



VRIJE
UNIVERSITEIT
BRUSSEL



Master thesis submitted in partial fulfilment of the requirements for
the degree of Master of Science in Applied Sciences and Engineer-
ing: Computer Science

INVESTIGATING FAIRNESS IN EPIDEMIC CONTROL

Sam Vanspringel

2024-2025

Promotors: Professor Ann Nowé, Professor Pieter Libin
Science and Bio-Engineering Sciences



VRIJE
UNIVERSITEIT
BRUSSEL



Proefschrift ingediend met het oog op het behalen van de graad
van Master of Science in Applied Sciences and Engineering: Com-
puter Science

FAIRNESS ASPECTEN BIJ HET BESTRIJDEN VAN EPIDEMIEËN

Sam Vanspringel

2024-2025

Promotors: Professor Ann Nowé, Professor Pieter Libin
Wetenschappen en Bio-ingenieurswetenschappen

Acknowledgements

I would like to express my gratitude to my promoters Pieter Libin and Ann Nowé for their continuous guidance, advice, and feedback in this research. Their valuable input helped shape this work profoundly.

Special thanks to my PhD mentor Alexandra Cimpean, who held weekly meetings and provided insightful feedback and guidance to ensure the gradual development of my Master's thesis.

My deep appreciation to my family for their support, encouragement, and for providing an environment in which this research could emerge.

Lastly, sincere thanks to my partner for her patience, motivation, and understanding, which made the process more bearable.

Abstract

Epidemic control requires a multifaceted strategy targeting both public health and societal impact. Since minimizing hospitalizations and deaths is a primary goal of mitigation measures, the social consequences are often overlooked in computational policy training. Furthermore, considering age-specific vulnerabilities further increases the complexity of learning optimal mitigation policies. To this end, we explore epidemic interventions in a multi-objective reinforcement learning (MORL) setting. This enables us to simultaneously minimize the public health concerns and social stress imposed on the population. While previous work considers a social burden objective to quantify the overall loss of social contacts, it does not account for a fair distribution of this social burden across all age-groups. To resolve this, we propose two fairness notions, *social burden fairness (SBF)* and *age-based fairness through unawareness (ABFTU)* to consider the fair distribution of these social consequences. We employ Pareto Conditioned Networks (PCN) to discover all non-dominated policies that optimize different combinations of fairness objectives alongside controlling the disease outbreak. To assess the effects of including the fairness notions, we compare the results of multiple distinct settings. Visualization tools are used to explore the learned policies and both their epidemiological and societal impact. We demonstrate how including fairness into the optimization process allows us to maintain the performance of previous work. The proposed fairness notions ensure social burden is distributed more fairly across and among the age-groups.

Contents

1	Introduction	5
2	Modelling the COVID-19 epidemic	6
2.1	Mathematical Compartmental Transmission model	6
2.2	Discrete time stochastic epidemic model	9
2.3	Interventions and Compliance	10
3	Reinforcement Learning	11
3.1	Multi-objective Reinforcement Learning	12
3.1.1	Multi-Objective Reinforcement Learning Setting	13
4	Pareto Conditioned Networks	14
4.1	PCN algorithm	14
4.2	PCN for Epidemiological Control	15
5	Environment	16
5.1	State Space	16
5.2	Action Space	16
5.3	Transition Dynamics	17
5.4	Reward Function	17
6	Fairness	17
6.1	Social Welfare Functions	17
6.2	Fairness in Machine Learning	18
6.3	Criteria for Selecting Fairness Notions	19
6.4	Existing Fairness notions	19
6.4.1	Group Fairness Notions	19
6.4.2	Individual Fairness Notions	20
7	Fairness in Reinforcement Learning	21
7.1	Fairness Framework	21
7.2	Epidemiological Fairness	21
7.3	Social Burden Fairness	22
7.3.1	Exploring Social Burden Fairness	23
7.4	Age-Based Fairness Through Unawareness	25
7.4.1	Algorithm	26
7.4.2	Exploring Age-Based Fairness Through Unawareness	27
8	Experiments	30
8.1	Experimental Scenarios	30
8.1.1	Social Burden	30
8.1.2	Social Burden & Social Burden Fairness	31
8.1.3	Social Burden Fairness	31
8.1.4	Age-based Fairness Through Unawareness	31
8.1.5	Social Burden Fairness and Age-based Fairness Through Unawareness . .	31
8.2	Overview	31

9	Results	32
9.1	Social Burden & Social Burden Fairness	32
9.1.1	Social Burden Fairness	34
9.1.2	$[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$	38
9.2	Age-based Fairness Through Unawareness	42
9.2.1	$[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$	46
9.3	Social Burden Fairness & Age-based FTU	49
9.3.1	$[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$	51
9.4	Overall Impact of Social Burden Fairness	55
10	Discussion	57
11	Future Work	58
12	Use of Generative AI	59
	References	60
A	Appendices	64
A.1	Social Burden Fairness for different action patterns	64
A.2	Age-Based FTU for different action patterns	65

1 Introduction

A major priority in the event of an epidemic is avoiding an excessive number of hospitalizations and deaths. However, when focusing on the limitation of the medical repercussions, additional impacts on the population should be considered. As the global COVID-19 pandemic is a few years old, we now observe the impact of the severe social restrictions [Witteveen et al., 2023, Mazrekaj and Witte, 2024]. In particular, it is important to consider multiple factors. On the one hand, quarantines should be used to ensure that the spread of the disease is limited. This minimizes the consequences and helps avoid long-lasting negative effects on public health. On the other hand, imposing severe social restrictions could have a broad psychological impact [Brooks et al., 2020]. A quarantine poses multiple challenges for the population, such as the duration, fear of infection, frustration, boredom, and potentially inadequate information and supplies. Brooks et al. suggest that the necessity of epidemic control should be accompanied by as bearable social restrictions as possible, clear communication and sufficient supplies. Moreover, the duration should be kept to a minimum while keeping the population busy. These mitigation strategies ensure a non-negative experience in a quarantine, minimizing the psychological consequences [Brooks et al., 2020].

Assessing these psychological effects is important. Therefore, it is crucial that the mental strain of the population as a result of social restrictions is considered. However, the population is not homogeneous. Different individuals have distinct social behaviors and are affected in different magnitudes. Therefore, it would be desirable to consider this heterogeneous nature in deciding upon mitigation measures. As researchers have extensively researched the age-dependent effects of COVID-19, it may be advantageous to consider the age of individuals [Davies et al., 2020].

Determining the social restrictions while considering these different objectives is a difficult task. Since epidemiological processes are complicated, they pose a great challenge to develop the best approach to mitigation strategies. A manual approach to this problem would require great effort from domain-experts and considerations for various points of view. Therefore, automated learning systems such as reinforcement learning systems could support such decision-making, reducing the workload and facilitating the process [Libin et al., 2021]. These systems enable automatic decision-making in epidemiological environments through simulation. Consequently, mitigation strategies can be constructed, while considering the complex nature of the epidemiological environment.

While classic reinforcement learning focuses on optimizing a single objective, this approach can be extended to include multiple objectives [Hayes et al., 2022]. Various multi-objective reinforcement learning (MORL) approaches exist, but the suitability is context-dependent [Hayes et al., 2022]. In an epidemiological setting, it would be desirable to explore multiple mitigation strategies, enabling domain-experts to assess different trade-offs. To this end, we employ a multi-policy approach that learns the Pareto front [Hayes et al., 2022]. Rather than returning one solution that maximizes or minimizes a metric, this approach results in a set of solutions that optimize the multifaceted issue. Utilizing this technique, we can determine policies that simultaneously suppress the medical repercussions and optimize the well-being of the population. This allows us to approach the decision of social restrictions in a more informed manner, rather than solely basing the regulations on the number of hospitalizations or deaths. This enables policy-makers to impose social restrictions balanced between public medical health and social expenses. Additionally, this strategy offers the possibility to even further extend this decision-making to include additional objectives in the future.

To quantify the burden of social restrictions on the population, we can utilize fairness notions. These express how fair we consider the measures taken to restrict an epidemic. The applicability of fairness notions has been extensively studied, but predominantly focused on machine learning [Makhlouf et al., 2020]. Applying these concepts to reinforcement learning, forces us to adapt or even develop measures to obtain the same descriptive power as previously seen techniques. Utilizing the fairness framework by Cimpean et al., we can incorporate these measures for consideration in an automated learning system. As previously mentioned, this work considers different age-groups and their susceptibility to the serious consequences of infection. We perform experiments to explore how the restriction of social interactions affects this fairness. This provides insight into the potentially negative effects of a policy when certain age-groups are health-wise less affected by an epidemic. Furthermore, we leverage the fairness notions to influence the RL-decision making to impose different social restrictions. This elucidates the difference in policies when an epidemic is approached as a more multifaceted problem.

In this research, we add to existing research by exploring the possibility of using reinforcement learning to optimize multiple objectives, including multiple fairness notions that serve as social welfare functions [Ogryczak et al., 2014]. We do this by experimenting with maximizing certain fairness measures alongside other objectives, while using existing techniques [Reymond et al., 2022].

2 Modelling the COVID-19 epidemic

The COVID-19 epidemic had a substantial impact on global life. To reduce the spread of the disease, intervention measures, such as social restrictions, were put in place to reduce social contacts. These reductions ensured the disruption of an unrestrained trajectory of the epidemic. Such unrestrained disease outbreaks would cause an overwhelming number of infections and subsequently hospitalizations and eventually deaths. Different models have been explored for modelling epidemics along with automated mitigation techniques [Libin et al., 2020]. The epidemiological models are categorized into three general groups. These are individual-based, compartmental, and meta-population models [Miranda et al., 2022]. The individual-based models provide epidemiological simulations on a personal level, while compartmental models monitor the population more broadly on how its distribution evolves over different health states. Compartmental models are efficient to evaluate, while providing general information, and individual-based models offer more detailed information, but are computationally costly. The meta-population models provide a balanced solution to this problem by segmenting the population and constructing multiple compartmental models for the different segments. In this work, the effects of the imposed reductions are simulated in a compartmental model constructed by Abrams et al.. This allows simulating and altering said epidemiological trajectory [Abrams et al., 2021]. Abrams et al. constructed an age-structured compartmental model that we utilize in this work. The model incorporates the different states, transitions and social contacts of the population to represent the epidemic in a detailed manner. Additionally, we discuss the stochastic extension of this model. However, due to computational convenience and the similarity between the variants, this model was omitted for the experiments.

2.1 Mathematical Compartmental Transmission model

To describe the dynamical aspect of the epidemic, Abrams et al. use a *SEIR* model. This model captures the trajectory of the epidemic in an age-structured way, while accounting for social contacts. We consider ten age-groups (Table 1) to account for differences in disease-specific characteristics and social behavior [Abrams et al., 2021].

Index	Age-group	Hospitalization Risk	School	Work	Leisure
1	0–10	0.0212	6.1	0.4	2.8
2	10–20	0.0060	6.6	1.0	4.7
3	20–30	0.0121	0.9	8.9	4.4
4	30–40	0.0098	0.2	13.0	4.7
5	40–50	0.0174	0.1	9.4	3.5
6	50–60	0.0219	0.2	8.2	4.3
7	60–70	0.0318	0.0	1.6	5.5
8	70–80	0.0559	0.0	0.8	2.8
9	80–90	0.0333	0.0	0.5	1.1
10	90+	0.0559	0.0	0.1	0.2

Table 1: Hospitalization risks and average total weekly contacts per age-group and area.

Table 1 depicts the differences in hospitalization risks per group and their distinct social patterns. We consider three domains where the population encounters their social contacts. Particularly, we display the average weekly contacts for each age-group for work, school, and leisure.

Considering the epidemiological model, all individuals can belong to one of five compartments for their age-group:

- **Susceptible (S):** Individuals that are susceptible to getting infected.
- **Exposed (E):** Individuals that got in contact with an infected person. They will then transition to the infected state at an age- and time-specific rate.
- **Infected (I):** Infected individuals are further categorized into different degrees of infection:
 - **Pre-symptomatic (I_{presym}):** Infected, but the individual experiences no symptoms yet.
 - **Asymptomatic (I_{asym}):** Individual experiences no symptoms.
 - **Mildly symptomatic (I_{mild}):** The individual exhibits mild symptoms.
 - **Severely symptomatic (I_{sev}):** Severe symptoms can lead to either hospitalization or intensive care (I_{hosp} & I_{icu}).
- **Recovered (R):** Recovered individuals. These individuals acquire immunity for the time considered [Abrams et al., 2021].
- **Deceased (D):** Individuals that died.

The different states are defined for each age-group $1, \dots, K$ at time t . We refer to the susceptible individuals for all age-groups at time t as $S(t)$. To describe the way individuals shift between these compartments, we use a set of ordinary differential equations (ODEs) [Abrams et al., 2021]. These equations use context-specific rates to govern how individuals switch states in the epidemic. The ODEs for moving individuals between the different compartments are described in the following equations:

$$\frac{dS(t)}{dt} = -\lambda(t)S(t) \quad (1)$$

$$\frac{dE(t)}{dt} = \lambda(t)S(t) - \gamma E(t) \quad (2)$$

$$\frac{dI_{\text{presym}}(t)}{dt} = \gamma E(t) - \theta I_{\text{presym}}(t) \quad (3)$$

$$\frac{dI_{\text{asym}}(t)}{dt} = \theta p I_{\text{presym}}(t) - \delta_1 I_{\text{asym}}(t) \quad (4)$$

$$\frac{dI_{\text{mild}}(t)}{dt} = \theta(1-p) I_{\text{presym}}(t) - \{\psi + \delta_2\} I_{\text{mild}}(t) \quad (5)$$

$$\frac{dI_{\text{sev}}(t)}{dt} = \psi I_{\text{mild}}(t) - \omega I_{\text{sev}}(t) \quad (6)$$

$$\frac{dI_{\text{hosp}}(t)}{dt} = \phi_1 \omega I_{\text{sev}}(t) - \{\delta_3 + \tau_1\} I_{\text{hosp}}(t) \quad (7)$$

$$\frac{dI_{\text{icu}}(t)}{dt} = (1 - \phi_1) \omega I_{\text{sev}}(t) - \{\delta_4 + \tau_2\} I_{\text{icu}}(t) \quad (8)$$

$$\frac{dD(t)}{dt} = \tau_1 I_{\text{hosp}}(t) + \tau_2 I_{\text{icu}}(t) \quad (9)$$

$$\frac{dR(t)}{dt} = \delta_1 I_{\text{asym}}(t) + \delta_2 I_{\text{mild}}(t) + \delta_3 I_{\text{hosp}}(t) + \delta_4 I_{\text{icu}}(t) \quad (10)$$

The constants λ , γ , θ , δ , ψ , τ , and ω control the rates at which individuals transition between states. The transitions between states are described in Figure 1.

