

# Preface

*Mattias Gunnemyr & Martin L. Jönsson*

In Gunnemyr, Mattias & Jönsson, Martin L. (2023) *Post Hoc Interventions: Prospects and Problems*.  
Lund: Department of Philosophy, Lund University. <https://doi.org/10.37852/oblu.184>

ISBN: 978-91-89415-60-7 (print)  
978-91-89415-61-4 (digital – pdf)  
978-91-89415-62-1 (digital – html)



## **Post Hoc Interventions** Prospects and Problems

Published by the Department of Philosophy, Lund University.  
Edited by: Mattias Gunnemyr and Martin L. Jönsson  
Cover layout by Cecilia von Arnold, Pufendorf Institute for Advanced Studies



This text is licensed under a Creative Commons Attribution-NonCommercial license. This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license does not allow for commercial use.

(License: <http://creativecommons.org/licenses/by-nc/4.0/>)

Text © Mattias Gunnemyr and Martin Jönsson 2023. Copyright of individual chapters is maintained by the chapters' authors.

# Preface

This volume is the tangible result of a research incubator (a ‘Pufendorf theme’) that took place in 2022 at the Pufendorf Institute of Advanced Studies in Lund, Sweden. During this time, members of the theme – Martin Jönsson, Mattias Gunnemyr, Jakob Bergman, Erik Girvan, Thore Husfeldt, Kasper Lippert Rasmussen, Anna Nilsson, and Una Tellhed – met at the Pufendorf Institute, once a week, for a yearlong interdisciplinary conversation on post hoc interventions (see Chapter 1 for a brief introduction). It is a testament to the broad expertise of the research group (containing as it did, epistemological, ethical, statistical, computational, legal, and psychological expertise), to the open-mindedness of the experts, and the fertility of the subject matter, that so much progress could be made in such a short time.

Scientific progress involves both uncovering new *prospects* – identifying embryos to new interventions, places of application, ways to overcome obstacles – as well as uncovering new *problems* – identifying restrictions and limitations, and disheartening dead ends at the end of once promising paths. As its name implies, this volume contains examples of both kinds.

One problem – discussed by Una Tellhed in Chapter 2 of this volume – is that GIIU – the post hoc intervention most often in focus during the year – qua structural social bias intervention, is likely to meet various forms of resistance once introduced into an organization; people might, for instance, deny that social bias exists, trivialize discrimination and segregation issues, and distrust an algorithm that offers suggestions at odds with their personal opinions. Tellhed describes some of these forms of resistance and lists some suggestions on how to proceed.

In Chapter 3, Jakob Bergman begins by offering a prospect. He discusses the possibility of bias that works in a non-linear way and outlines a test to detect different types of biases of this kind, which might be used as a failsafe when applying GIIU. He also briefly discusses situations where candidates are

## *Post Hoc Interventions: Prospects and Problems*

assessed on several dimensions, or by several evaluators, and shares some ideas on how to proceed in such cases.

In Chapter 4, Erik Girvan considers what would happen if we used GIIU to adjust for potential ethnic biases in criminal sentencing outcomes in the U.S. In particular, he considers whether GIIU's presuppositions could be satisfied in practice and argues that while the requirements may not all be satisfied in their strong form, they can likely be satisfied in many circumstances.

The Law, by its very nature, is restrictive, and it is thus unsurprising that post hoc interventions might face legal problems. In addition to improving the accuracy of evaluations, GIIU can also help to reduce discrimination. Doing so is permitted, and even prescribed, in most legal systems in liberal democracies. Still, many legal systems prohibit affirmative action such as the use of quotas or other preferential treatment, at least in some areas. While GIIU is importantly different from affirmative action, it is still important to decide whether it is different enough from a legal perspective. In chapter 5, Anna Nilsson examines the implications of Swedish and EU discrimination law for the use of post hoc interventions during recruitment processes. She argues that post hoc interventions such as GIIU might be thought of in two different ways: either as a tool that corrects for biases and prejudice or as a form of preferential treatment. If it is best characterized as a tool that corrects for biases and prejudice, there is nothing in the Swedish Discrimination Act that prevents an employer from using GIIU. There is, however, a risk that a Swedish court could find that the use of GIIU is a form of preferential treatment. If so, it would not be allowed to use GIIU to compensate for ethnic discrimination in Sweden. (Still, it would be allowed to use GIIU to compensate for prejudice against persons with disabilities and persons with transgender identity or expression.) Further, because of EU regulations, it would not be allowed to use GIIU to automatically update biased decisions when the bias concerns the sex of the applicants. Instead, in such cases, GIIU must be used as a decision support device.

