Why study explanations?

Philosophical Foundations of Explanation

Dean McHugh

Institute of Logic, Language and Computation
University of Amsterdam

January 10, 2024

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There's more to causality than language
3. Causation in the history of philosophy

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There’s more to causality than language
3. Causation in the history of philosophy

Causation and explanation

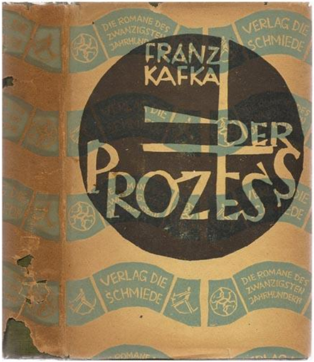
Two questions

1. Is an explanation correct?
   → Study the meaning of *cause* and *because*
2. Is an explanation useful?
   → Study causal selection
   (see e.g. Franklin-Hall 2015, Icard, Kominsky, and Knobe 2017)

For an overview of the relation between causation and explanation see Lipton (2009).

Franz Kafka's *The Trial* (1915)

Figure: Wolfgang Letti, *The Trial* (1981)


DIE ROMANE DES ZWANZIGSTEN JAHRHUNDERTS
VERLAG DIE SCHMIEDE
FRANZ KAFKA
DER PROZESS
VERLAG DIE SCHMIEDE
DIE ROMANE DES ZWANZIGSTEN JAHRHUNDERTS

The right to explanation

GDPR Article 13.2(f) affords a right to information about

> *"the existence of automated decision making and ... meaningful information about the logic involved".*

The right to explanation

GDPR Article 13.2(f) affords a right to information about

> *"the existence of automated decision making and ... meaningful information about the logic involved".*

Kate Vredenburgh (2022) argues for a right to explanation.

Figure: Kate Vredenburgh

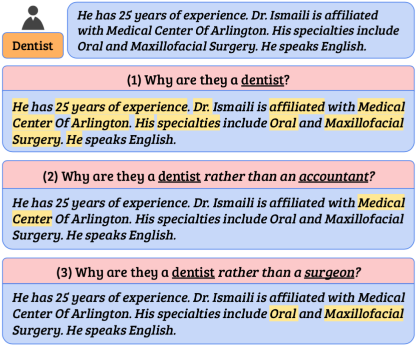
Explanation across domains

Philosophy. Schaffer (2005, 2010) proposes that causation is a contrastive relation:

c rather than c' caused e rather than e'.

Linguistics. Dretske (1972) and Beaver and Clark (2008).

Artificial intelligence. Mittelstadt, Russell, and Wachter (2019).



Dentist

He has 25 years of experience. Dr. Ismaili is affiliated with Medical Center Of Arlington. His specialties include Oral and Maxillofacial Surgery. He speaks English.

(1) Why are they a dentist?

He has 25 years of experience. Dr. Ismaili is affiliated with Medical Center Of Arlington. His specialties include Oral and Maxillofacial Surgery. He speaks English.

(2) Why are they a dentist *rather than an* accountant?

He has 25 years of experience. Dr. Ismaili is affiliated with Medical Center Of Arlington. His specialties include Oral and Maxillofacial Surgery. He speaks English.

(3) Why are they a dentist *rather than a* surgeon?

He has 25 years of experience. Dr. Ismaili is affiliated with Medical Center Of Arlington. His specialties include Oral and Maxillofacial Surgery. He speaks English.

Figure: From Jacovi et al. (2021)

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There’s more to causality than language
3. Causation in the history of philosophy

Causal theories of knowledge

Goldman (1967), ‘A causal theory of knowing’.

a knows that *p*

just in case

what makes *p* true caused *a* to believe *p*.

Causal theories of knowledge

Goldman (1967), ‘A causal theory of knowing’.

a knows that *p*

just in case

what makes *p* true caused *a* to believe *p*.

Gettier case (Gettier 1963):

> *Across the street, Madonna is hiding behind a cardboard cutout of Madonna. Alice sees the cardboard cutout and concludes that Madonna is across the street.*

Causal theories of reference

Geach (1969), Donnellan (1970), and Kripke (1972):

- An initial baptism takes place
- Beginning with this baptism, the name is passed ‘from link to link’.

Causal theories of reference

Geach (1969), Donnellan (1970), and Kripke (1972):

- An initial baptism takes place
- Beginning with this baptism, the name is passed ‘from link to link’.