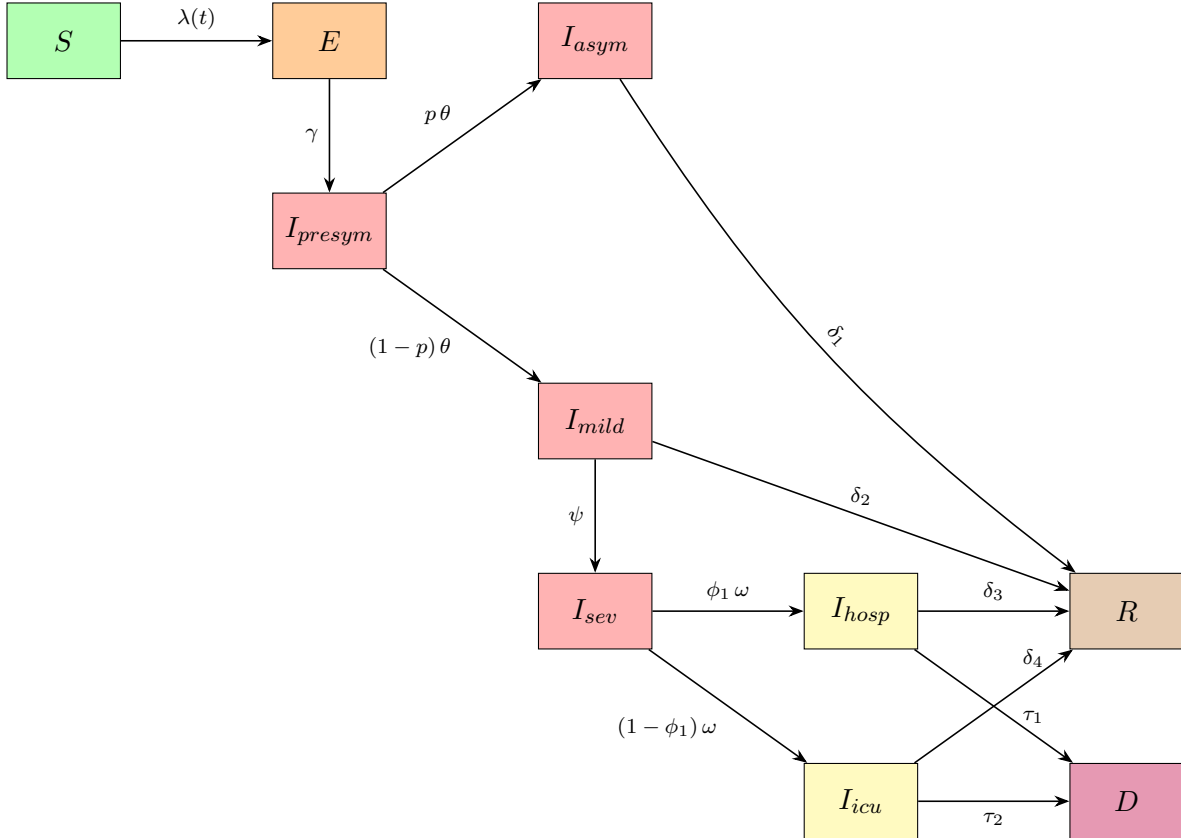


Figure 1: The different epidemiological compartments for each age-group defined by Abrams et al.. The model consists of ten compartments per age-group: susceptible (S), exposed (E), pre-symptomatic infection (I_{presym}), asymptomatic infection (I_{asym}), infection with mild symptoms (I_{mild}), infection with severe symptoms (I_{sev}), hospitalized (I_{hosp}), ICU (I_{icu}), death (D), and recovered (R).

The transmission rate, λ specifically, is based on social contact data between the age-groups. The force of infection for an individual of age-group k at the time t is defined as:

$$\lambda(k, t) = \sum_{k'=1}^K \beta(k, k') I(k', t),$$

where $\beta(k, k')$ represents the average transmission rate at which an infectious individual from age-group k' makes an effective contact with an individual of age-group k . $I(k', t)$ describes the total number of infectious individuals in an age-group k' at time t . The transmission rate $\beta(k, k')$ incorporates the social contacts between age-group k and k' by relying on the social contact hypothesis [Wallinga et al., 2006] and is determined by:

$$\beta(k, k') = qc(k, k')$$

with $c(k, k')$ the rate at which an individual of age-group k makes social contact with an individual of age-group k' at time t . The value q accounts for variability in social contacts due to infections and susceptibility. The structure of $c(k, k')$ is a $K \times K$ -matrix that integrates how the different age-groups interact with each other in the epidemic. The different values of $c(k, k')$ are derived from social contact data collected in 2010 by Kifle et al., Willem et al., Willem et al., and Hoang et al.. To account for the difference in social contacts between symptomatic and non-symptomatic individuals, two social contact matrices are defined. Two distinct matrices should be utilized, since social behavior differs between symptomatic and asymptomatic individuals. Consequently, this changes the transmission rate and the age-dependent force of infection:

$$\begin{aligned}\beta_{sym}(k, k') &= q_{sym} c_{sym}(k, k') \\ \beta_{asym}(k, k') &= q_{asym} c_{asym}(k, k') \\ \lambda(k, t) &= \beta_{sym}(k, k') + \beta_{asym}(k, k')\end{aligned}$$

2.2 Discrete time stochastic epidemic model

Since humans and viruses do not behave deterministically, the previously defined model may be insufficient to capture this nondeterministic behavior. Therefore, the authors update the original ODEs to incorporate stochastic components. The stochastic epidemic model accounts for both variability in the model parameters and for the stochasticity of the epidemic. In the stochastic discrete-time chain binomial model by Bailey, the model is evaluated at each time interval $(t, t+h]$, with $h = 1/24$ day. Each time-step can be interpreted as a Bernoulli trial, since individuals either transition to different states, or they do not. Altogether, these independent trials can be formulated binomially. The model describes the newly exposed individuals as:

$$E_{t+h}^{\text{new}}(k) \sim \text{Binomial}(S_t(k), p_t^*(k) = 1 - \{1 - p_t(k)\}^{I_t})$$

where $S_t(k)$ are the susceptible individuals at time t in age-group k and I_t the amount of infected individuals at time t . Assuming $p_t(k)$ the probability of transmission and $q_t(k) = 1 - p_t(k)$ the probability of escaping it, the probability of a person not being exposed when contacting m individuals is $q_t^m(k)$. This results in an infection probability of:

$$p_t^*(k) = 1 - \exp \left[-h \sum_{k'=1}^K (\beta_{asym}(k, k') \{I_{\text{presym},t}(k') + I_{\text{asym},t}(k')\} + \beta_{sym}(k, k') \{I_{\text{mild},t}(k') + I_{\text{sev},t}(k')\}) \right]$$

The number of individuals in the different compartments is described by the following equations:

$$\begin{aligned}
S_{t+h}(k) &= S_t(k) - E_{t+h}^{\text{new}}(k), \\
E_{t+h}(k) &= E_t(k) + E_{t+h}^{\text{new}}(k) - I_{\text{presym},t+h}^{\text{new}}(k), \\
I_{\text{presym},t+h}(k) &= I_{\text{presym},t}(k) + I_{\text{presym},t+h}^{\text{new}}(k) - I_{\text{asym},t+h}^{\text{new}}(k) - I_{\text{mild},t+h}^{\text{new}}(k), \\
I_{\text{asym},t+h}(k) &= I_{\text{asym},t}(k) + I_{\text{asym},t+h}^{\text{new}}(k) - R_{\text{asym},t+h}^{\text{new}}(k), \\
I_{\text{mild},t+h}(k) &= I_{\text{mild},t}(k) + I_{\text{mild},t+h}^{\text{new}}(k) - I_{\text{sev},t+h}^{\text{new}}(k) - R_{\text{mild},t+h}^{\text{new}}(k), \\
I_{\text{sev},t+h}(k) &= I_{\text{sev},t}(k) + I_{\text{sev},t+h}^{\text{new}}(k) - I_{\text{hosp},t+h}^{\text{new}}(k) - I_{\text{icu},t+h}^{\text{new}}(k), \\
I_{\text{hosp},t+h}(k) &= I_{\text{hosp},t}(k) + I_{\text{hosp},t+h}^{\text{new}}(k) - D_{\text{hosp},t+h}^{\text{new}}(k) - R_{\text{hosp},t+h}^{\text{new}}(k), \\
I_{\text{icu},t+h}(k) &= I_{\text{icu},t}(k) + I_{\text{icu},t+h}^{\text{new}}(k) - D_{\text{icu},t+h}^{\text{new}}(k) - R_{\text{icu},t+h}^{\text{new}}(k), \\
D_{t+h}(k) &= D_t(k) + D_{\text{hosp},t+h}^{\text{new}}(k) + D_{\text{icu},t+h}^{\text{new}}(k), \\
R_{t+h}(k) &= R_t(k) + R_{\text{asym},t+h}^{\text{new}}(k) + R_{\text{mild},t+h}^{\text{new}}(k) + R_{\text{hosp},t+h}^{\text{new}}(k) + R_{\text{icu},t+h}^{\text{new}}(k).
\end{aligned}$$

With the transitions between compartments within the specified interval h described by the following derivations:

$$\begin{aligned}
I_{\text{presym},t+h}^{\text{new}}(k) &\sim \text{Binomial}(E_t(k), 1 - \exp\{-h\gamma\}), \\
I_{\text{asym},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{presym},t}(k), 1 - \exp\{-hp(k)\theta\}), \\
I_{\text{mild},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{presym},t}(k), 1 - \exp\{-h[1 - p(k)]\theta\}), \\
I_{\text{sev},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{mild},t}(k), 1 - \exp\{-h\psi(k)\}), \\
I_{\text{hosp},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{sev},t}(k), 1 - \exp\{-h\phi_1(k)\omega(k)\}), \\
I_{\text{icu},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{sev},t}(k), 1 - \exp\{-h[1 - \phi_1(k)]\omega(k)\}), \\
D_{\text{hosp},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{hosp},t}(k), 1 - \exp\{-h\tau_1(k)\}), \\
D_{\text{icu},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{icu},t}(k), 1 - \exp\{-h\tau_2(k)\}), \\
R_{\text{asym},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{asym},t}(k), 1 - \exp(-h\delta_1)), \\
R_{\text{mild},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{mild},t}(k), 1 - \exp\{-h\delta_2(k)\}), \\
R_{\text{hosp},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{hosp},t}(k), 1 - \exp\{-h\delta_3(k)\}), \\
R_{\text{icu},t+h}^{\text{new}}(k) &\sim \text{Binomial}(I_{\text{icu},t}(k), 1 - \exp\{-h\delta_4(k)\}).
\end{aligned}$$

where $1/\gamma$ depicts the latency period [Abrams et al., 2021]. Since the behavior of the binomial chain model matches the ODE model, we utilize the latter in this work.

2.3 Interventions and Compliance

Trying to disrupt the above described process, social contact restrictions are imposed. These mitigation measures are carried out by manipulating the number of contacts in the previously mentioned social contact matrices. Therefore, the imposed restrictions affect the structure of the social contact matrices defined for the different age-groups. Consequently, this influences the trajectory of the disease transmission. Since the social contact matrices are fundamental to the transmission rates defined in the compartmental model, this approach inevitably affects the epidemiological trajectory. To account for human behavior, the compliance to the social contact restrictions is defined using a logistic compliance function:

$$p_c(t) = \frac{\exp \{\beta_0^* + \beta_1^*(t - t_I)\}}{1 + \exp \{\beta_0^* + \beta_1^*(t - t_I)\}},$$

where t_I denotes the time of the restrictions.

Abrams et al. explore different exit strategies using these interventions to investigate various ways of slowing down the COVID-19 epidemic to ultimately resume normal social life. In this work, we explore a different approach to researching the different intervention strategies based on the work of Reymond et al.. We utilize social restrictions to mitigate the epidemic. These social restrictions are continuous actions on three distinct areas in which the individuals encounter each other.

3 Reinforcement Learning

In reinforcement learning, an agent progressively learns decision-making by interacting with, and receiving feedback from, a certain environment. This feedback can be either rewards or penalties. As discussed by Sutton and Barto, reinforcement learning is mostly concerned with optimizing a single objective. More specifically, maximizing or minimizing a scalar cumulative reward signal. They discuss a framework in which an agent interacts with an environment in a discrete-time manner and receives a feedback signal or reward. If a reinforcement learning task has the Markov property¹, this framework can be described as a Markov Decision Process or MDP [Sutton and Barto, 2018]. A MDP can be formulated as a 5-tuple:

- $\mathcal{S}$: The set of states of the environment.
- $\mathcal{A}$: The set of actions an agent can execute.
- $\mathcal{R}$: The reward function that produces the feedback signal when transitioning states due to an executed action.

$$\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S}' \rightarrow \mathbb{R}$$

- $\mathcal{P}$: The state-transition probability function that determines how the state changes when an action is executed. This function is updated as the agent learns.

$$\mathcal{P}(s' | s, a) = \mathbf{P}(S_{t+1} = s' | S_t = s, A_t = a)$$

- γ : The discount factor for the rewards that controls the importance of short-term rewards.

In the MDP, the agent repeatedly interacts with the environment by executing actions. These actions induce transitions in the state and trigger a reward from the environment. The goal of the agent is to learn from these rewards to ultimately maximize the expected sum of future rewards. This expected cumulative reward can be formulated as the sum of discounted immediate rewards at each timestep:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

¹Markov Property: The next chosen action depends solely on the current state of the environment and not on past states.

with G_t the cumulative reward at timestep t , R_{t+k} the immediate reward at timestep $t+k$, and $\gamma \in [0, 1]$ a discount factor that controls the importance of future rewards. A smaller γ ensures that the agent favors short-term rewards, since the contribution of future rewards degrades as the discount is applied repeatedly. If gamma is higher, long-term rewards are more important, as the future rewards lose less of their contribution throughout the trajectory. When a γ of 1 is utilized, all rewards contribute equally to the cumulative reward, causing the agent to completely prioritize the maximization of the cumulative reward. The length of the trajectory, over which the cumulative reward is computed, can be finite or infinite. In the case of finite episodes, multiple executions may be necessary for the process. Eventually, the learning leads to the agent having a plan or *policy* of which action to choose in which state. This policy defines how the agent behaves in the environment to achieve the best possible reward. To maximize the return of the policies, the agent can employ different optimization methods such as Q-learning, SARSA, and policy gradient algorithms [Sutton and Barto, 2018]. Therefore, the effectiveness of these methods depends heavily on the reward structure. Upon maximizing the expected cumulative reward, a particular policy π takes shape. The policy maps the combination of a state and action to a probability to define the behavior at hand. A policy that maximizes the cumulative reward by following it, is an optimal policy π^* . The value of a policy π is denoted by:

$$V^\pi(s) = \mathbb{E}_\pi [G_t \mid S_t = s]$$

which corresponds to the expected total discounted reward following policy π in the environment.

3.1 Multi-objective Reinforcement Learning

Most real-world settings are inherently more complex than optimizing a single objective. Therefore, multi-objective reinforcement learning (MORL) is often associated with trade-offs. For example, when optimizing an investment portfolio, we desire the maximization of returns and the minimization of risk [Markowitz, 1952]. In our epidemic mitigation setting, reinforcement learning provides a way to automatically learn prevention strategies using complex models. Classic research in reinforcement learning focuses primarily on optimizing a single objective [Raghu et al., 2017, Yu et al., 2020]. However, epidemic mitigation is a multifaceted problem. It affects public health, the economy, education, public mental health, and more [Deb et al., 2022, Kao et al., 2024, Hamatani et al., 2022, Mazrekaj and Witte, 2024, Ceylan et al., 2020]. We intend to consider multiple and possible conflicting objectives to find the best way to mitigate an epidemic, as it is crucial to learn balanced policies. Minimizing infections/hospitalizations and social burden have already been explored by Reymond et al.. However, this research remained simplistic in quantifying the social burden on the population, as no attention was given to the heterogeneous nature of the population [Reymond et al., 2024]. We elucidate on this subject by improving and extending this research using fairness notions that consider this particular aspect. To this end, we quantify fairness notions that can be utilized as an objective in epidemic control. We use a MORL approach to learn these balanced solutions incorporating fairness. We concentrate on exploring exit strategies in the first wave of the COVID-19 pandemic and focus on minimizing hospitalizations and maximizing the fairness of social contact restrictions.