Chapter 6 also discusses legal questions in relation to post hoc interventions. Martin Jönsson and Jonas Ledendal investigate the possible tension between GIIU and GDPR (the General Data Protection Regulation, which restricts data processing within the EU). They conclude that many applications of GIIU are in compliance with the GDPR, even without the specific consent for this processing of data subjects, but that others might not be, specifically those where the processing includes special categories of personal data that is considered sensitive.

## *Preface*

The final three chapters concern the ethics of post hoc interventions, broadly construed. In Chapter 7, Mattias Gunnemyr considers three potential ethical problems in relation to post hoc interventions: that post hoc interventions might infringe on the decision makers freedom to make decisions in morally problematic way, that GIU might indicate that a certain decision is biased even if it is not, and that GIU might rely on probabilistic evidence that does not tell us anything about whether the decision at hand is biased. Gunnemyr concludes on a positive note that 1) while post hoc interventions might infringe on the freedom of the decision makers, they do not do so in a problematic way – especially not if GIU is implemented as decision support system, that 2) GIU actually requires an additional presupposition, or should at least be constrained so that it is only applied in a specific way to avoid incorrect updates of evaluations, and that 3) post hoc interventions do not rely on probabilistic evidence in a problematic way.

In the US, courts have used the algorithmic tool COMPAS to assess potential recidivism risk. Critics have argued that the court’s use of COMPAS unfairly disadvantages black offenders since it lacks calibration across groups. In particular, it generates higher false positive rates of predicted recidivism for black offenders than white offenders. Kasper Lippert-Rasmussen argues in Chapter 8 that we do not think that lack of calibration entails unfair bias in non-algorithmic contexts, such as hirings, and that we therefore should reject the view that calibration is necessary for fairness in an algorithmic context.

In the final chapter, Thore Husfeldt considers whether fairness requires calibration across groups, but from another perspective. He starts off formalizing some of our most common notions of fairness and shows that they are incompatible. For instance, we might think that fairness requires that group membership (in terms of sex, ethnicity, etc.) does not influence who gets recruited for a certain position, and that all groups should be represented in proportion to their part of the entire population (at a certain job, in parliament, etc.). These kinds of fairness are sometimes called “equal odds” and “democratic parity”. Husfeldt shows that it is impossible to be fair in both ways simultaneously in non-trivial cases.

Most of the contributions in this volume (with the exception of chapters 3 and 6) were presented at a conference which shares its name on October 5<sup>th</sup>-6<sup>th</sup> 2022 in Lund. We are thankful to the participants of that conference for their many suggestions, questions and requests that indirectly helped improve this volume. We are particularly indebted to the invited commentators at the conference for their generous feedback. Nazar Akrami, Fredrik Björklund, Eric

## *Post Hoc Interventions: Prospects and Problems*

Brandstedt, Leila Brännström, Jenny Magnusson, Boel Nelson, Maria Stanfors, and Frej Klem Thomsen. Thank you!

We would also be remiss if we forgot to thank everyone who has visited our weekly seminar during the theme's run: Nazar Akrami, Ramón Alvaro, György Barabás, Fredrik Björklund, Michaela Digneus, Jonas Ledendal, Fredrik Lindstrand, Alex Madva, Sara Martinsson, Gregor Noll, Julian Schuessler, Anders Sjöberg, András Szigeti, Kim Mannemar Sønderskov, and Joan Williams. Thank you for sharing your work and for tuning in to our conversation.

Moreover, we would like to thank everyone at the Pufendorf Institute in Lund for their financial support, and for giving the group the opportunity to spend its Wednesdays at their beautiful premises. Johanna Albiñ, Ann-Katrin Bäcklund, Stephan Choquette, Eva Persson, Sune Sunesson, Stacey Ristinmaa Sörensen och Cecilia von Arnold. Thank you for all your help.

Last but not least, we would like to thank the other wonderful participants of the theme: Anna, Erik, Jakob, Kasper, Thore and Una. For your efforts, patience and insights. Thank you.

/ Mattias Gunnemyr and Martin Jönsson, January 2023