게 되었지만, 동시에 인공지능의 의사결정 과정은 블랙박스가 되어 문제가 어디에 기인하는지 찾아내는 것도 쉽지 않다. 인공지능 기반의 시스템에는 편향의 위험이 실재하므로(Mittelstadt et al., 2016), 인공지능 시스템을 도입할 때에는 편향가능성과 무결성에 대해 면밀하고 다양한 관점에서의 검토가 필요하며 도입 이후에도 지속적인 모니터링이 필요하다.

인공지능 도입의 영향은 인공지능이 사용되는 상황맥락, 개인적 특성, 문화차 등의 다양한 요인에 의해 이질적으로 나타날 수 있다. 이를 긍정적으로 수용하는 사람도 있겠지만, 누군가는 수용을 못하거나 불공정하게 받아들일 수도 있다. 인공지능과 기술도입의 추세가 사회통합과 발맞추어 진행되기 위해서는, 실

제 기술의 설명가능성과 비편향성을 추구해야 함은 물론 사회전반의 기술에 대한 이해를 높이는 노력이 동시에 필요하다. 사용자의 인식과 수용성을 개선할 수 있는 요인을 연구하여 시스템에 반영할 필요가 있다.

참고문헌

- Ababneh, K. I., Hackett, R. D., & Schat, A. C. (2014). The role of attributions and fairness in understanding job applicant reactions to selection procedures and decisions. *Journal of Business and Psychology, 29*(1), 111-129. <https://doi.org/10.1007/s10869-013-9304-y>.
- Acemoglu, D., & Restrepo, P. (2017). Robots and jobs: Evidence from US labor markets. *NBER working paper 23285*. <https://doi.org/10.3386/w23285>.
- Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. sage.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (May 2016). *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*. Propublica. Retrieved October 07, 2020, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Anseel, F., & Lievens, F. (2009). The mediating role of feedback acceptance in the relationship between feedback and attitudinal and performance outcomes. *International Journal of Selection and Assessment, 17*(4), 362-376. <https://doi.org/10.1111/j.1468-2389.2009.00479.x>
- Assaker, G. (2020). Age and gender differences in online travel reviews and user-generated-content (UGC) adoption: extending the technology acceptance model (TAM) with credibility theory. *Journal of Hospitality Marketing & Management, 29*(4), 428-449. <https://doi.org/10.1080/19368623.2019.1653807>
- Baert, S. (2018). Facebook profile picture appearance affects recruiters' first hiring decisions. *New Media & Society, 20*(3), 1220-1239. <https://doi.org/10.1177/1461444816687294>
- Bauer, T. N., Maertz Jr, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of applied Psychology, 83*(6), 892-903. <https://doi.org/10.1037/0021-9010.83.6.892>
- Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology, 24*(7), 738-743. <https://doi.org/10.1016/j.cub.2014.02.009>
- Beierlein, C., Werner, C. S., Preiser, S., & Wermuth, S. (2011). Are just-world beliefs compatible with justifying inequality? Collective political efficacy as a moderator. *Social Justice Research, 24*(3), 278-296. <https://doi.org/10.1007/s11211-011-0139-2>
- Bell, B. S., Wiechmann, D., & Ryan, A. M. (2006). Consequences of organizational justice expectations in a selection system. *Journal of Applied Psychology, 91*(2), 455-466. <https://doi.org/10.1037/0021-9010.91.2.455>.

- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21-34. <https://doi.org/10.1037/0021-9010.91.2.455>.
- Bóo, F. L., Rossi, M. A., & Urzúa, S. S. (2013). The labor market return to an attractive face: Evidence from a field experiment. *Economics Letters*, 118(1), 170-172. <https://doi.org/10.1016/j.econlet.2012.10.016>.
- Chernyak-Hai, L., Halabi, S., & Nadler, A. (2014). "Justified dependency": Effects of perceived stability of social hierarchy and level of system justification on help-seeking behavior of low-status group members. *Group Processes & Intergroup Relations*, 17(4), 420-435. <https://doi.org/10.1177/1368430213507320>.
- Choi, Y. (2019). Comparing Probability of Computerisation of Korean Jobs: An Empirical Analysis with KNOW Data. [Master's thesis, Seoul National University. Research Information Sharing Service.
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved August 18, 2020, from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G?fbclid=IwAR3dWIANNaEuHYRuX41UOIJeAgD7Dr2Fa_5fkCSwQ4NygjoHre10p5WIMVo.
- Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*, 33, 372-376. <https://doi.org/10.1016/j.chb.2013.08.013>.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation?. *Technological forecasting and social change*, 114, 254-280. <https://doi.org/10.1016/j.techfore.2016.08.019>.
- Friedman, B., Kahn Jr, P. H., & Hagman, J. (2003, April). Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 273-280). ACM. <https://doi.org/10.1145/642611.642660>.
- Friesen, J. P., Laurin, K., Shepherd, S., Gaucher, D., & Kay, A. C. (2019). System justification: Experimental evidence, its contextual nature, and implications for social change. *British Journal of Social Psychology*, 58(2), 315-339. <https://doi.org/10.1111/bjso.12278>.
- Gamliel, E., & Peer, E. (2009). Effect of framing on applicants' reactions to personnel selection methods. *International Journal of Selection and Assessment*, 17(3), 282-289.
- Gessl, A. S., Schlögl, S., & Mevenkamp, N. (2019). On the perceptions and acceptance of artificially intelligent robotics and the psychology of the future elderly. *Behaviour & Information Technology*, 38(11), 1068-1087. <https://doi.org/10.1080/0144929x.2019.1566499>.
- Georgia State University (2018). *Artificial intelligence gets its day in court*. Retrieved December 12, 2018, from <<https://phys.org/news/2018-03-artificial-intelligence-day-court.html>>
- Goetz, J., Kiesler, S., & Powers, A. (2003, October). Matching robot appearance and

- behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on* (pp. 55-60). IEEE.
<https://doi.org/10.1109/roman.2003.1251796>.
- Hafer, C. L., & Sutton, R. (2016). Belief in a just world. In *Handbook of social justice theory and research* (pp. 145-160). Springer.
- Jost, J. T., & Hunyady, O. (2005). Antecedents and consequences of system-justifying ideologies. *Current directions in psychological science, 14*(5), 260-265.
<https://doi.org/10.1111/j.0963-7214.2005.00377.x>.
- Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior, 71*, 291-298.
<https://doi.org/10.1016/j.chb.2017.02.022>.
- Khosla, R., Chu, M., & Nguyen, K. (2016). Human-Robot Interaction Modelling for Recruitment and Retention of Employees. *HCI in Business, Government, and Organizations: Information Systems Lecture Notes in Computer Science, 302-312*.
https://doi.org/10.1007/978-3-319-39399-5_29.
- Kim, E., Kim, D., Park, H., Kim, S., & Kim, J. (2017). Validation of the Korean version of the Belief in a Just World Scale (K-BJWS). *The Korean Journal of Counseling and Psychotherapy, 29*(3), 689-710.
<https://doi.org/10.23844/kjcp.2017.08.29.3.689>.
- Kim, T. Y., & Leung, K. (2007). Forming and reacting to overall fairness: A cross-cultural comparison. *Organizational Behavior and Human Decision Processes, 104*(1), 83-95.
<https://doi.org/10.1016/j.obhdp.2007.01.004>.
- Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology, 8*(1), 3-25.
<https://doi.org/10.1007/bf02230391>.
- Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R. M. Q., & MacDonald, B. A. (2009, September). Age and gender factors in user acceptance of healthcare robots. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 214-219). IEEE.
<https://doi.org/10.1109/roman.2009.5326292>.
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment, 27*(3), 217-234.
<https://doi.org/10.1111/ijsa.12246>.
- Laurin, K., Shepherd, S., & Kay, A. C. (2010). Restricted emigration, system inescapability, and defense of the status quo: System-justifying consequences of restricted exit opportunities. *Psychological Science, 21*(8), 1075-1082.
<https://doi.org/10.1177/0956797610375448>.
- Lee, C. (2017). *The quantitative analysis between the occupational interests, values, knowledges, skills and the occupations'computerisable probabilities due to the*

- fourth industrial revolution*. [Master's thesis, The Cyber University of Korea]. Research Information Sharing Service.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1-16. <https://doi.org/10.1177/2053951718756684>.
- Lerner, M. J. (1980). The belief in a just world. In *The Belief in a just World* (pp. 9-30). Springer.
- Lind, E. A. (2001). Advances in organizational justice. In J. Greenberg & R. Cropanzano (Eds.). *Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations* (pp. 56-88). Stanford University Press.
- Lucas, T., Zhdanova, L., & Alexander, S. (2011). Procedural and distributive beliefs for self and others: Assessment of a four-factor individual differences model. *Journal of Individual Differences*, 32(1), 14-25. <https://doi.org/10.1027/1614-0001/a000032>.
- Makkonen, M., Frank, L., & Koivisto, K. (2017). Age differences in technology readiness and its effects on information system acceptance and use: the case of online electricity services in Finland. Bled 2017: *Proceedings of the 30th Bled eConference. Digital Transformation: From Connecting Things to Transforming Our Lives*, ISBN 978-961-286-043-1. <https://doi.org/10.18690/978-961-286-043-1.28>.
- Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 117-124). IEEE. <https://doi.org/10.1145/2696454.2696458>.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21. <https://doi.org/10.1177/2053951716679679>.
- Min, J., Kim, S., Park, Y., & Sohn, Y. W. (2018). A Comparative Study of Potential Job Candidates' Perceptions of an AI Recruiter and a Human Recruiter. *Journal of the Korea Convergence Society*, 9(5), 191-202. <https://doi.org/10.15207/JKCS.2018.9.5.191>.
- Morris, M. G., & Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing work force. *Personnel psychology*, 53(2), 375-403. <https://doi.org/10.1111/j.1744-6570.2000.tb00206.x>.
- Nam, T. (2019). Technology usage, expected job sustainability, and perceived job insecurity. *Technological Forecasting and Social Change*, 138, 155-165. <https://doi.org/10.1016/j.techfore.2018.08.017>.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1), 81-103. <https://doi.org/10.1111/0022-4537.00153>.
- Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in Human-Machine Interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, 89,