Since the epidemiological transmission model explicitly defines the force of infection using the degree of social contact, reducing the hospitalizations is fairly straightforward. By restricting the population’s contacts, the transition of susceptible to exposed would lower, and this would ultimately decrease the number of hospitalizations. However, unrestrictedly reducing social

contacts would result in a complete lockdown. While a lockdown is a useful tool to disrupt the trajectory of an epidemic, the effects are broad. It causes a halt of macroeconomic growth, impacts mental health, and can even increase the rate of domestic violence [Meo et al., 2020, Onyeaka et al., 2021, Auray and Eyquem, 2020]. Therefore, it is important to account for other factors in restricting social contact such as social burden, i.e., the fairness of the restrictions. Since the number of hospitalizations and the social burden should be minimized while the fairness should be maximized, this proposes an interesting multi-objective optimization problem. To this end, we can utilize different RL algorithms to learn prevention strategies while interacting with complex epidemic models [Hayes et al., 2022].

3.1.1 Multi-Objective Reinforcement Learning Setting

We utilize the ODE-compartmental model extended to a Multi-Objective oriented Markov Decision process (MOMDP) to use it in a MORL environment [Hayes et al., 2022]. We transform the ordinary Markov Decision Process of form $M = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R} \rangle$ to a Multi-Objective MDP $M = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R}, \mu \rangle$ to support multiple objectives. With $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S}' \rightarrow \mathbb{R}^d$ now a reward vector denoting d objectives and $\mu : \mathcal{S} \rightarrow [0, 1]$ a probability distribution over initial states. In a standard MDP, the initial-state distribution is often omitted or assumed to be a fixed start state. To avoid ambiguity in our multi-objective setting, we include μ explicitly.

In single objective RL, the aim is to optimize the discounted sum of the cumulative reward. However, when dealing with multiple objectives, we cannot assume that one solution is better than another. In case we have a 3-dimensional reward vector, we could end up with optimal solutions $(0, 0, 1)$, $(1, 0, 0)$, and $(0, 1, 0)$. In this scenario, there is no complete ordering of the solution space. This illustrates why MORL finds an optimal solution set of policies rather than a single optimal solution. Moreover, the value function for policies π in an MOMDP employs a reward vector.

Assuming there is no utility function that allows us to scalarize the value function and thus transforming the problem into a single-objective problem, there is only a partial ordering over the solution space. A utility function is formulated by:

$$u : \mathbb{R}^d \rightarrow \mathbb{R} : V_u^\pi = u(\mathbf{V}^\pi)$$

In the epidemiological setting, a utility function is unavailable. There is no way to unequivocally translate the objectives to a single scalar. Therefore, we must use a RL approach that explores the solution space to determine the best possible policies. To reason about the solution sets, we introduce some definitions [Hayes et al., 2022].

Definition 3.1. In the set of all solutions Π , there is a subset $U(\Pi)$ or the *undominated set*, for which there exists a utility function that maximizes the scalar value of a policy:

$$U(\Pi) = \left\{ \pi \in \Pi \mid \exists u, \forall \pi' \in \Pi : u(V^\pi) \geq u(V^{\pi'}) \right\}$$

Definition 3.2. Assuming a subset of utility functions $\mathcal{U}$, the *coverage set* contains the policies with the maximal scalar value for all $u \in \mathcal{U}$:

$$CS(\Pi) \subseteq U(\Pi) \wedge \left(\forall u, \exists \pi \in CS(\Pi), \forall \pi' \in \Pi : u(V^\pi) \geq u(V^{\pi'}) \right)$$

Definition 3.3. A solution $\mathbf{V}^\pi$ of policy π in the solution set is *Pareto-efficient* if no objective of the solution can be improved without negatively impacting another objective.

$$\exists i : \mathbf{V}_i^\pi < \mathbf{V}_i^{\pi'} \implies \exists j : \mathbf{V}_j^\pi > \mathbf{V}_j^{\pi'}$$

Definition 3.4. A policy π that yields return $\mathbf{V}^\pi$ *Pareto-dominates* a policy π' with return $\mathbf{V}^{\pi'}$ if all objectives have a higher or equal value.

$$\mathbf{V}^\pi \succ_P \mathbf{V}^{\pi'} \iff (\forall i : \mathbf{V}_i^\pi \geq \mathbf{V}_i^{\pi'}) \wedge (\exists i : \mathbf{V}_i^\pi > \mathbf{V}_i^{\pi'}).$$

Definition 3.5. If we assume the utility function u to be monotonically increasing, the set of all Pareto-efficient or non-dominated solutions is the *Pareto front*.

$$PF(\Pi) = \{\pi \in \Pi \mid \nexists \pi' \in \Pi : \mathbf{V}^{\pi'} \succ_P \mathbf{V}^\pi\},$$

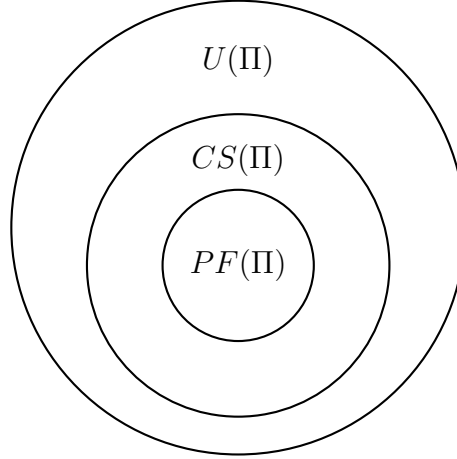


Figure 2: Relationship between solution sets: the undominated set $U(\Pi)$, the coverage set $CS(\Pi)$, and the Pareto front $PF(\Pi)$.

In this case, our goal is to find the Pareto front of policies that yield the best values for the different combinations of objectives. These are further explained in Section 8.

4 Pareto Conditioned Networks

Our goal is to learn the set of best policies in the epidemiological setting. A policy’s performance is evaluated by its return vector, and is Pareto-optimal if no other policy yields a better return in at least one objective without being worse in another. The set of all these non-dominated solutions forms the Pareto front [Reymond et al., 2024]. By examining the Pareto front, decision-makers can understand the trade-offs that it implies. A point on the Pareto front indicates a policy where improving one objective would degrade another [Reymond et al., 2024]. PCN utilizes a single neural network to explore all non-dominated policies. The model is conditioned on the returns of epidemiological episodes. This entails that generating transitions for a certain return, generates the same returns and therefore the same epidemiological trajectory. This enables us to inspect particular policies by specifying the desired return of the episode.

4.1 PCN algorithm

Pareto Conditioned Networks (PCN) is an approach to utilize supervised machine learning to improve policies in MORL. The key idea is to encode all Pareto-optimal policies in one network by conditioning the policy on a desired outcome. The model trains a network on a dataset of

experiences labelled with the actions that lead to particular return outcomes [Reymond et al., 2022]. This approach ensures stable learning by avoiding unstable iterative value updates and moving targets.

PCN involves using a single neural network that operates on input $s, \hat{h}, \hat{\mathbf{R}}$. These correspond to the current state of the environment, the time span in which the desired return should be achieved, and the desired return itself. By specifying an input, PCN outputs an action $a_i \in \mathcal{A}$ to reach this return in the specified time. Since the actions in the epidemiological setting are continuous, this is similar to supervised regression. PCN takes an input for which instances $(s, \hat{h}, \hat{\mathbf{R}})$ should get the correct prediction (a_i) [Reymond et al., 2022, 2024].

To enable regression, we need to build a dataset to learn from. At first, this is done by generating trajectories with random actions. These trajectories are stored and processed further. For each timestep in the trajectories, an instance is created from the cumulative reward from that point until the end of the episode. This entails storing $\langle s, \hat{h}, \hat{\mathbf{R}} \rangle = \langle s_t, h_t, \mathbf{R}_t \rangle \rightarrow a_t$ with $\mathbf{R}_t = \sum_{i=t}^T \gamma^i \mathbf{r}_i$. Storing this instance, ensures that choosing an action a_t at timestep t ensures cumulative reward $\mathbf{R}_t$ in h_t time steps, i.e., is the optimal action to choose [Reymond et al., 2022].

The neural network utilized by PCN, consists of multiple layers. These layers are used to transform the embedding of the state, horizon, and desired return to an executable action. The model is trained periodically by comparing the output of dataset samples to the predicted output. The policy is then updated according to the Mean Squared Error loss. Since PCN learns policies based on the executed trajectories, it is important to explore policies that yield higher returns. To avoid executing the same returns or not improving the trajectories, the model is conditioned on increasingly higher returns. More specifically, a random non-dominated return is selected from the current coverage set. By improving a single objective of this return within the boundaries of the known returns, PCN can execute different actions to achieve this newly formed return. Based on this new starting point $\langle s, \hat{h}, \hat{\mathbf{R}}' \rangle$, a new episode is executed to add the newly generated trajectories to the dataset according to the aforementioned method [Reymond et al., 2022].

To keep the dataset up-to-date with relevant policies, PCN needs to remove some. To achieve this, the model only stores policies that span different parts of the objective-space. From these policies, the model prunes trajectories based on a combination of the distance to the coverage set and how similar they are to comparable trajectories [Reymond et al., 2022].

4.2 PCN for Epidemiological Control

Given that epidemic control involves multiple goals or objectives, PCN is suitable for application in this setting. As Reymond et al. proposed, the network can be utilized to optimize multiple objectives such as hospitalizations, infections, and social burden. PCN is trained to learn a Pareto front that displays the different trade-offs of the corresponding policies. In this work, we continue to consider the epidemiological setting and expand it. We leverage PCN to include additional objectives that grasp the strain of imposed social restrictions on a more detailed level.

5 Environment

The epidemiological setting that PCN operates in, consists of multiple modules. The first and most fundamental is the epidemiological model by Abrams et al.. Their ODE-compartmental model provides an age-structured framework that runs the dynamics of the epidemic throughout time for ten age-groups (Table 1). This process is impacted by six social contact matrices $C_{\text{home}}, C_{\text{work}}, C_{\text{transport}}, C_{\text{school}}, C_{\text{leisure}},$ and C_{other} that describe the social environments for these age-groups. Alongside this model, we define a MOMDP to train a Pareto conditioned network [Reymond et al., 2022]. Collectively, these represent the MOBelCov environment [Reymond et al., 2024]. We employ this model and align it with the fairness framework proposed by Cimpean et al.. This allows us to investigate a selection of fairness notions to illustrate the effects of the MOMDP on the fairness (and vice versa) throughout the epidemic. In the following section, we describe the different aspects of this MOMDP and how the different segments represent the fairness-extended MOBelCov environment.

5.1 State Space

The different states of the epidemic are represented by the age-structured compartments in the compartmental model [Abrams et al., 2021]. The tuple $\{S_k, E_k, I_k^{\text{presym}}, I_k^{\text{asym}}, I_k^{\text{mild}}, I_k^{\text{sev}}, I_k^{\text{hosp}}, I_k^{\text{icu}}, H_k^{\text{new}}, D_k, R_k\}$ is defined for each age-group $k \in \{1 : [0, 10[, 2 : [10, 20[, 3 : [20, 30[, 4 : [30, 40[, 5 : [40, 50[, 6 : [50, 60[, 7 : [60, 70[, 8 : [70, 80[, 9 : [80, 90[, 10 : [90, \infty[\}$. H_k^{new} represents the newly hospitalized individuals for age-group k . This results in a matrix s_m with dimensions 10×11 :

$$s_m = \begin{bmatrix} S_1 & E_1 & I_1^{\text{presym}} & I_1^{\text{asym}} & I_1^{\text{mild}} & I_1^{\text{sev}} & I_1^{\text{hosp}} & I_1^{\text{icu}} & H_1^{\text{new}} & D_1 & R_1 \\ S_2 & E_2 & I_2^{\text{presym}} & I_2^{\text{asym}} & I_2^{\text{mild}} & I_2^{\text{sev}} & I_2^{\text{hosp}} & I_2^{\text{icu}} & H_2^{\text{new}} & D_2 & R_2 \\ S_3 & E_3 & I_3^{\text{presym}} & I_3^{\text{asym}} & I_3^{\text{mild}} & I_3^{\text{sev}} & I_3^{\text{hosp}} & I_3^{\text{icu}} & H_3^{\text{new}} & D_3 & R_3 \\ S_4 & E_4 & I_4^{\text{presym}} & I_4^{\text{asym}} & I_4^{\text{mild}} & I_4^{\text{sev}} & I_4^{\text{hosp}} & I_4^{\text{icu}} & H_4^{\text{new}} & D_4 & R_4 \\ S_5 & E_5 & I_5^{\text{presym}} & I_5^{\text{asym}} & I_5^{\text{mild}} & I_5^{\text{sev}} & I_5^{\text{hosp}} & I_5^{\text{icu}} & H_5^{\text{new}} & D_5 & R_5 \\ S_6 & E_6 & I_6^{\text{presym}} & I_6^{\text{asym}} & I_6^{\text{mild}} & I_6^{\text{sev}} & I_6^{\text{hosp}} & I_6^{\text{icu}} & H_6^{\text{new}} & D_6 & R_6 \\ S_7 & E_7 & I_7^{\text{presym}} & I_7^{\text{asym}} & I_7^{\text{mild}} & I_7^{\text{sev}} & I_7^{\text{hosp}} & I_7^{\text{icu}} & H_7^{\text{new}} & D_7 & R_7 \\ S_8 & E_8 & I_8^{\text{presym}} & I_8^{\text{asym}} & I_8^{\text{mild}} & I_8^{\text{sev}} & I_8^{\text{hosp}} & I_8^{\text{icu}} & H_8^{\text{new}} & D_8 & R_8 \\ S_9 & E_9 & I_9^{\text{presym}} & I_9^{\text{asym}} & I_9^{\text{mild}} & I_9^{\text{sev}} & I_9^{\text{hosp}} & I_9^{\text{icu}} & H_9^{\text{new}} & D_9 & R_9 \\ S_{10} & E_{10} & I_{10}^{\text{presym}} & I_{10}^{\text{asym}} & I_{10}^{\text{mild}} & I_{10}^{\text{sev}} & I_{10}^{\text{hosp}} & I_{10}^{\text{icu}} & H_{10}^{\text{new}} & D_{10} & R_{10} \end{bmatrix}$$

Furthermore, a representation of the social contact behavior and budgets are added to this epidemic state to reflect the MOBelCov state:

$$s = s_m \cup \hat{C} \cup b$$

With $\hat{C}$ the global social contact matrix and b the budget left for each action. We elucidate on the matrix and budget in Section 5.2.

5.2 Action Space

The action space of the environment is defined as restrictions on the six social contact matrices. These reflect the remaining proportion of social contacts allowed as the epidemic is simulated. The possible actions $a = [p_w, p_s, p_l]$ are limited to the 3-dimensional vector $[0, 1]^3$. The global social contact matrix is affected by the restrictions in the following way:

$$\hat{C} = C_{\text{home}} + p_w (C_{\text{work}} + C_{\text{transport}}) + p_s (C_{\text{school}}) + p_l (C_{\text{leisure}} + C_{\text{other}})$$

Each action is associated with a budget. The vector $b = [b_w, b_s, b_l]$ quantifies how often the corresponding actions can be modified over the course of the trajectory. The aforementioned PCN architecture is utilized to generate the distinct actions or social reductions [Reymond et al., 2022].