Case from Kripke (1972, p. 96):

“I hear the name ‘Napoleon’ and decide it would be a nice name for my pet aardvark.”

“Napoleon is hungry today.”

Figure: Napoleon

For criticism of causal theories of reference see Evans (1973) and Dickie (2015).

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There's more to causality than language
3. Causation in the history of philosophy

Discrimination law

> *It shall be an unlawful employment practice for an employee to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin.*
>
> — Civil Rights Act (1964), section 703(a)(1) [link]

See Murray and Eastwood (1965) for a good analysis of the relationship between sex discrimination and the Civil Rights Act.

Tort law

Examples

- one commits simple battery only if one "intentionally causes bodily contact" with another;
- trespasses only if one "intentionally enters or causes tangible entry upon the land in possession of another"
- is liable for negligence only if one "causes personal injury or property damage" to another.

(Harvard Law Review 2017)

- Hart and Honoré (1959) distinguish *causation in fact* and *legal causation*
- Two tests for causation in fact:
  1. but for test
  2. substantive factor test

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There's more to causality than language
3. Causation in the history of philosophy

Moral responsibility

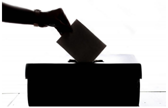
Braham and van Hees (2012): being morally responsible for an event requires causally contributing to it.

Moral responsibility

Braham and van Hees (2012): being morally responsible for an event requires causally contributing to it.

Example (Counting votes)

- A committee with two members: Chair and CEO
- New policies require both to vote Yes to pass
- The CEO's envelope is opened first: it says 'No'.
- The proposal is declared failed without ever opening the Chair's envelope.
- The proposal's failure bankrupts the company.



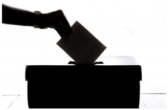
1. Did the Chair's vote causally contribute to the outcome?

Moral responsibility

Braham and van Hees (2012): being morally responsible for an event requires causally contributing to it.

Example (Counting votes)

- A committee with two members: Chair and CEO
- New policies require both to vote Yes to pass
- The CEO's envelope is opened first: it says 'No'.
- The proposal is declared failed without ever opening the Chair's envelope.
- The proposal's failure bankrupts the company.



1. Did the Chair's vote causally contribute to the outcome?
2. Is the Chair morally responsible for the outcome?

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There's more to causality than language
3. Causation in the history of philosophy

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There’s more to causality than language
3. Causation in the history of philosophy

Formal semantics

1. Designate a language, such as
   - A (fragment of) natural language
   - A formal language
2. Define a model
3. Interpret the sentences of the language in that model

Goal of formal semantics

Find a translation from natural language into a formal language such that, for every sentence S of the language, S is intuitively true at world w just in case $[\![S]\!]^M = 1$ in the model that represents w.

Why focus on language?

Natural languages are

- expressive
- compositional

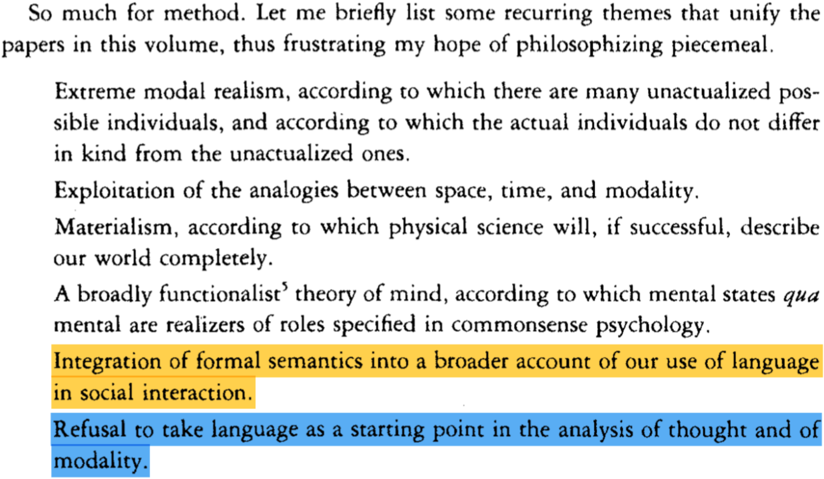
The principle of compositionality

The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined. (Partee 1984)

Figure: Barbara Partee

(1) a. Socrates died because he drank poison.
 b. Socrates only died because he drank poison.

David Lewis's philosophical method



So much for method. Let me briefly list some recurring themes that unify the papers in this volume, thus frustrating my hope of philosophizing piecemeal.

Extreme modal realism, according to which there are many unactualized possible individuals, and according to which the actual individuals do not differ in kind from the unactualized ones.