- 27-39.
<https://doi.org/10.1016/j.chb.2018.07.022>.
- Park, J., & Jung, Y. (2020). The effects of perceived Artificial Intelligence's competence on employee's job insecurity and work cynicism: Moderated mediation by work meaningfulness. *Korean Journal of Industrial and Organizational Psychology*, 33(2), 145-175.
<https://doi.org/10.24230/kjiop.v33i2.145-175>.
- Parkin, S. (2016). *The artificially intelligent doctor will hear you now* - MIT Technology Review. Retrieved December 12, 2018, from <<https://www.technologyreview.com/s/600868/the-artificially-intelligent-doctor-will-hear-you-now/>>.
- Peters, M. A. (2017). Deep Learning, the final stage of automation and the end of work (Again)? *Psychosociological Issues in Human Resource Management*, 5(2), 154-168.
- Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior*, 65, 23-30.
<https://doi.org/10.22381/pihrm5220176>.
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.
- Ployhart, R. E., & Harold, C. M. (2004). The applicant attribution reaction theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, 12(1 2), 84-98.
<https://doi.org/10.1111/j.0965-075x.2004.00266.x>.
- Schermerhorn, P., Scheutz, M., & Crowell, C. R. (2008, March). Robot social presence and gender: Do females view robots differently than males?. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 263-270).
<https://doi.org/10.1145/1349822.1349857>.
- Schroth, H. A., & Pradhan Shah, P. (2000). Procedures: Do we really want to know them? An examination of the effects of procedural justice on self-esteem. *Journal of Applied Psychology*, 85(3), 462-471.
<https://doi.org/10.1037/0021-9010.85.3.462>.
- Stone, D. L., Deadrick, D. L., Lukaszewski, K. M., & Johnson, R. (2015). The influence of technology on the future of human resource management. *Human Resource Management Review*, 25(2), 216-231.
<https://doi.org/10.1016/j.hrmr.2015.01.002>.
- Suen, H. Y., Chen, M. Y. C., & Lu, S. H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes?. *Computers in Human Behavior*, 98, 93-101.
<https://doi.org/10.1016/j.chb.2019.04.012>.
- Sundar, S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. Metzger & A. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73-100). MIT Press.
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1), 52-72.
<https://doi.org/10.1111/j.1460-2466.2001.tb02872.x>.
- Tung, F. W. (2016). Child perception of humanoid

- robot appearance and behavior. *International Journal of Human-Computer Interaction*, 32(6), 493-502.
<https://doi.org/10.1080/10447318.2016.1172808>.
- Van der Kaa, H. A. J., & Krahmer, E. J. (2014). Journalist versus news consumer: The perceived credibility of machine written news. In *Proceedings of the computational Journalism conference*. NY.
- Voiklis, J., Kim, B., Cusimano, C., & Malle, B. F. (2016, August). Moral judgments of human vs. robot agents. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on* (pp. 775-780). IEEE.
<https://doi.org/10.1109/roman.2016.7745207>.
- White, K., MacDonnell, R., & Ellard, J. H. (2012). Belief in a just world: Consumer intentions and behaviors toward ethical products. *Journal of Marketing*, 76(1), 103-118.
<https://doi.org/10.1509/jm.09.0581>.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036-1040.
<https://doi.org/10.1073/pnas.1418680112>.
- Zielinski, D. (2017, February 13). Recruiting Gets Smart Thanks to Artificial Intelligence. Retrieved September 20, 2017, from <https://www.shrm.org/resourcesandtools/hrtopics/technology/pages/recruiting-gets-smart-thanks-to-artificial-intelligence.aspx>

1차원고접수 : 2020. 10. 15.

2차원고접수 : 2020. 12. 28.

최종게재결정 : 2020. 12. 29.

Special issue: Psychology and Fourth Industrial Revolution 2

The effect of Belief in a Just World on the acceptance of AI technology

Sinae Kim

Young Woo Sohn

Department of Psychology, Yonsei University

With the development of artificial intelligence (AI) technology, AI has been introduced in various jobs to replace people. This study viewed AI as an alternative to the current human-led system and examined the effect of Belief in a Just World (BJW) on AI acceptance based on the system justification theory. We expected that participants with stronger BJW would prefer the current human-led system, and those with weaker BJW would be more likely to accept the AI-based system. Study 1 examined the effect of BJW on the perception of job competence by making the participants choose between human and AI. Study 2 examined the effect of BJW on the selection outcome acceptance in a 2 (Human vs. AI) X 2 (accepted vs. rejected) between-subjects design. Results showed that BJW predicted higher competence perception of AI, mediated by higher fairness perception of AI (Study1), and those with weaker BJW showed higher acceptance of selection results based on the AI system in the accepted condition, but not in the rejected condition (Study 2). Based on our findings, we discussed the factors affecting AI acceptance, limitations of the present study, and suggestions for future research.

Key words : HCI, Just World Belief, System Justification, Artificial intelligence technology acceptance, AI interview