5.3 Transition Dynamics

The transitions of the Abrams compartmental model are incorporated in the MOBelCov environment. This entails that transitions are based on the current model state and social contact matrix at timestep t . The transitions per timestep are generated per week. As the social contact matrix is modified using action a_t , the transition incorporates the effects of the action. Abrams' model provides two methods to simulate the trajectories of the epidemic, a deterministic ODE model and a chain-binomial model. In this work, we focus on the ODE model as it compares to the chain-binomial model [Abrams et al., 2021].

5.4 Reward Function

To enable learning in a MOMDP, we need to define a multidimensional reward. In this case, the reward incorporates multiple interests. On the one hand, we aim to restrain the progression of the epidemic as much as possible. On the other hand, we strive to keep the social burden of the contact reductions reasonable.

To control the epidemic, we can utilize the negative attack rate on two fronts. First, we consider the infection attack rate as the difference in total new susceptible individuals over all age-groups between states [Reymond et al., 2024].

$$\mathcal{R}_{ARI}(s, a, s') = - \left(\sum_{k=1}^K S_k(s) - \sum_{k=1}^K S_k(s') \right).$$

Alongside this, the new hospitalizations can be considered as a reward for the hospitalization attack rate [Reymond et al., 2024].

$$\mathcal{R}_{ARH}(s, a, s') = - \sum_{k=1}^K H_k^{\text{new}}(s)$$

To define the social burden as a reward function, we utilize the modification of the social contact matrix as a result of the imposed contact reductions. This measure entails the total amount of lost contacts across the non-infected individuals (S, R) caused by the imposed restrictions [Reymond et al., 2024].

$$\mathcal{R}_{SB}(s, a, s') = \sum_{i=1}^K \sum_{j=1}^K (\hat{C} - C)_{ij} S_i(s) S_j(s) + \sum_{i=1}^K \sum_{j=1}^K (\hat{C} - C)_{ij} R_i(s) R_j(s),$$

6 Fairness

6.1 Social Welfare Functions

A way to ensure fairness in reinforcement learning systems is using social welfare functions [Fan et al., 2023, Siddique et al., 2020]. When fulfilling social welfare functions between groups or

individuals, each group or individual should be in the best possible situation. In most cases, artificial intelligence systems are built to optimize a specific goal or task. When optimizing such systems, often a cumulative measure is used to quantify the performance. However, this could lead to undesirable results in multi-objective situations [Fan et al., 2023, Hayes et al., 2022]. Satisfying social welfare functions resembles multi-objective reinforcement learning, where multiple goals should be considered [Hayes et al., 2022, Van Moffaert and Nowé, 2014]. Instead of utilizing a single scalar to represent the performance and rewards, a reward vector is used. When considering fairness in reinforcement learning, it should be included alongside the other objectives.

Former research explored maximizing social welfare functions, such as the Nash social welfare function [Kaneko and Nakamura, 1979, Fan et al., 2023] or the Gini evaluation function [Siddique et al., 2020]. Siddique et al. focus primarily on optimizing a single objective. Fan et al. consider fairness in a multi-objective environment, but only add a single social welfare function. Realistically, situations are more complex and additional factors and fairness aspects should be considered. For instance, an autonomous delivery system needs to balance fairness across individuals and geographical areas with optimizing service rates, cost, and time per person [Fan et al., 2023, Chen et al., 2023]. In the epidemiological setting, we wish to limit the amount of new hospitalizations while making sure that the needed social restrictions are as fair as possible for different age-groups. This highlights the importance and potential effectiveness of interpreting such multi-objective situations as a multi-objective reinforcement learning problem [Hayes et al., 2022, Reymond et al., 2024, 2022].

Optimizing multiple social welfare functions in a multi-objective reinforcement learning context has not been explored extensively. However, exploring fairness is crucial, as social justice is correlated with overall life satisfaction [Di Martino and Prilleltensky, 2020]. Therefore, it is important that autonomous decision-making systems are trained on multiple objectives that represent real-life circumstances before deployment.

6.2 Fairness in Machine Learning

Contrary to reinforcement learning, training in machine learning happens offline. Fairness in machine learning refers to the extent to which algorithms objectively evaluate individuals or groups in their predictions. It is possible that machine learning models will judge subjectively and discriminate by associating negative outcomes with certain sensitive features. In this setting, the model might associate young people with more social contacts and therefore restrict them more socially to keep an epidemic in check. However, when their risk of being seriously affected by the disease is low, they should not necessarily be the group that experiences the highest social burden. To this end, fairness notions are used to quantify and ensure objective and fair evaluations [Makhlouf et al., 2020]. The applicability and effectiveness of different fairness notions is dependent on the context [Grgić-Hlača et al., 2018]. These fairness measures can be categorized into group-based and individual-based. Group fairness ensures equal treatment across groups that share a sensitive attribute, while individual fairness ensures that similar individuals receive similar outcomes regardless of sensitive features. A limitation of group fairness is its focus on specific sensitive features. Despite group fairness notions being satisfied, individual discrimination may still occur [Dwork et al., 2012]. For example, ensuring equal treatment of men and women in a particular situation, does not warrant the avoidance of discrimination of one woman against the other. Individual fairness notions try to address these problems by quantifying whether similar individuals receive similar outcomes.

6.3 Criteria for Selecting Fairness Notions

Since fairness measures are context-dependent in both online and offline learning context, criteria for choosing fairness notions are crucial. This section outlines different considerations when determining suitable fairness notions [Makhlouf et al., 2020].

Ground truth: The availability of a reliable ground truth affects the calculation of fairness. Ground truth refers to objectively accurate data or outcomes. If absent, outcomes depend on subjective or potentially biased evaluations. Therefore, fairness notions should be determined in regard to the availability of a ground truth. Most fairness notions are based on comparing the chosen action to the objectively correct action [Makhlouf et al., 2020]. In our case, this approach is not feasible, as we do not have information on the correct social restrictions at a certain point in an epidemic. Additionally, comparing the action to the correct action is more difficult when these actions are continuous.

Unreliable outcomes: When outcomes are inferred from subjective evaluations or potentially biased processes, they are considered unreliable. This entails the absence of a ground truth. This criterion is applicable to the epidemiological environment. We do not have the ability to label specific actions or restrictions as the correct approach. In the case of unreliable outcomes, fairness notions that do not consider the true values are more appropriate.

Emphasis on recall versus precision: Depending on context, the emphasis of positive instances can be on precision or recall. Precision focuses on how accurate the positive outcomes are, while recall prioritizes capturing as many true positive cases as possible. Depending on the meaning of a positive instance, the emphasis on precision or recall influences the definition of fairness. Since labelling actions as correct or incorrect is impossible, this criterion is not applicable in our setting.

Fixed versus adaptive thresholds: Thresholds determine whether an instance is considered positive or negative. Fixed thresholds remain constant, while floating thresholds adjust dynamically based on context-specific factors, such as economic conditions. The choice between fixed and adaptive thresholds influences the choice of appropriate fairness measures. In this setting, this would translate to the severity of the epidemic. Based on the magnitude of the outbreak, an adaptive threshold could be utilized to classify social restrictions. However, without the evaluation of actions at hand, this criterion is ineffective in this case.

Likelihood of intersectionality: Intersectionality occurs when individuals face discrimination on multiple features simultaneously. If intersectional discrimination is probable, individual fairness measures should be prioritized to capture nuanced discrimination more effectively. Since the main distinctive feature in this setting is age, this criterion is difficult to evaluate.

6.4 Existing Fairness notions

6.4.1 Group Fairness Notions

Group fairness notions can be used to see whether two or more groups are treated equally. In this subsection, we discuss some notable examples.

Statistical parity: Statistical parity requires equal rates of a certain positive outcome across all groups [Makhlouf et al., 2020]. Statistical parity does not consider the ground truth or base rates. Therefore, it is well-suited for scenarios in which there is an unreliable outcome. This would imply that statistical parity is suitable for use in the epidemiological setting. However, translating this to practice, this would correspond to restricting all age-groups in the same way. Considering the groups are not affected the same by the epidemic, this would be an undesirable outcome. Therefore, statistical parity is unsuitable for use.

Equal opportunity: Equal opportunity ensures equal proportions of correctly predicted positives across groups. It focuses on minimizing the unfair negative classification of individuals across groups when the actual outcome should be positive. Therefore, it is suitable where false negatives carry significant ethical or practical consequences. This would translate to comparing whether the different age-groups had the same proportions of correct actions applied to them. Since this entails considering the ground truth, it is ineffective in this case.

Predictive equality: Predictive equality is similar to equal opportunity, but focuses on reducing wrongful inclusion (false positives) and is most suitable for contexts where false positives are particularly harmful. Similarly to equal opportunity, this fairness notion requires a ground truth, which makes it unusable in our epidemiological environment.

6.4.2 Individual Fairness Notions

This section discusses some fairness notions regarding individuals. Group fairness notions mostly look at whether groups of people are treated fairly compared to each other. However, this can potentially cause discrimination on a more fine-grained level going unnoticed. To this end, we utilize individual fairness notions. These consider the fair treatment of individuals compared to others.

Causal discrimination: This fairness notion is particularly suitable when considering intersectionality. Satisfying this measure, ensures that no individuals are discriminated against. If two individuals only differ in one feature, causal discrimination expects them to have the same outcome [Makhlouf et al., 2020]. In the epidemiological setting, we only have a select number of features that distinguish the different age-groups. Since this fairness notion requires more detailed information, it is unsuitable in our scenario.

Fairness through unawareness: This individual fairness notion generalizes causal discrimination. Satisfying this, measure states that similar individuals have a similar prediction [Makhlouf et al., 2020, Dwork et al., 2020, Vanspringel et al., 2023]. To compute this fairness notion, we consider individual i and individual j with feature vectors v_i and, v_j respectively. These vectors are associated with outcome probability distributions $M(v_i)$ and $M(v_j)$. Assuming $d(v_i, v_j)$ as the similarity distance between the candidates and $D(M(v_i), M(v_j))$ the distance between their outcome distributions. Fairness through unawareness is satisfied when the following equation is satisfied for each possible pairwise combination of individuals i and j :

$$D(M(v_i), M(v_j)) \leq d(v_i, v_j)$$

Concretely, the satisfaction of this fairness notion can be quantified using the following equation:

$$-1 + \frac{1}{n} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}, j \neq i} d(i, j) - D(M_i || M_j) = 0$$

In this work, we consider a variant of this individual fairness notion, which is discussed in Section 5.

7 Fairness in Reinforcement Learning

Considering the unsuitability of the majority of existing fairness notions, we must broaden our perspective. Similar to Siddique et al., we aim to ensure optimal as well as fair solutions. To this end, the aforementioned authors utilize the existing *generalized Gini social welfare function* [Weymark, 1981]. Analogously, we introduce new fairness notions that resemble the working of such functions. The proposed metrics quantify the welfare of the population according to the current social state. In this setting, the fairness notions quantify how different age-groups experience the social restrictions as a result of the epidemiological conditions.

7.1 Fairness Framework

To exploit these fairness notions for usage in RL, we utilize a fairness framework developed by Cimpean et al.. The framework extends the existing MDP process even further to flexibly incorporate fairness notions and assess them over different window sizes. Leveraging this *fMDP*, we can expand the epidemiological setting to include fairness signals for PCN to interpret. This extension results in the reward vector being extended to include this fairness feedback. Utilizing this framework, we can effectively integrate existing fairness measures or implement variations on existing ones. The *fMDP* allows us to discuss distinct scenarios and combinations of fairness notions and subsequently investigate trade-offs made.

7.2 Epidemiological Fairness

It is vital that epidemics are addressed fairly. In difficult times, the population should work towards a common goal. Therefore, it is essential that social restrictions do not cause division among the masses. To achieve unity in settings like this, we must impose the necessary restrictions while ensuring fairness. Since the previously mentioned measure $\mathcal{R}_{SB}$ describes social burden broadly across the whole population, the imposed social restrictions do not consider the diversity of the population. Therefore, considering these social strains in a more fine-grained way, could support unity among individuals. To this end, we propose two fairness notions that capture the social restrictions for different age-groups and compared to each other. These measures enable us to leverage PCN to consider the differences in nature of a disease. Particularly, when differing age-groups do not suffer comparable adverse effects from getting sick, they should not necessarily experience parallel negative effects on other aspects. The proposed fairness notions particularly take the age of the individuals and their susceptibility to the serious consequences of infection into account. To further improve our understanding of the environment, we examine the total contacts of each age-group in the different areas of restriction.

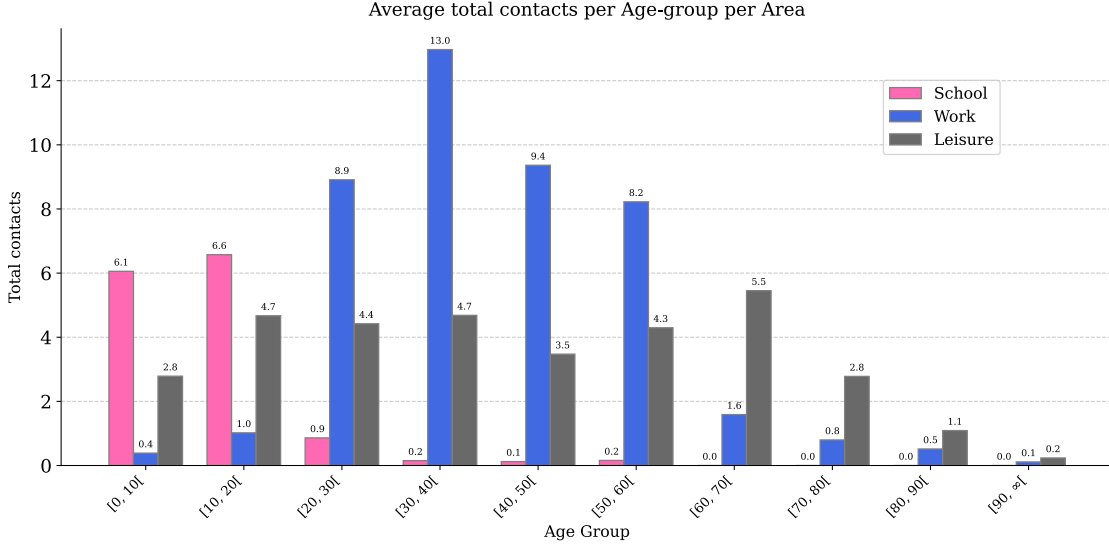


Figure 3: The total contacts made in one week per area for each age-group.

Figure 3 denotes how the different contacts are distributed across the different areas of restriction for the age-groups. In this figure, we observe that the majority of school contacts are made in the two youngest groups. The contacts at work are more spread out, but concentrated in the middle-aged groups. The contacts in leisure are the most spread out across all age-groups, with the most for the 60 – 70 year olds. This graph illustrates which age-groups are primarily impacted when imposing social restrictions in the different areas.

7.3 Social Burden Fairness

To capture fairness on a similar level as the aforementioned group fairness notions in Section 6.4.1, we propose a fairness notion similar to a social welfare function. Fairness notion *social burden fairness* (*SBF*) is an adaptation of $\mathcal{R}_{SB}$. It extends social burden by considering the hospitalization risk of the different age-groups. It sums the amount of lost contacts together among the uninfected population and divides it by the hospitalization risks. The fairness notion is denoted by the following equation:

$$\begin{aligned}
 \mathcal{F}_{SBF}(s, a, s') = & - \left(\sum_{i=1}^K \sum_{j=1}^K [(\hat{C} - C)_{ij} S_i(s) \frac{1}{h_i} + (\hat{C} - C)_{ij} S_j(s) \frac{1}{h_j}] \right. \\
 & \left. + \sum_{i=1}^K \sum_{j=1}^K [(\hat{C} - C)_{ij} R_i(s) \frac{1}{h_i} + (\hat{C} - C)_{ij} R_j(s) \frac{1}{h_j}] \right)
 \end{aligned} \tag{11}$$

where h_i denotes the age-specific hospitalization risk. These risks are derived from the parameters in the epidemiological model by Abrams et al.. As depicted in Figure 1, the transition to being severely ill (I_{sev}) and consequently hospitalized or recovering (R) is determined by age-specific probabilities ψ and δ_2 . Since the transition to compartment I_{sev} ultimately implies hospitalization, the vector ψ is adopted as the hospitalization risks. The hospitalization risks per group are depicted in Figure 4.

