

# Insurance AI and Fairness

## From Proxy Discrimination to Governed Trade-offs

**Arthur Charpentier**, with Olivier Côté and Marie-Pier Côté

IFoA AI Ethics Governance & Risk Management discussion, April 2026



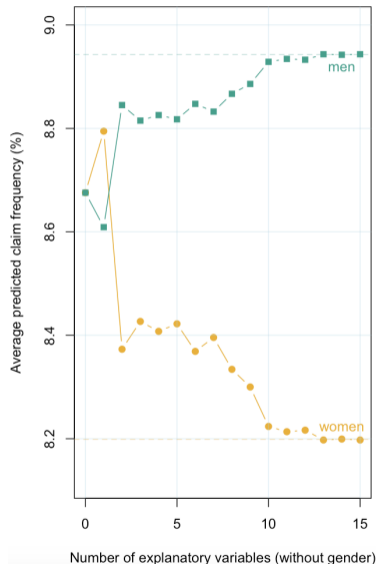
# Fairness by Unawareness

On a French motor dataset, average claim frequencies are **8.94%** (men) and **8.20%** (women).

Consider some logistic regression to estimate annual claim frequency, on  $k$  explanatory variables **excluding gender**.

	men	women
$k = 0$	8.68%	8.68%
$k = 2$	8.85%	8.37%
$k = 8$	8.87%	8.33%
$k = 15$	8.94%	8.20%
empirical	8.94%	8.20%

Models simply tend to **reproduce what was observed** in the data (see “**is-ought**” problem, in **Hume (1739)**).

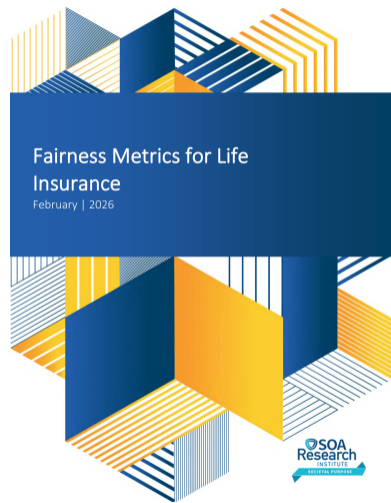


# Classical Dilemma

- ▶ Supreme Court Justice Harry Blackmun stated, in 1978,  
*“In order to get beyond racism, we must first take account of race. There is no other way. And in order to treat some persons equally, we must treat them differently,”* Knowlton (1978), cited in Lippert-Rasmussen (2020)
  - To counteract **disparate impact**, intentional **disparate treatment** is necessary
  - See philosophical discussions about **affirmative action**, e.g., Rubinfeld (1997); Pojman (1998); Anderson (2004)
- ▶ In 2007, John G. Roberts of the U.S. Supreme Court submits  
*“The way to stop discrimination on the basis of race is to stop discriminating on the basis of race,”* Sabbagh (2007) and Turner (2015)
  - Corresponds to the “colorblind” approach
  - Rejects any form of **disparate treatment**, even for corrective purposes, and reproduction of historical inequalities will lead to **disparate impact**

## “Fairness Metrics for Life Insurance”, Krafcheck et al. (2026)

- ▶ Fairness in life insurance is not a single concept.
- ▶ The paper takes a practical approach: define what fairness means before choosing a metric.
- ▶ It distinguishes between individual fairness and group fairness.
- ▶ Individual fairness is close to the actuarial view: similar risks should be treated similarly.
- ▶ Group fairness seeks parity across groups, which may conflict with risk-based pricing.
- ▶ On the individual side, it reviews unawareness, awareness, and no omitted-variable bias.
- ▶ On the group side, it discusses independence, sufficiency, and separation.
- ▶ Main takeaway: no single fairness metric is enough; trade-offs must be made explicit and discussed.

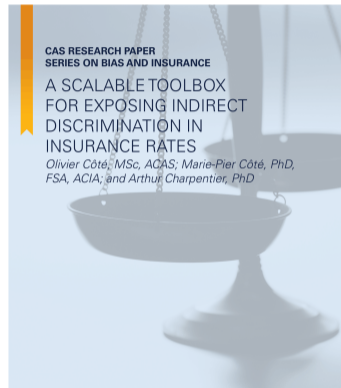


## “Governing the Fairness Trilemma in Insurance”, Charpentier et al. (2026)

- ▶ Structural tension: insurance is both risk-based pricing and risk-sharing.
- ▶ It proposes a fairness trilemma between actuarial fairness, solidarity, and causal legitimacy.
  - **Actuarial fairness** means aligning premiums with expected losses.
  - **Solidarity** means allowing selected departures from pure risk-based pricing to preserve access and redistribution.
  - **Causal legitimacy** means that every price differentiation must be justified by an admissible causal relationship, not just by predictive power.
- ▶ These three goals are individually attractive but cannot be fully satisfied at the same time.
- ▶ So the real issue is not to “solve” fairness once and for all, but to **govern trade-offs explicitly**, using scorecards, causal due diligence, and a comply-or-explain framework.