Exploitation of the analogies between space, time, and modality.

Materialism, according to which physical science will, if successful, describe our world completely.

A broadly functionalist[5] theory of mind, according to which mental states *qua* mental are realizers of roles specified in commonsense psychology.

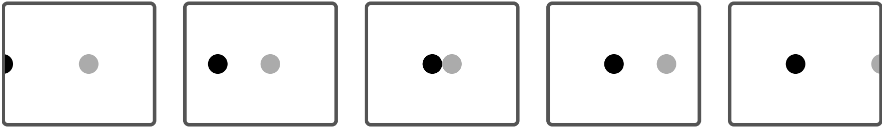
Integration of formal semantics into a broader account of our use of language in social interaction.

Refusal to take language as a starting point in the analysis of thought and of modality.

Figure: From Lewis (1983, p. xi)

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There’s more to causality than language
3. Causation in the history of philosophy



Causal reasoning in infants

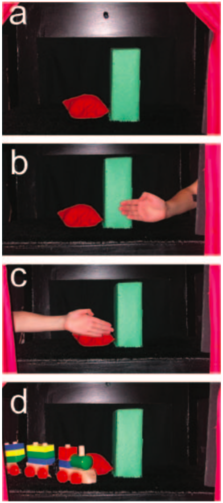


Figure: Experimental setup from Saxe, Tenenbaum, and Carey (2005)

Causal reasoning in infants

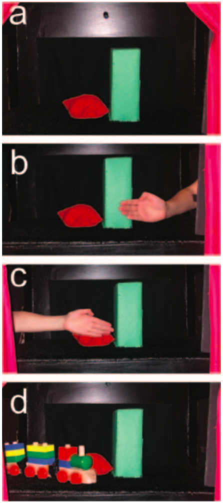


Figure: Experimental setup from Saxe, Tenenbaum, and Carey (2005)

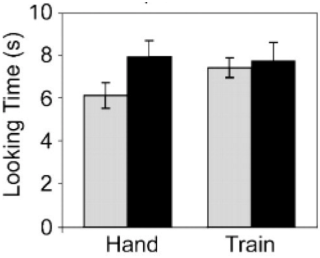


Figure: Object came from same side (gray) or opposite side (black).

Causal reasoning in infants

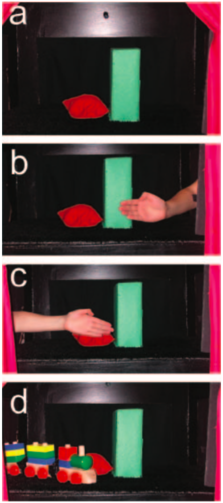


Figure: Experimental setup from Saxe, Tenenbaum, and Carey (2005)

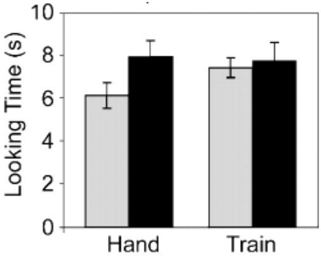
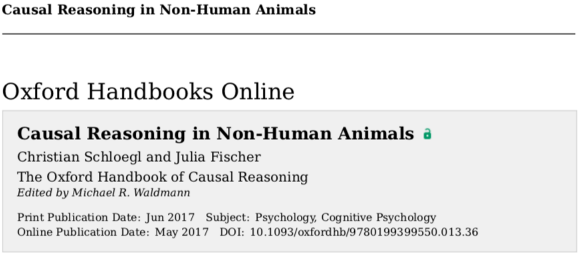


Figure: Object came from same side (gray) or opposite side (black).

"In the current study, infants made a distinction between causal agents and inanimate objects." (Saxe, Tenenbaum, and Carey 2005, p. 1000)

Causal reasoning without language



Causal Reasoning in Non-Human Animals

Oxford Handbooks Online

Causal Reasoning in Non-Human Animals

Christian Schloegl and Julia Fischer

The Oxford Handbook of Causal Reasoning

Edited by Michael R. Waldmann

Print Publication Date: Jun 2017 Subject: Psychology, Cognitive Psychology
Online Publication Date: May 2017 DOI: 10.1093/oxfordhb/9780199399550.013.36

Figure: Schloegl and Fischer (2017)