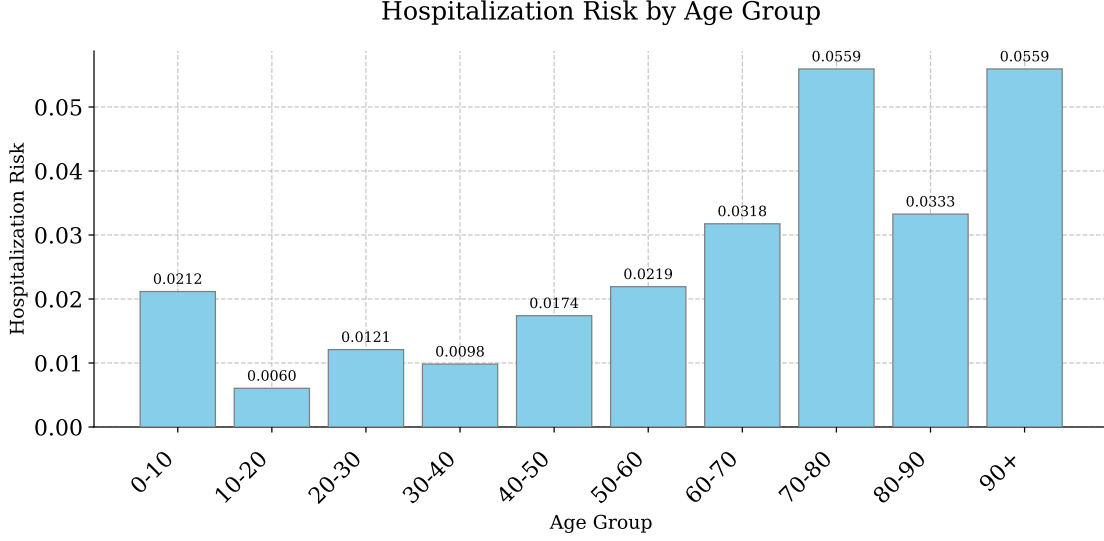


Figure 4: The probability or risk of being hospitalized for each age-group.

Figure 4 can be combined with Figure 3 to inspect how the contacts are weighed by *social burden fairness*. By dividing the average total contacts per area by the hospitalization risk, we can observe how PCN targets the different groups by imposing restrictions.

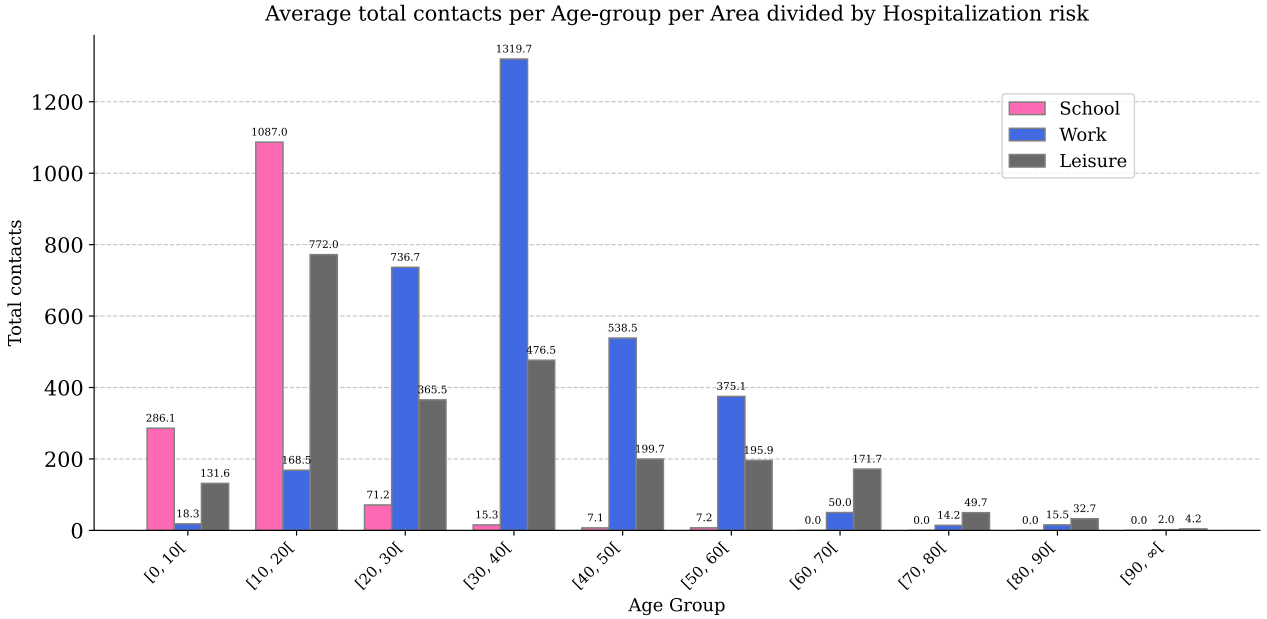


Figure 5: The average total contacts per age-group per area in one week divided by their respective hospitalization risks.

Figure 5 denotes which groups affect the value of $\mathcal{F}_{SBF}$ more significantly. We identify that age-groups $[10, 20[$ and $[30, 40[$ have the potentially highest impact factor for the fairness notion. Therefore, the measures affecting these groups impact the fairness notion the most. However, imposing mitigation measures affects all the age-groups in that area. Therefore, this effect requires PCN to find trade-offs in which mitigation measures to choose.

7.3.1 Exploring Social Burden Fairness

To gain a more in-depth understanding of $\mathcal{F}_{SBF}$, we investigate its behavior according to various executed actions. Since reasoning about negative values is intuitively more difficult, we depict

the fairness notions as *unfairness*. Therefore, the smaller values, closer to zero, are considered more fair. Figure 6 depicts the absolute values of the fairness notion when different contrasting actions are executed.

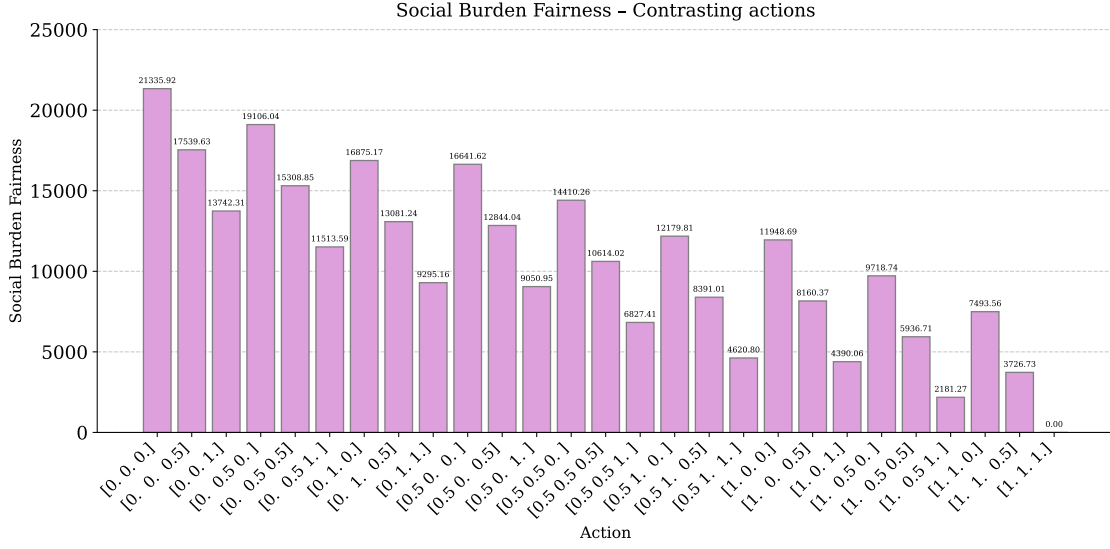


Figure 6: The absolute unfairness values of $\mathcal{F}_{SBF}$ for different contrasting actions.

We observe a general downwards or increasingly fair trend when social restrictions are loosened. This is expected as the amount of lost contacts is lower in this situation. However, we can identify a varying behavior when actions affect different areas. To analyze this, we observe a selection of four contrasting actions. To illustrate the difference between these actions, we display the amount of lost contacts per age-group. This enables us to investigate which social restrictions are considered more unfair. Figure 7 denotes the proportionally lost contacts per age-group for four contrasting actions.

Proportionally lost contacts per age group for actions considering Social Burden Fairness

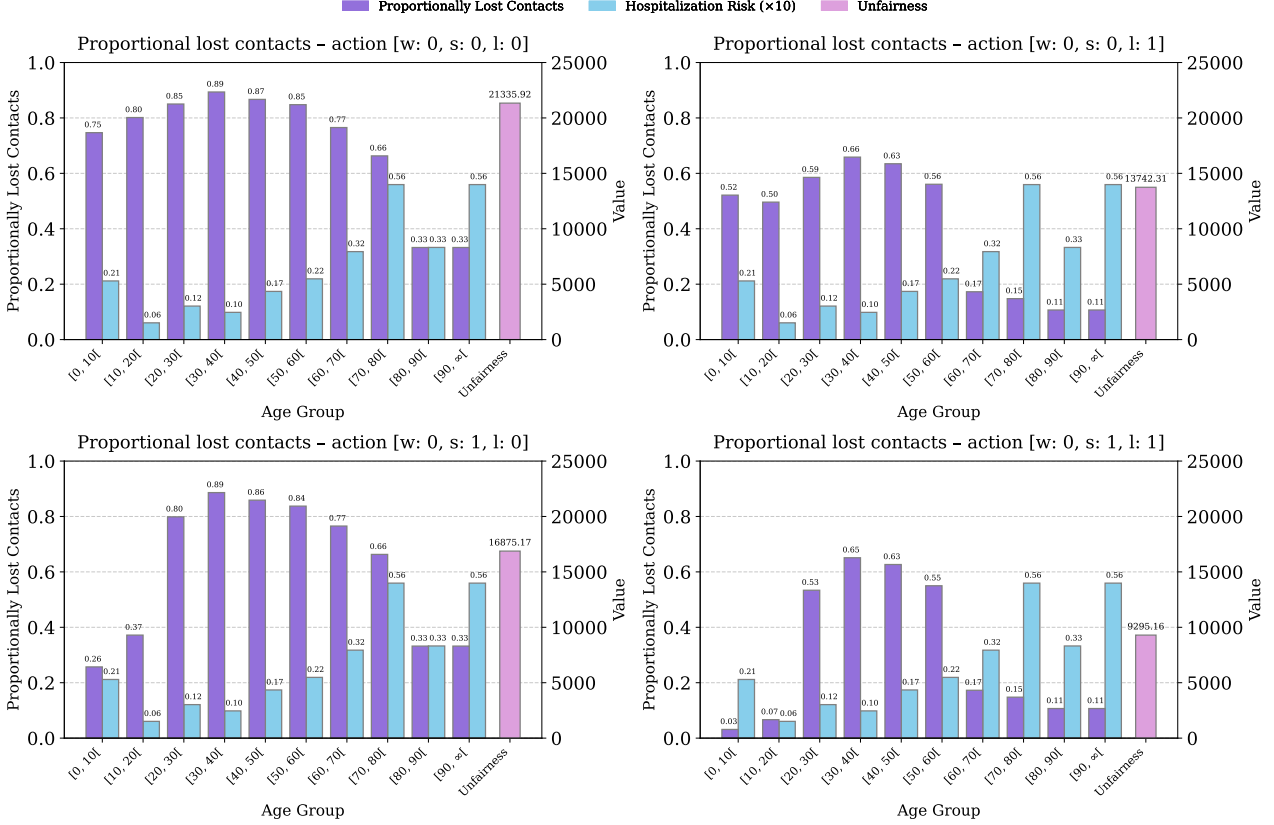


Figure 7: The average proportionally lost contacts for each age-group for a selection of distinct actions (purple) along with the hospitalization risks (blue) and unfairness (pink) measured for the action. The hospitalization risks are scaled up to improve readability and improve expressiveness of the graph. A lower value for the unfairness indicates a more fair outcome for the specific action.

In Figure 7, we observe the different values of $\mathcal{F}_{SBF}$ for different social restrictions. Here, we observe that the unfairness is the highest when mostly younger and less hospital-prone age-groups are restricted. Moreover, the unfairness remains high when social restrictions are more concentrated on the middle-aged groups. The unfairness decreases when the restrictions of less risk-prone age-groups are loosened in comparison to the older age-groups. This insight confirms the expected behavior of $\mathcal{F}_{SBF}$. Moreover, it provides a way to employ PCN to consider hospitalization risks along with the social burden in choosing the best possible actions at each moment. Different patterns of executed actions can be found in Appendix A.1.

7.4 Age-Based Fairness Through Unawareness

Fairness should not only be measured in a coarse-grained manner. It can alternatively be computed on a more fine-grained level. In classical fairness computation, this distinction is generally done by group fairness notions and individual fairness notions. In this setting, the different age-groups are the maximum reachable level of detail. However, this still leaves the opportunity to display fairness from a different perspective. To this end, we can adapt an existing individual fairness notion to fit this particular scenario. We alter the individual fairness notion *fairness through unawareness* for use in the continuous action space and on the age-level granularity in the absence of a ground truth. As mentioned in Section 6.4.2, fairness through unawareness is an individual fairness notion that states that similar individuals should have a similar outcome [Makhlouf et al., 2020, Dwork et al., 2012]. We adapt this fairness notion

to fit the different age-groups. Satisfying *age-based fairness through unawareness* (ABFTU) requires similar age-groups to be socially restricted similarly.

We consider age-groups i, j from the set of age-groups with a distance of $d(i, j)$, based on their hospitalization risks. To quantify the needed outcome distributions M_i and M_j , we consider the effects or outcome of the actions on each age-group. We define M_i and M_j as the probability distributions of actions impacting age-groups i and j . Concretely, we inspect how the actions lead to a distribution of social contact reductions. The distance between these probability distributions is defined as $D(M_i||M_j)$ and is calculated using the Kullback-Leibler divergence [Kullback and Leibler, 1951]. To achieve perfect individual fairness between two age-groups, the difference of their hospitalization risks should be equal to the difference between their social contact reductions. This would indicate that the impact on their social behavior is justifiable by their difference in hospitalization risks. Overall, for all age-groups this would imply the following equation to be zero:

$$\mathcal{F}_{ABFTU} = -\frac{1}{n(n-1)} \sum_i \sum_{j, j \neq i} |d(i, j) - D(M_i||M_j)| \quad (12)$$

with n the number of pairwise combinations of age-groups that can be made.