# “A Scalable Toolbox for Exposing Indirect Discrim...”, Côté et al. (2025a)

- ▶ Removing a sensitive variable does not remove discrimination, because other variables may still act as proxies.
- ▶ Its goal is to make fairness in insurance operational and measurable in actuarial terms.
- ▶ It organizes fairness around three dimensions: actuarial fairness, solidarity, and causality.
- ▶ It studies these issues in a real auto insurance case study from Québec province, using credit score as a sensitive variable.
- ▶ A key contribution is to express fairness problems in dollars and people, not only in abstract statistical metrics.



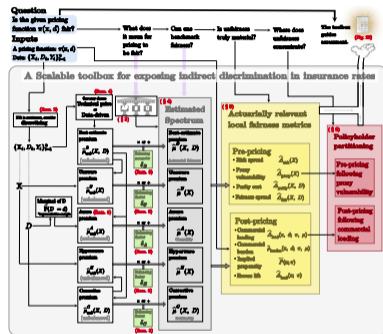
CASUALTY ACTUARIAL SOCIETY

Commissioned by the Canada Race and Insurance Pricing Task Force



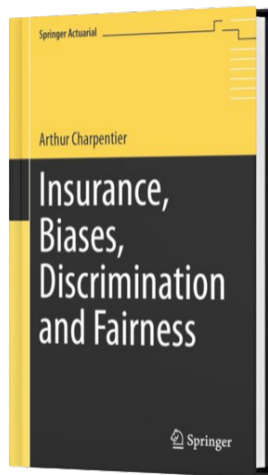
# “A Scalable Toolbox for Exposing Indirect Discrimin...”, Côté et al. (2025a)

- ▶ The paper introduces local metrics such as risk spread, **proxy vulnerability**, fairness range, and parity cost.
- ▶ Proxy vulnerability measures how much a segment may be over- or under-priced because neutral variables indirectly reconstruct the sensitive attribute.
- ▶ The main takeaway is that the toolbox helps actuaries detect, quantify, and communicate indirect discrimination
- ▶ Following Côté et al. (2024), “A Fair price to pay: exploiting causal graphs for fairness in insurance”



## To go further...

- ▶ Charpentier (2024) *Insurance, Biases, Discrimination and Fairness*. Springer.
- ▶ A pricing rule that looks fair on one insurer's portfolio may still be unfair at the market level, because **portfolios are not representative** of the insured population, Côté et al. (2025b).
- ▶ Enforcing one notion of fairness may end up robbing Peter to pay Paul.
- ▶ When **sensitive variables are unobserved**, the question is no longer only “Is the model fair?”, but also “How can fairness be audited, evidenced, and governed?”



# References

- Anderson, T. H. (2004). *The pursuit of fairness: A history of affirmative action*. Oxford University Press.
- Charpentier, A. (2024). *Insurance, Biases, Discrimination and Fairness*. Springer, Switzerland.
- Charpentier, A., Côté, O., and Côté, M.-P. (2026). Governing the fairness trilemma in insurance pricing: Actuarial adequacy, social solidarity and causal legitimacy. *to appear soon*.
- Côté, O., Côté, M.-P., and Charpentier, A. (2024). A fair price to pay: exploiting causal graphs for fairness in insurance. *Journal of Risk and Insurance*.
- Côté, O., Côté, M.-P., and Charpentier, A. (2025a). A scalable toolbox for exposing indirect discrimination in insurance rates. *Casualty Actuarial Society*.
- Côté, O., Côté, M.-P., and Charpentier, A. (2025b). Selection bias in insurance: Why portfolio-specific fairness fails to extend market-wide. *SSRN*.
- Hu, F., Ratz, P., and Charpentier, A. (2023). Fairness in multi-task learning via wasserstein barycenters. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases – ECML PKDD*.
- Hu, F., Ratz, P., and Charpentier, A. (2024). A sequentially fair mechanism for multiple sensitive attributes. *Annual AAAI Conference on Artificial Intelligence*.

# References

- Hume, D. (1739). *A Treatise of Human Nature*. Cambridge University Press Archive.
- Knowlton, R. E. (1978). Regents of the university of california v. bakke. *Arkansas Law Review*, 32:499.
- Krafcheck, E., Balnozan, I., and Huang, F. (2026). Fairness metrics for life insurance. *SOA Research Institute*.
- Lippert-Rasmussen, K. (2020). *Making sense of affirmative action*. Oxford University Press.
- Pojman, L. P. (1998). The case against affirmative action. *International Journal of Applied Philosophy*, 12(1):97–115.
- Rubinfeld, J. (1997). Affirmative action. *Yale Law Journal*, 107:427.
- Sabbagh, D. (2007). *Equality and transparency: A strategic perspective on affirmative action in American law*. Springer.
- Turner, R. (2015). The way to stop discrimination on the basis of race. *Stanford Journal of Civil Rights & Civil Liberties*, 11:45.