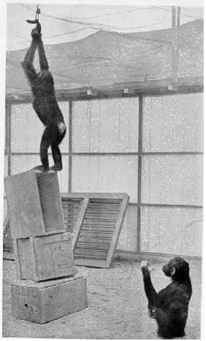


Figure: From Wolfgang Köhler *The Mentality of Apes* (1917)

Causal inference in non-human animals

- Causally implausible events elicit longer looking times in chimpanzees and bonobos (O'Connell and Dunbar 2005) and rooks (Bird and Emery 2010)

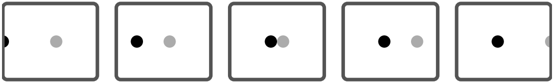


Figure: Michotte's experiment (Michotte 1963)

Causal inference in non-human animals

- Causally implausible events elicit longer looking times in chimpanzees and bonobos (O'Connell and Dunbar 2005) and rooks (Bird and Emery 2010)

- When searching for food in opaque bottles, champanzees learned to discriminate between lighter and heavier bottles faster than to discriminate between bottles of different colours (Hanus and Call 2011)

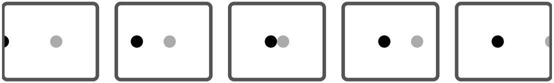


Figure: Michotte's experiment (Michotte 1963)

Causal inference in non-human animals

- Causally implausible events elicit longer looking times in chimpanzees and bonobos (O’Connell and Dunbar 2005) and rooks (Bird and Emery 2010)

- When searching for food in opaque bottles, champanzees learned to discriminate between lighter and heavier bottles faster than to discriminate between bottles of different colours (Hanus and Call 2011)
- Mascalzoni, Regolin, and Vallortigara (2010) showed freshly hatched chicks a launching experiment. The chicks preferred to follow the self-propelled object over the non-self-propelled object.

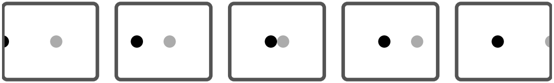
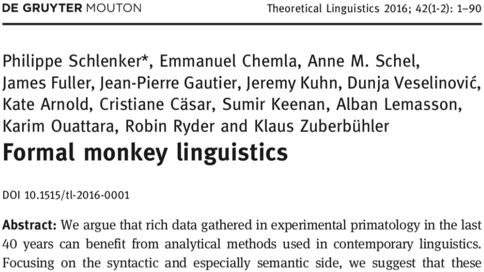


Figure: Michotte’s experiment (Michotte 1963)

...but monkeys seem to have some language

Figure: A putty nose monkey



DE GRUYTER MOUTON — Theoretical Linguistics 2016; 42(1-2): 1–90

Philippe Schlenker*, Emmanuel Chemla, Anne M. Schel, James Fuller, Jean-Pierre Gautier, Jeremy Kuhn, Dunja Veselinović, Kate Arnold, Cristiane Cäsar, Sumir Keenan, Alban Lemasson, Karim Ouattara, Robin Ryder and Klaus Zuberbühler

Formal monkey linguistics

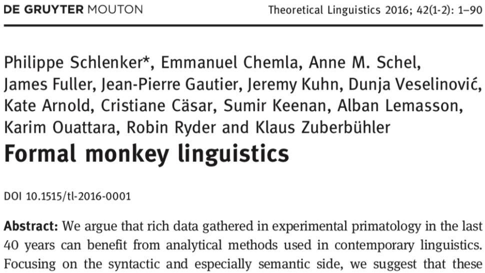
DOI 10.1515/tl-2016-0001

Abstract: We argue that rich data gathered in experimental primatology in the last 40 years can benefit from analytical methods used in contemporary linguistics. Focusing on the syntactic and especially semantic side, we suggest that these

Figure: Schlenker et al. (2016)

...but monkeys seem to have some language

Figure: A putty nose monkey



DE GRUYTER MOUTON Theoretical Linguistics 2016; 42(1-2): 1–90

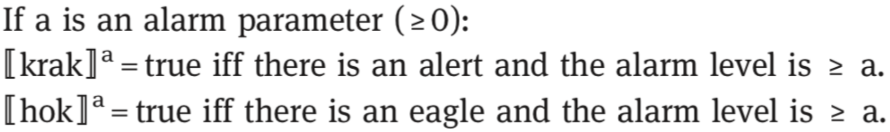
Philippe Schlenker*, Emmanuel Chemla, Anne M. Schel, James Fuller, Jean-Pierre Gautier, Jeremy Kuhn, Dunja Veselinović, Kate Arnold, Cristiane Cäsar, Sumir Keenan, Alban Lemasson, Karim Ouattara, Robin Ryder and Klaus Zuberbühler

Formal monkey linguistics

DOI 10.1515/tl-2016-0001

Abstract: We argue that rich data gathered in experimental primatology in the last 40 years can benefit from analytical methods used in contemporary linguistics. Focusing on the syntactic and especially semantic side, we suggest that these

Figure: Schlenker et al. (2016)



If a is an alarm parameter (≥ 0):

$[\![\text{krak}]\!]^a$ = true iff there is an alert and the alarm level is $\geq$ a.