To compute the fairness notion, we assume that the distance metric is an age-dependent factor determined based on domain expertise [Makhlouf et al., 2020]. Particularly, the difference in hospitalization risk between two age-groups. As the continuous actions are imposed on the entire population, we require a way to differentiate between the concrete effects on the different age-groups to determine their outcome distributions. To this end, we inspect the resulting contact restrictions for each age-group on each social contact matrix as a result of the continuous actions. By transforming this to contact reduction distributions over the aforementioned six areas in Section 5.2 we have access to M_i and M_j . This enables us to compare the social contact reductions in distributions M_i and M_j for age-groups i and j to determine $D(M_i||M_j)$. Specifically, we compute the needed metrics with the following equations:

$$d(i, j) = |h_i - h_j|$$

$$D(M_i||M_j) = \text{KL}(\mathbf{R}_i, \mathbf{R}_j)$$

We elucidate on the details of the computation in Section 7.4.1.

7.4.1 Algorithm

The computation of *age-based fairness through unawareness* involves multiple steps. First, the amount of reduced contacts on the six social contact matrices from Section 5.2 is totaled for each age-group k .

$$\mathbf{R}_k = \left[\sum_{k'} (\hat{C}^a - C^a)[k][k'] \right]_{a \in A}$$

with $A = \{\text{home}, \text{work}, \text{transport}, \text{school}, \text{leisure}, \text{other}\}$ the different areas of reduction impacted by restricting work, school, and leisure. Next, these arrays and the array of hospitalization risks are transformed to probability distributions. This ensures a better scale and preserves the relative proportions.

$$\mathbf{h} = \mathbf{h} / \sum_k h[k]$$

$$\mathbf{R}_k = \mathbf{R}_k / \sum_{a \in A} r_a$$

where r_a denotes the reduction per area for age-group k . Furthermore, two matrices are constructed for efficient computation. The first matrix represents d and contains all pairwise subtractions of differences in hospitalization risks between the age-groups.

$$\mathbf{d}[i][j] = |h_i - h_j|, \quad \text{for all age-groups } i, j$$

The second matrix contains all possible pairwise combinations of $D(M_i || M_j)$. These distances are computed using the Kullback-Leibler divergence [Kullback and Leibler, 1951].

$$\mathbf{D}[i][j] = \text{KL}(\mathbf{R}_i, \mathbf{R}_j), \quad \text{for all age-groups } i, j$$

with KL the Kullback-Leibler divergence:

$$\text{KL}(P || Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

Finally, to ensure the comparability of $d(i, j)$ and $D(M_i || M_j)$, we apply min-max normalization to the matrix $\mathbf{D}$ and scale these values to the range of $\mathbf{d}$.

$$\mathbf{D}[i][j] = \frac{\mathbf{D}[i][j] - \mathbf{D}^{\min}}{\mathbf{D}^{\max} - \mathbf{D}^{\min}} \times (\mathbf{d}^{\max} - \mathbf{d}^{\min}) + \mathbf{d}^{\min}$$

with $\mathbf{D}^{\max}$ and $\mathbf{D}^{\min}$ the maximum and minimum values determined by a brute-force search using 10^6 uniquely generated continuous actions ². The eventual computation of *age-based fairness through unawareness* can then be described using matrix operations:

$$\mathcal{F}_{ABFTU} = -\frac{1}{n(n-1)} \sum_i \sum_{j, j \neq i} |\mathbf{d}[i][j] - \mathbf{D}[i][j]|$$

7.4.2 Exploring Age-Based Fairness Through Unawareness

To assess how $\mathcal{F}_{ABFTU}$ functions in different settings, we perform similar experiments as for *social burden fairness*. First, we analyze how the fairness notion and the KL-divergence behave for several contrasting actions. Again, we quantify the fairness as *unfairness* for visualization and reasoning purposes. A higher value for *age-based fairness through unawareness* indicates a more unfair treatment.

² $\mathbf{D}^{\min} = 0$; $\mathbf{D}^{\max} = 28$

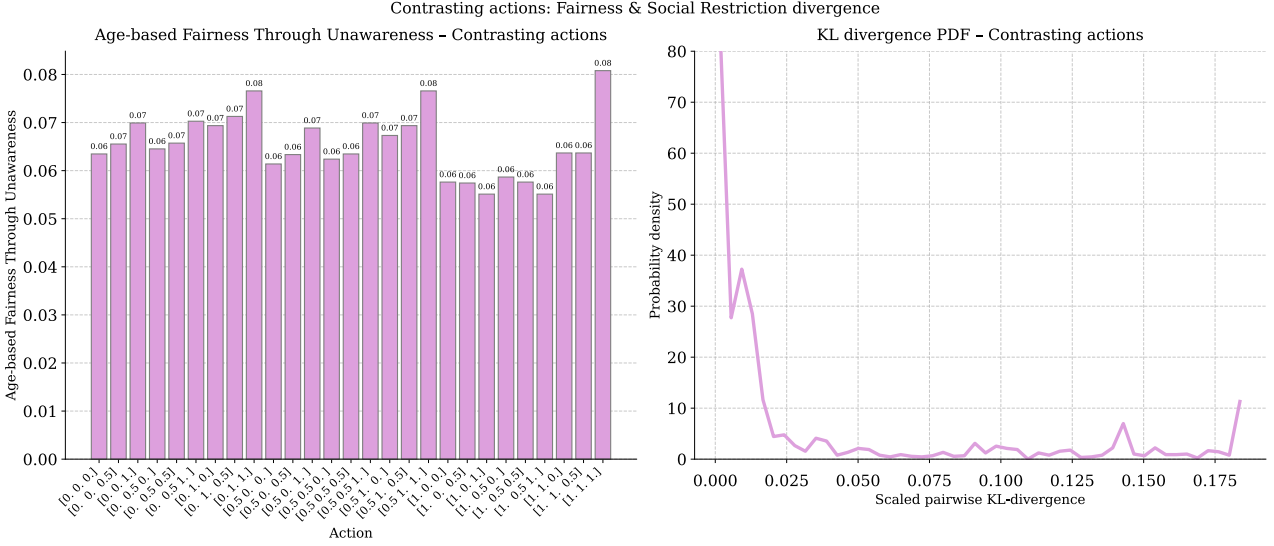


Figure 8: The absolute unfairness values of $\mathcal{F}_{ABFTU}$ (left) and an average probability density plot (right) for different contrasting actions.

In Figure 8 we identify how the fairness notion does not fluctuate drastically for the different actions. Since the continuous actions target multiple age-groups simultaneously, the mapping of the three-dimensional mitigation measures to the ten age-groups causes a limited difference in the KL-divergence. Additionally, the hospitalization risks are constant throughout the simulation. This bounds the possible range of the unfairness, as the different age-groups cannot be individually restricted and have a completely separate experience based on their risk. Despite this, a pattern is identifiable. We observe that in general, the unfairness increases when the social restrictions on leisure are loosened. This is due to the leisure impacting more distinct age-groups as we can observe in Figure 3. When restrictions are imposed on school and work, these apply to inherently similar age-groups. Therefore, the restrictions increase the difference in treatment between similar age-groups less. In contrast, restricting leisure contacts causes more age-groups to be affected. Since the different age-groups do not have the same distribution of social contacts, this causes the distance between their restrictions to increase. To provide additional insight, we plot the proportionally lost contacts per age-group, along with the average difference between hospitalization risks and KL-divergences for a selection of contrasting actions.

Proportionally lost contacts per age group for actions considering Age-based Fairness Through Unawareness

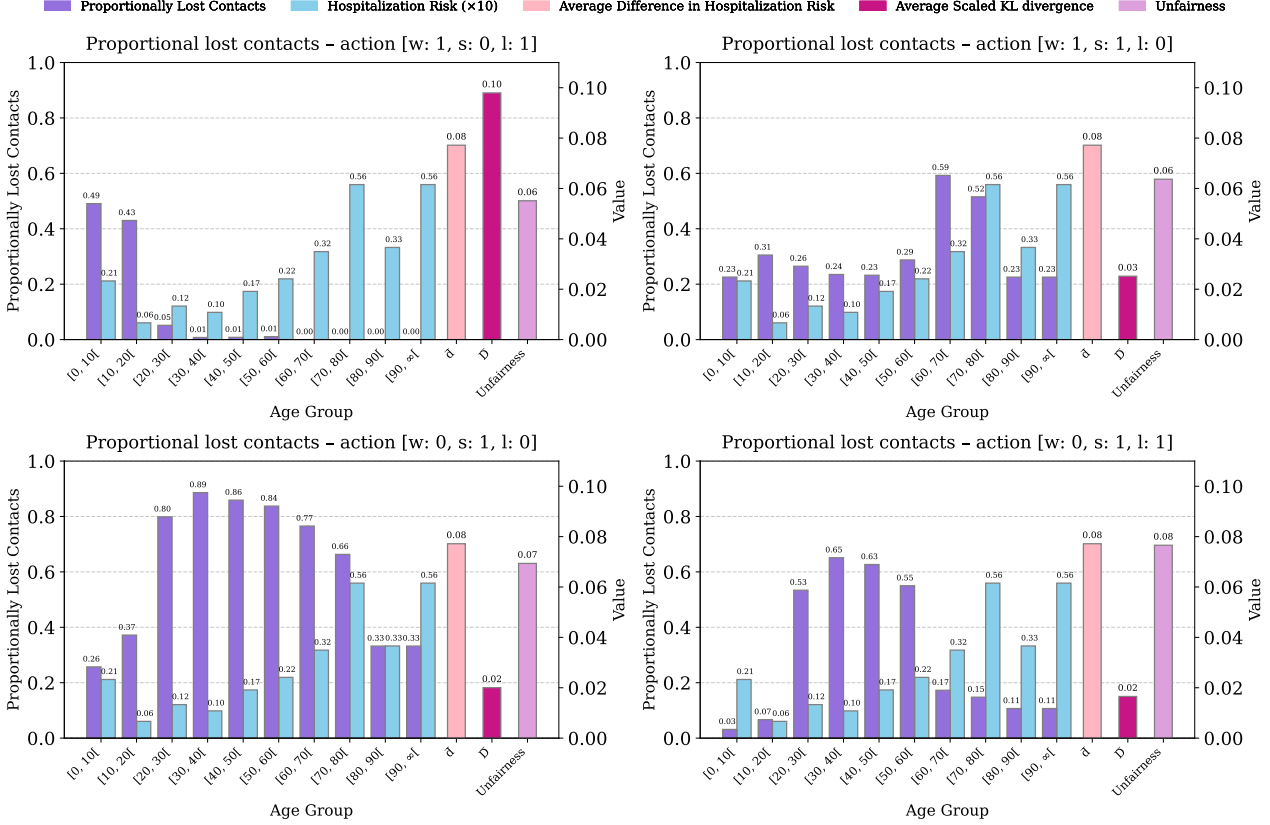


Figure 9: The average proportionally lost contacts for each age-group for a selection of distinct actions (purple) along with the hospitalization risks (blue) and unfairness (pink) measured for the action. Additionally, the average distances between the hospitalization risks (light pink) and the KL-divergences (magenta) are shown. The hospitalization risks are scaled up to improve readability and improve expressiveness of the graph. A lower value for the unfairness indicates a more fair outcome for the specific action.

In Figure 9 we observe that the unfairness increases when there is a larger discrepancy between the average hospitalization difference and the average disparity in social restrictions. In the bottom two subplots, we find that when groups with a similar hospitalization risk have a disproportional amount of lost contacts, the overall unfairness increases. On the other hand, when the restrictions follow the general trend of the hospitalization risks, we note that the unfairness decreases. These patterns indicate that this fairness notion could serve as a metric for PCN to control fair treatment of age-groups relative to each other. Different action patterns can be found in Appendix A.2. To better visualize how the age-groups differ in terms of their hospitalization, we plot a heatmap to display the disparities.

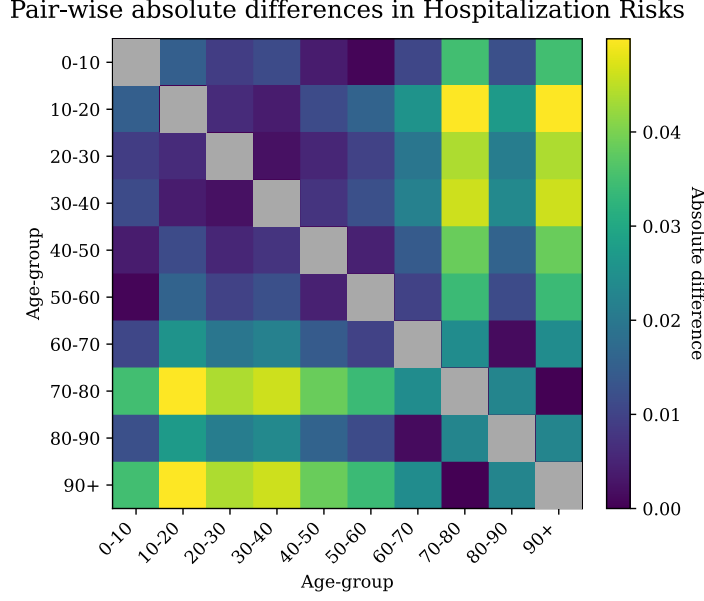


Figure 10: The absolute intergroup differences in hospitalization risks.

Figure 10 visualizes how the distance varies between the groups in this setting. This illustration facilitates reasoning about the impact of the incorporation of $\mathcal{F}_{ABFTU}$. Using this visualization, we can assess whether the similar age-groups receive a similar treatment due to the mitigation measures.

8 Experiments

To proceed, we incorporate the proposed fairness notions as objectives. We consider five distinct scenarios and inspect how PCN performs and handles epidemiological situations. The focus in the experiments is on the new hospitalizations and the included fairness notions. Using these contrasting objectives provides an interesting perspective to see how PCN adapts to these rewards.

8.1 Experimental Scenarios

We integrate the rewards in different ways to gain insight into their influence in the decision-making of PCN. These are chosen from a vector with all the returned values that illustrate the epidemiological trajectory. The complete feedback signal by the environment contains the following elements:

$$[\mathcal{R}_{ARI}, \mathcal{R}_{ARH}, \mathcal{R}_{SBW}, \mathcal{R}_{SBS}, \mathcal{R}_{SBL}, \mathcal{R}_{SB}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$$

with the opportunity to focus on the social burden per area using $\mathcal{R}_{SBW}$, $\mathcal{R}_{SBS}$, and $\mathcal{R}_{SBL}$. We select different subsets of this reward vector to investigate how it changes the epidemiological trajectories generated by PCN. Next, we discuss these distinct formats of the setup.

8.1.1 Social Burden

At first, we inspect the reproduction of the results by Raymond et al.. Here, we set up an important baseline that serves as a reference point for the ensuing results. In this scenario, PCN is only instructed to optimize on both hospitalization attack rate and the overall social

burden of the population. More specifically, the model focuses on the $\mathcal{R}_{ARH}$ and $\mathcal{R}_{SB}$ entries in the reward vector.

8.1.2 Social Burden & Social Burden Fairness

In this scenario, we start incorporating the aforementioned fairness notions. This step is necessary as it illustrates whether *social burden fairness* impacts social burden, and whether the fairness notion could potentially serve as a standalone objective and achieve similar results. In this case, the considered objectives are $\mathcal{R}_{ARH}$, $\mathcal{R}_{SB}$, and $\mathcal{F}_{SBF}$.

8.1.3 Social Burden Fairness

Furthermore, we use social burden fairness only as an objective along the hospitalization attack rate. This experiment verifies whether this objective can serve as a sole measure of social discomfort to optimize. Moreover, it allows us to compare the setting that optimizes $\mathcal{R}_{ARH}$ and $\mathcal{R}_{SB}$ with the setting of objectives $\mathcal{R}_{ARH}$ and $\mathcal{F}_{SBF}$. This informs us whether this measure would be an applicable improvement to the naïve social burden and achieve similar results.

8.1.4 Age-based Fairness Through Unawareness

This set of experiments considers the fairness notion that considers less the overall impact on age-groups. In this setting, PCN focuses more on comparable treatment of similar groups. More specifically, age-groups with a comparable hospitalization risk should have a comparable reduction in social contacts. This fairness notion therefore impacts less how the groups themselves are affected, and more how they are impacted compared to each other. PCN is optimized on $\mathcal{R}_{ARH}$ and $\mathcal{F}_{ABFTU}$ in this case.

8.1.5 Social Burden Fairness and Age-based Fairness Through Unawareness

In this setting, we let PCN consider both fairness notions. Decision-making should be done considering both the overall social impact on age-groups and how they are affected compared to each other. These results give us insight in how compatible the fairness notions are with each other and how they impact the overall treatment of the age-groups by PCN. Here, the model considers the objectives $\mathcal{R}_{ARH}$, $\mathcal{F}_{SBF}$, and $\mathcal{F}_{ABFTU}$.

8.2 Overview

We provide an overview of the configuration settings of the environments.

Parameter	Value
Model	ODE
Steps	500,000
Scale	[800000, 11000, 50, 20, 50, 120, 240000, 1]
Budget	2, 3, 4, 5, ∞
Window Size	1 week
Episode Length	17 weeks

Table 2: Hyperparameters used for each set of experiments.

All experiments were repeated for 10 different seeds and adopted the same learning parameters as Raymond et al.. To assess fairness, we repeatedly utilize a window of 1, which translates to one week. The usage of generative AI is discussed in Section 12.

9 Results

We consider the five distinct scenarios to explore how PCN can be applied in epidemiological environments. At first, we inspect the baseline, where we reproduce and discuss the results obtained by Reymond et al.. Next, we elaborate on these results by extending the objectives with the *social burden fairness* that attempts to capture a similar purpose as the baseline. Following, we inspect whether this fairness notion can be used as a standalone objective and yields similar results. Subsequently, we introduce the age-based variant of *fairness through unawareness*. We explore the impact of adding this fairness notion as an objective, followed by a scenario in which both *social burden fairness* and *age-based fairness through unawareness* are optimized. For each experiment, we observe the interpolated Pareto fronts of policies found by PCN over the different seeds with the standard deviation. Displaying the standard deviation was omitted in the 3D plots for visibility purposes. These plots visualize the unfairness achieved in regard to the caused hospitalizations during the epidemiological trajectories. Accompanying this plot, we discuss radar plots generated by the fairness framework by Cimpean et al. to gain insight into the allocation of the distinct objectives. The policies in the radar plots are scaled individually to fit a range of $[-1, 0]$. This allows us to properly display the different trade-offs made for different objectives. To ensure transparency, the minimum and maximum values of the objectives over all policies are indicated on the axes. The illustrated policies are the best for the optimized objectives (solid), complemented with a few uniformly chosen policies in the Pareto fronts (dashed). The results depicted in the radar plots are marked correspondingly in the plot of the Pareto fronts.

Since PCN is trained by specifying a desired return for an episode, this allows us to experiment with different goals. To compare different versions of PCN, we examine a selection of trajectories for each setting. To obtain these trajectories, we specify different goals or desired returns and assess how PCN behaves. For each setting, we illustrate the effective mitigation measures chosen by PCN when it has a budget of 5 changes for each restriction. Here, we first consider the policies that perform best and worst in terms of the hospitalization attack rate. The amount of available ICU beds were the driving factor in deciding which mitigation measures to take [Faes et al., 2021, Belgian Health Care Knowledge Centre (KCE), 2024]. Therefore, we additionally consider a policy that achieves a peak level of approximately 1,500 ICU patients over the epidemiological trajectory of seventeen weeks. Furthermore, we repeatedly compare different goals that cause different amounts of peak ICU patients to see how the different combinations of objectives impact the epidemiological trajectories generated by PCN. Furthermore, we assess how the distinct settings impact different age-groups by examining how the contacts of some different groups are effectively reduced.

Additionally, we discuss the impact of utilizing *social burden fairness* as a measure for social distress compared to social burden.

9.1 Social Burden & Social Burden Fairness

First, we inspect the reproduction of the results obtained by Reymond et al.. We visualize the Pareto front by plotting the social burden according to the hospitalizations caused at the end of an epidemiological trajectory of seventeen weeks. More specifically, we utilize a version of PCN that is optimized for both $\mathcal{R}_{ARH}$ and $\mathcal{R}_{SB}$. Additionally, we evaluate the different trade-offs made by PCN using the radar plots generated by the fairness framework by Cimpean et al..

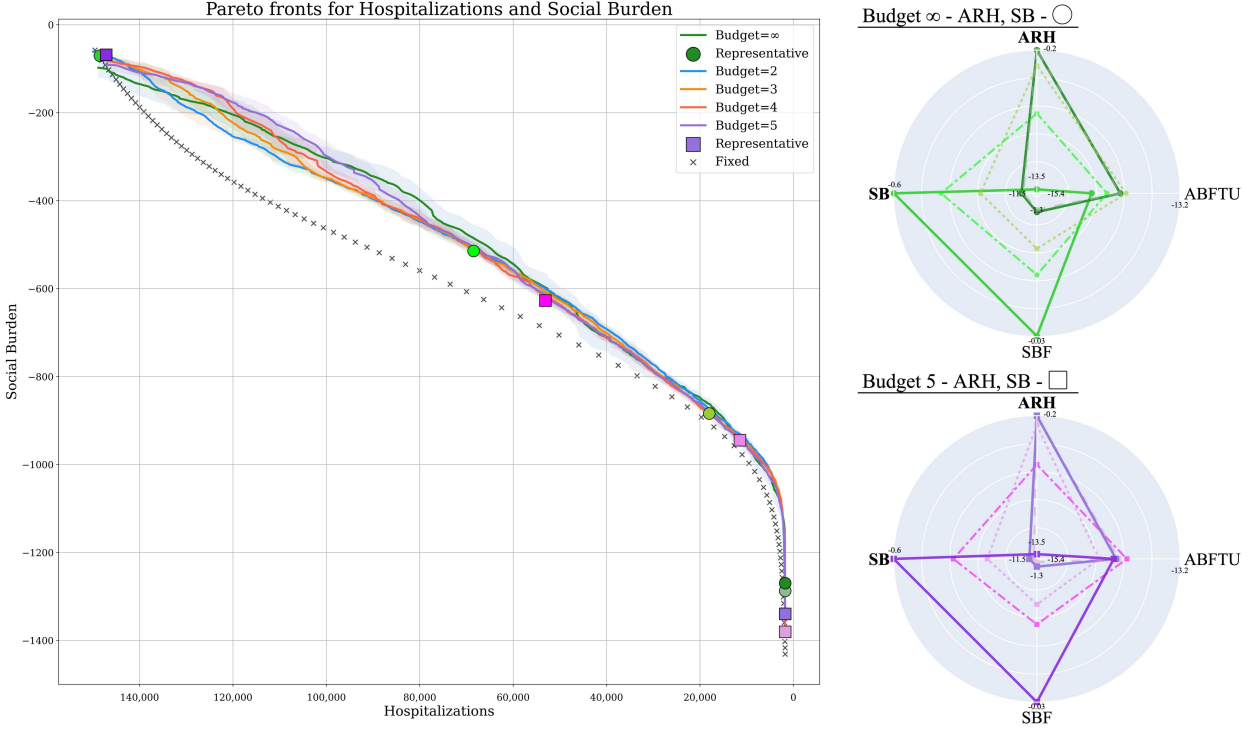


Figure 11: Interpolated Pareto fronts for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ for budgets $\infty, 2, 3, 4, 5$. Radar plots visualize the different trade-offs achieved by PCN for budgets ∞ and 5 while optimizing on given objectives. Each objective is shown scaled individually to a range $[-1, 0]$ along with the actual minimum and maximum values. The best performing policies for the optimized objectives are drawn solid and the supplementary policies dashed with less opacity. Each policy is marked in the corresponding color on the Pareto fronts. The x-axis visualizes the amount of new hospitalizations at the end of an epidemiological episode of 17 weeks, while the y-axis displays the cumulative average lost contacts for all age-groups. The outcome for fixed policies is marked by crosses.

Figure 11 displays the superior performance of PCN compared to the results possible by following fixed policies during an epidemic. At nearly each possible outcome regarding hospitalizations, PCN outperforms the fixed policies in terms of social burden. In the radar plots, we notice the various trade-offs made. When policies are leaning more towards maximizing $\mathcal{R}_{ARH}$, we observe the detrimental effects on social burden. Conversely, when the policies put less strain on the population regarding social life, we identify less optimal results regarding the hospitalizations. Next, we inspect some epidemiological trajectories to inspect how PCN changes its restrictions to optimize the desired objectives. We will discuss the policies that perform best and worst in terms of hospitalizations, along with a policy that causes a maximum peak of 1,500 ICU patients in an epidemiological trajectory of seventeen weeks.

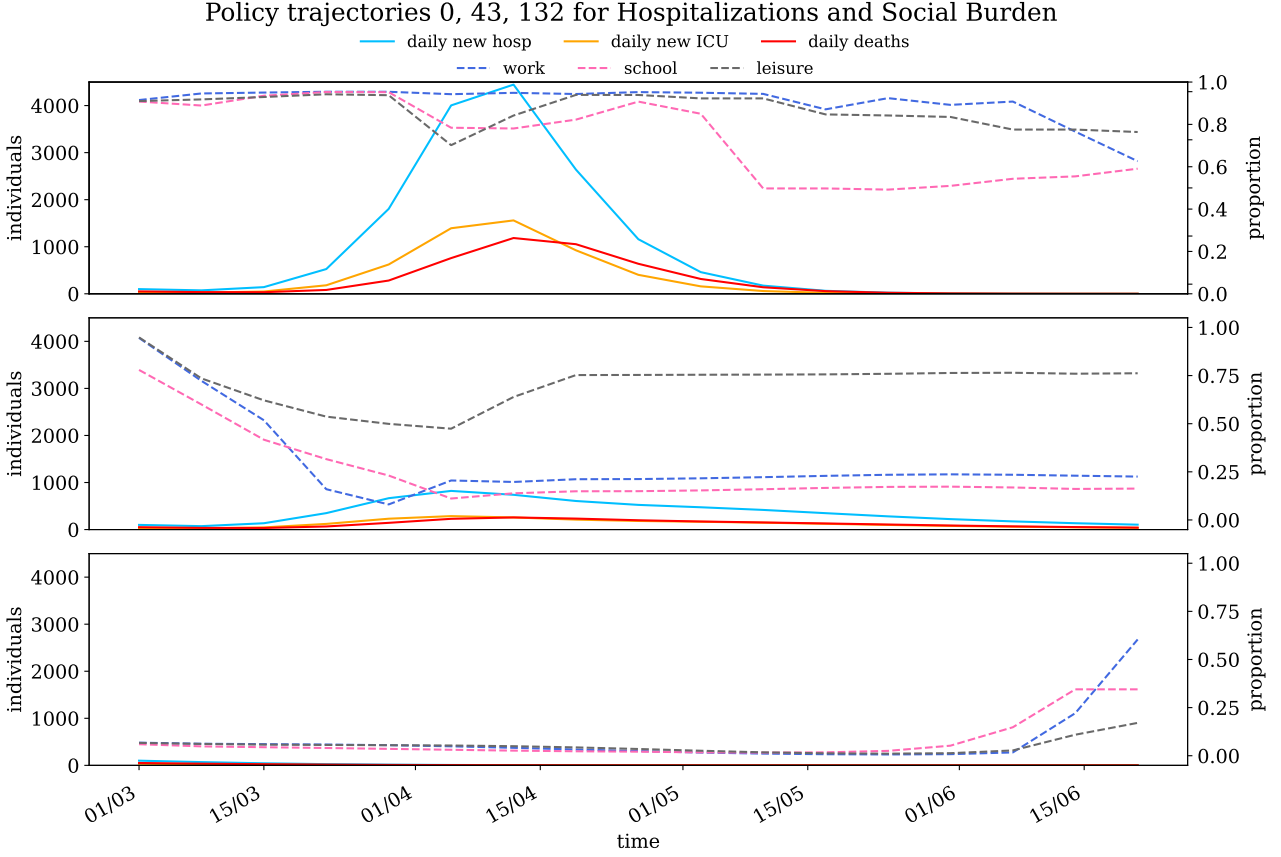


Figure 12: Epidemiological trajectories of policies achieving the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom). PCN is trained on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ with a budget of 5 changes. The epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

Figure 12 denotes how the choice of actions affects the progression of the epidemic. In the top trajectory, we observe the detrimental effects of minimal social restrictions. We identify the catastrophic effects, as there would be more than 4,000 daily deaths for multiple weeks. In the bottom panel, we observe that the imposed social restrictions avoid this death toll, but put great social restrictions on the population. Since no social contacts would be allowed for multiple months, this would require great social effort of the population [Onyeaka et al., 2021]. In the center panel, we note how PCN would restrict the population to ensure a maximal spike of 1,500 ICU patients in seventeen weeks. Work and school are restricted the most over the complete trajectory, while leisure less. The restrictions are collectively increased in the first month and then loosened to keep a consistent level of restriction for the rest of the trajectory.

9.1.1 Social Burden Fairness

To explore the effects of the proposed fairness notions, we start by incorporating *social burden fairness* as an objective. Here, PCN is optimized on $\mathcal{R}_{ARH}$, $\mathcal{R}_{SB}$ and $\mathcal{F}_{SBF}$ initially and subsequently only on $\mathcal{R}_{ARH}$ and $\mathcal{F}_{SBF}$. First, we look at the Pareto fronts compared to a fixed policy when the fairness notion is added as an objective. This way, we observe the effect of the fairness notion on the existing social burden metric and whether it impacts it.