$[\![\text{hok}]\!]^a$ = true iff there is an eagle and the alarm level is $\geq$ a.

Figure: From Schlenker et al. (2016, p. 17)

Some questions

- What do we talk about when we talk about causation?
- When we call something *causal* reasoning, how do we know it is causal?
- Do we always talk about something that can be expressed in language?

Plan

1. Motivations
   - Causation and explanation
   - Causation in philosophy
   - Causation in the law
   - Moral responsibility
2. Methods
   - Formal semantics
   - There's more to causality than language
3. Causation in the history of philosophy

Aristotle’s Four Causes

Physics II.3 (194b17-195a4), *Metaphysics* V.2

Four answers to the question ‘Why?’.
“we think we do not have knowledge of a thing until we have grasped its why, that is to say, its cause [*aitia*]” (*Physics* 194 b 17-20)

Material cause (hūlē, *nature*) ”that out of which” it is made. The material cause of Rodin’s *The Thinker* is bronze.

Efficient Cause (kinoûn, *cause* in its modern sense) The source of change.

Formal Cause (eîdos, *form* or *idea*) The essence of the object, the pattern the object exemplifies.

Final Cause (télos, *goal* or *end*) That for the sake of which a thing is done. The goal of a seed is to become a plant.

The Greek word *aitia* is often translated as ‘cause’, but is better translated as ‘explanation’.

Hume, *Enquiry* Section 7

“Of all the ideas that occur in metaphysics, none are more obscure and uncertain than those of power, force, energy or necessary connection”

“All our ideas are merely copies of our impressions, so it is impossible for us to think of anything that we haven’t previously felt through either our external or our internal senses.”

Hume, *Enquiry* Section 7

"When we look around us at external objects, and think about the operation of causes, we are *never* able to discover any power or necessary connection, any quality that ties the effect to the cause and *makes it an infallible consequence of* it. All we find is that the one event *does in fact follow* the other. The impact of one billiard-ball is accompanied by motion in the other. This is all that appears to the outer senses. The mind feels no sentiment or inward impression from this sequence of events: so in no single particular instance of cause and effect is there anything that can suggest the idea of power or necessary connection."

Hume, *Enquiry* Section 7

“Summing up, then: throughout the whole of nature there seems not to be a single instance of connection that is conceivable by us. All events seem to be entirely loose and separate. ”

“So the feeling or impression from which we derive our idea of power or necessary connection is a feeling of connection in the mind—a feeling that accompanies the imagination’s habitual move from observing one event to expecting another of the kind that usually follows it. That’s all there is to it”

Kant on Hume

Kant aims to rescue the
“a priori origin of the pure concepts of the understanding and the validity of the general laws of nature as laws of the understanding, in such a way that their use is limited only to experience, because their possibility has its ground merely in the relation of the understanding to experience, however, not in such a way that they are derived from experience, but that experience is derived from them, a completely reversed kind of connection which never occurred to Hume.”
(*Prolegomena to any future metaphysics*)

References I

- Beaver, David and Brady Clark (2008). *Sense and sensitivity: How focus determines meaning*. Vol. 12. Wiley. DOI: 10.1002/9781444304176.
- Bird, Christopher D and Nathan J Emery (2010). Rooks perceive support relations similar to six-month-old babies. *Proceedings of the Royal Society B: Biological Sciences* 277.1678, pp. 147–151. DOI: 10.1098/rspb.2009.1456.
- Braham, Matthew and Martin van Hees (2012). An Anatomy of Moral Responsibility. English. *Mind* 121.483, pp. 601–634. DOI: 10.1093/mind/fzs081.
- Dickie, Imogen (2015). *Fixing reference*. Oxford University Press.
- Donnellan, Keith S. (1970). Proper names and identifying descriptions. *Synthese* 21, pp. 335–358.
- Dretske, Fred (1972). Contrastive statements. *The Philosophical Review* 81.4, pp. 411–437. DOI: 10.2307/2183886.

References II

- Evans, Gareth (1973). The causal theory of names. *Proceedings of the Aristotelian Society* 47, pp. 187–225.
- Franklin-Hall, Laura R (2015). Explaining causal selection with explanatory causal economy: Biology and beyond. *Explanation in biology*. Springer, pp. 413–438. DOI: 10.1007/978-94-017-9822-8_18.
- Geach, Peter Thomas (1969). The perils of Pauline. *The Review of Metaphysics* 23.2, pp. 287–300.
- Gettier, Edmund (1963). Is knowledge justified true belief? *Analysis* 23.6, pp. 121–123.
- Goldman, Alvin (1967). A causal theory of knowing. *The Journal of Philosophy* 64 (12), pp. 138–153. DOI: 10.2307/2024268.
- Hanus, Daniel and Josep Call (2011). Chimpanzee problem-solving: contrasting the use of causal and arbitrary cues. *Animal Cognition* 14.6, pp. 871–878. DOI: 10.1007/s10071-011-0421-6.

References III

- Hart, H. L. A. and Tony Honoré (1959). *Causation in the Law*. Clarendon Press. DOI: 10.1093/acprof:oso/9780198254744.001.0001.
- Icard, Thomas F, Jonathan F Kominsky, and Joshua Knobe (2017). Normality and actual causal strength. *Cognition* 161, pp. 80–93.
- Jacovi, Alon et al. (2021). Contrastive explanations for model interpretability. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1597–1611.
- Kripke, Saul (1972). *Naming and necessity*. Harvard University Press.
- Lewis, David (1983). *Philosophical papers*. Vol. 1. Oxford University Press.
- Lipton, Peter (Nov. 2009). Causation and Explanation. *The Oxford Handbook of Causation*. Oxford University Press. DOI: 10.1093/oxfordhb/9780199279739.003.0030.

References IV

- Mascalzoni, Elena, Lucia Regolin, and Giorgio Vallortigara (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proceedings of the National Academy of Sciences* 107.9, pp. 4483–4485. DOI: 10.1073/pnas.0908792107.
- Michotte, Albert (1963). *The perception of causality*. Basic Books.
- Mittelstadt, Brent, Chris Russell, and Sandra Wachter (2019). Explaining explanations in AI. *Proceedings of the conference on fairness, accountability, and transparency*, pp. 279–288.
- Murray, Pauli and Mary O Eastwood (1965). Jane Crow and the Law: Sex Discrimination and Title VII. *George Washington Law Review* 34, p. 232. URL: https://heinonline.org/HOL/P?h=hein.journals/gwlr34&i=246.
- O'Connell, Sanjida and RIM Dunbar (2005). The perception of causality in chimpanzees. *Animal Cognition* 8.1, pp. 60–66. DOI: 10.1007/s10071-004-0231-1.

References V

- Partee, Barbara (1984). Compositionality. *Varieties of Formal Semantics, Proceedings of the Fourth Amsterdam Colloquium*. Dordrecht: Foris, pp. 281–312.
- Review, Harvard Law (2017). Rethinking actual causation in tort law. *Harvard Law Review* 130, pp. 2163–2182. URL: https://harvardlawreview.org/2017/06/rethinking-actual-causation-in-tort-law/.
- Saxe, Rebecca, JB Tenenbaum, and Susan Carey (2005). Secret agents: Inferences about hidden causes by 10-and 12-month-old infants. *Psychological science* 16.12, pp. 995–1001. DOI: 10.1111/j.1467-9280.2005.01649.x.
- Schaffer, Jonathan (2005). Contrastive causation. *The Philosophical Review* 114.3, pp. 327–358. DOI: 10.1215/00318108-114-3-327.
- — (2010). Contrastive causation in the law. *Legal Theory* 16.4, pp. 259–297. DOI: 10.1017/S1352325210000224.

References VI

- Schlenker, Philippe et al. (2016). Formal monkey linguistics. *Theoretical Linguistics* 42.1-2, pp. 1–90. DOI: 10.1515/tl-2016-0001.
- Schloegl, Christian and Julia Fischer (2017). Causal reasoning in non-human animals. *The Oxford handbook of causal reasoning*. Ed. by Michael R. Waldmann. Oxford University Press, pp. 699–715. DOI: 10.1093/oxfordhb/9780199399550.013.36.
- Vredenburgh, Kate (2022). The right to explanation. *Journal of Political Philosophy* 30.2, pp. 209–229. DOI: 10.1111/jopp.12262.