Pareto fronts for Hospitalizations, Social Burden and Social Burden Fairness
(Fair = 0)

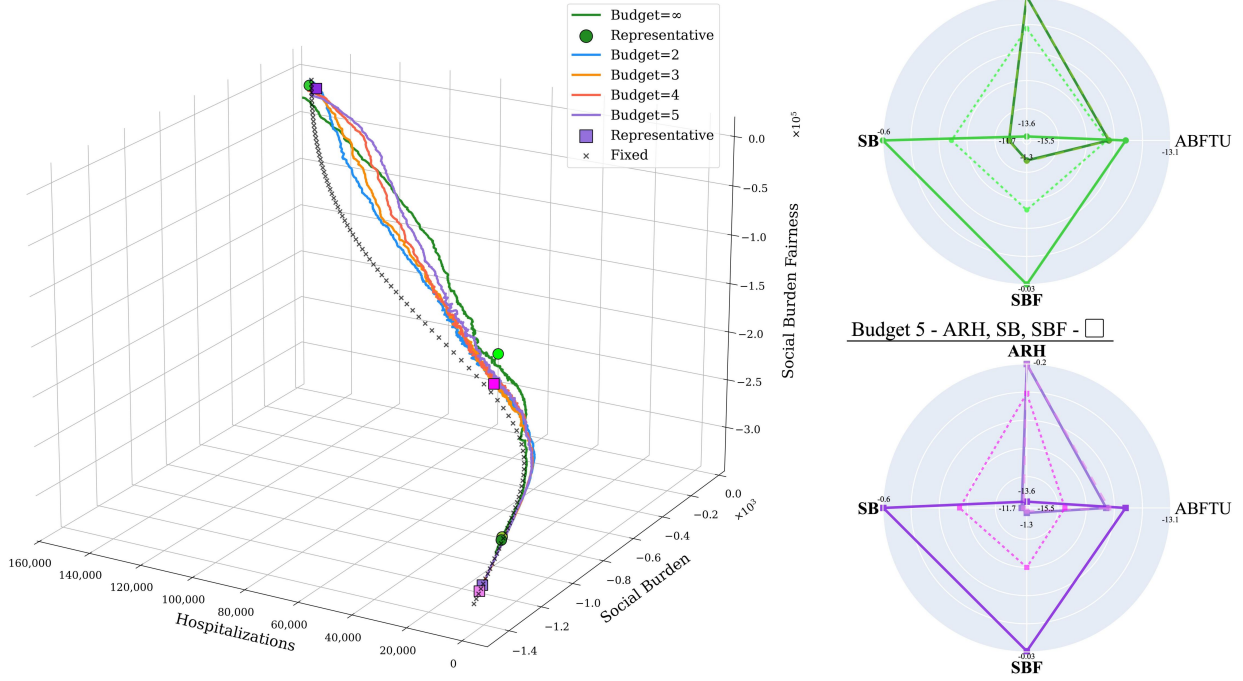


Figure 13: Interpolated Pareto fronts for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}, \mathcal{F}_{SBF}]$ for budgets $\infty, 2, 3, 4, 5$. Radar plots visualize the different trade-offs achieved by PCN for budgets ∞ and 5 while optimizing on given objectives. Each objective is shown scaled individually to a range $[-1, 0]$ along with the actual minimum and maximum values. The best performing policies for the optimized objectives are drawn solid and the supplementary policies dashed with less opacity. Each policy is marked in the corresponding color on the Pareto fronts. The x-axis visualizes the amount of new hospitalizations at the end of an epidemiological episode of 17 weeks, and the y-axis displays the cumulative average lost contacts for all age-groups. The z-axis shows the scaled down unfairness, measured by social burden fairness. If the fairness notion measures a result closer to zero, this indicates a more fair trajectory for the population. The outcome for fixed policies is marked by crosses. Displaying the standard deviation was omitted for visibility purposes.

Figure 13 shows that the policies learned by PCN ultimately outperform the fixed policies. Both in terms of social burden and *social burden fairness*, the model puts less strain on the population to achieve the same number of hospitalizations. In the radar plots, the policies learned for budget ∞ span a larger area. This is due to the model having more freedom to adjust the actions taken throughout the epidemic. The radar plots confirm that $\mathcal{R}_{SB}$ and $\mathcal{F}_{SBF}$ are not conflicting objectives. The policies that perform best on these objectives, yield similar or even identical results. Additionally, we observe the impact of the bigger budgets. The policies performing best on hospitalization attack rate, span a larger area for greater budgets. This illustrates how these budgets allow for PCN to improve on other objectives while achieving this result for hospitalizations.

Policy trajectories 0, 72, 258 for Hospitalizations, Social Burden and Social Burden Fairness

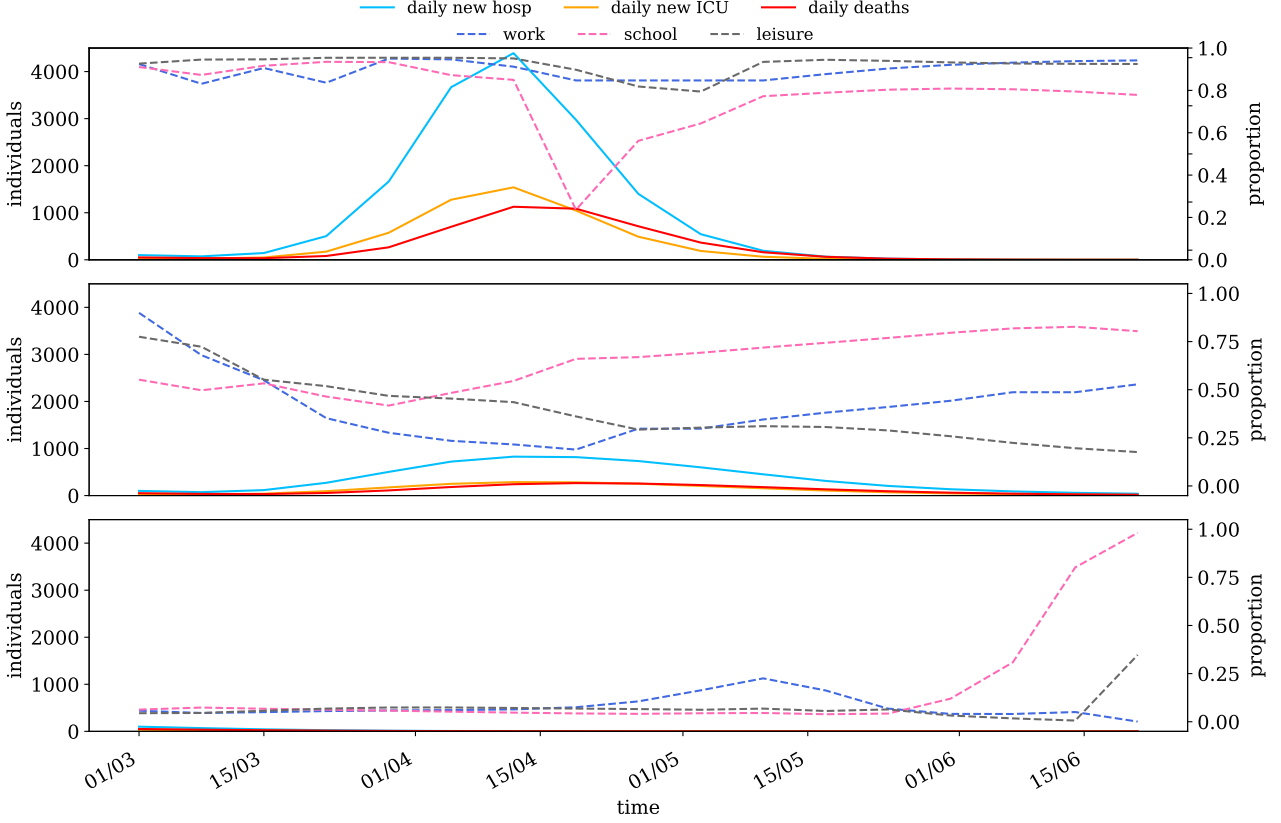


Figure 14: Epidemiological trajectories of policies achieving the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom). PCN is trained on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}, \mathcal{F}_{SBF}]$ with a budget of 5 changes. The epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In the trajectories depicted in Figure 14, we identify a similar pattern for the hospitalizations, deaths, and ICU patients. In the top trajectory, in which PCN causes the highest hospitalization attack rate, there is a large peak in the hospitalizations around one month into the epidemic trajectory. In the setting that causes the least hospitalizations, we observe a drastic loosening of the imposed restrictions in the schools at the end of the trajectory. This is due to PCN considering the *social burden fairness*. The incorporation of this metric causes the model to first lower the restrictions in this area to optimize $\mathcal{F}_{SBF}$ along with social burden. School going people are the best group in the population for this, since the restriction on schools targets almost exclusively lower hospitalization risk groups with many social contacts. Additionally, it impacts the unfairness less as it does not include many age-groups with higher risks. In the trajectory that achieves a peak of 1,500 ICU patients, we note a similar shift in the restrictions. In this trajectory, PCN first restricts work the most and then prefers to loosen the social restrictions put on the schools and work. Moreover, the restrictions on leisure life are the highest at the end. Next, we investigate whether $\mathcal{F}_{SBF}$ could serve as a standalone objective.

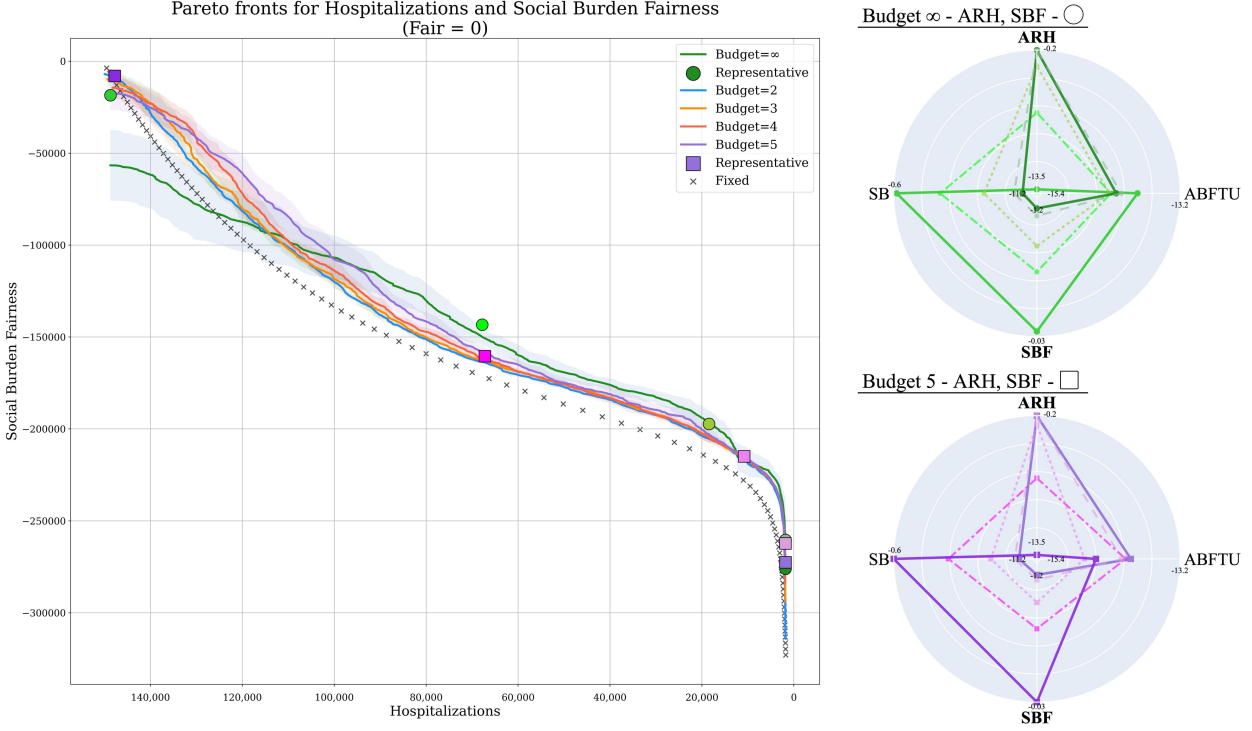


Figure 15: Interpolated Pareto fronts for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ for budgets $\infty, 2, 3, 4, 5$. Radar plots visualize the different trade-offs achieved by PCN for budgets ∞ and 5 while optimizing on given objectives. Each objective is shown scaled individually to a range $[-1, 0]$ along with the actual minimum and maximum values. The best performing policies for the optimized objectives are drawn solid and the supplementary policies dashed with less opacity. Each policy is marked in the corresponding color on the Pareto fronts. The x-axis visualizes the amount of new hospitalizations at the end of an epidemiological episode of 17 weeks, and the y-axis shows the scaled down unfairness, measured by social burden fairness. If the fairness notion measures a result closer to zero, this indicates a more fair trajectory for the population. The outcome for fixed policies is marked by crosses.

Figure 15 displays similar results as before. PCN mostly outperforms the fixed policies. Only when achieving a high number of hospitalizations, the unlimited budget performs worse. In the radar plots, we again recognize the slightly worse performance of lower budgets. The lower budget policies are less capable of spanning a larger area to perform better on multiple policies. In purple, for the policies with budget 5, we note that the policies perform overall better in the Pareto front compared to budget 2. Furthermore, we identify how the proposed *social burden fairness* is compatible with regular social burden. The policy that achieves the best result for the fairness notion, is able to hit similarly high values for social burden. This indicates that optimizing $\mathcal{F}_{SBF}$, induces similar effects for the social burden on people in general. Additionally, the higher budgets perform better on $\mathcal{F}_{ABFTU}$, which suggests that both fairness notions might be moderately compatible. Next, we look at some trajectories to assess how the social restrictions are approached concretely.

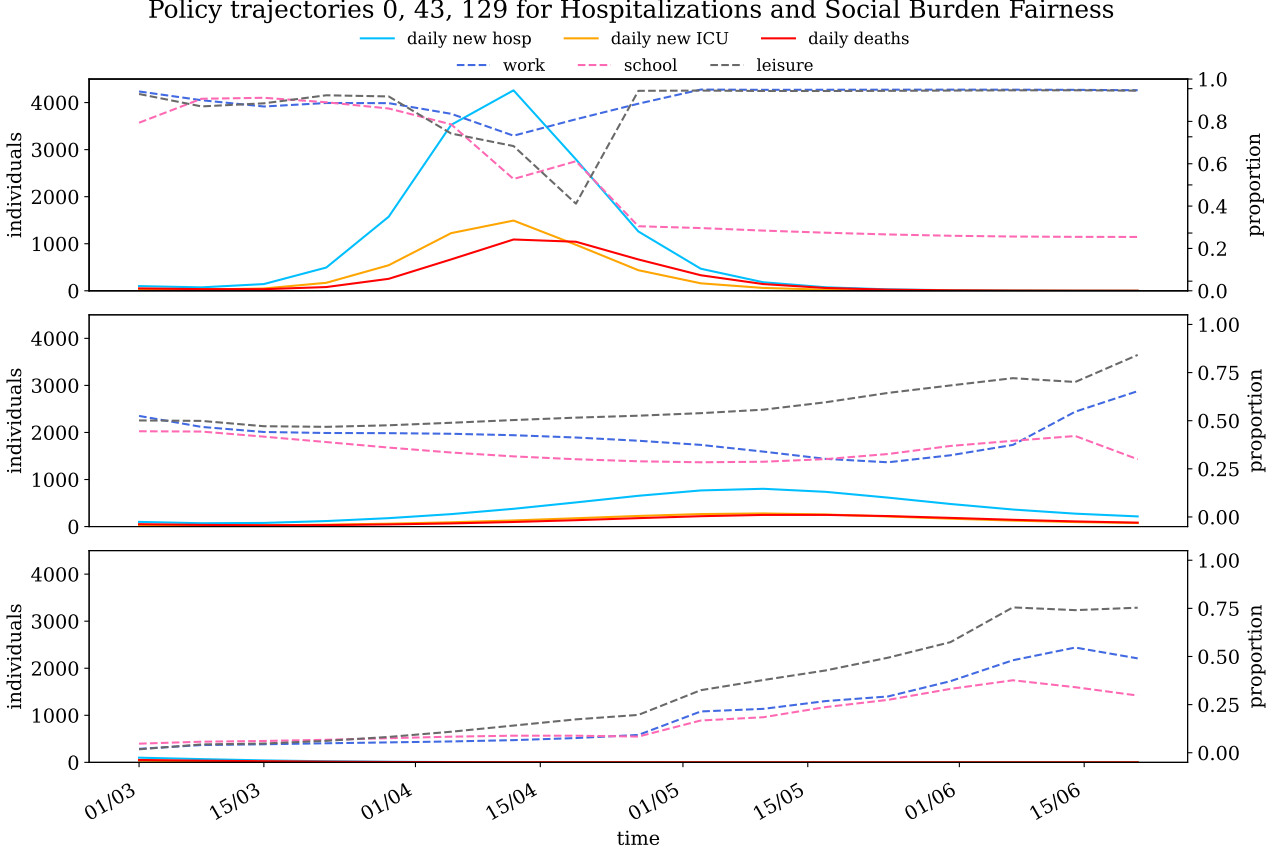


Figure 16: Epidemiological trajectories of policies achieving the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom). PCN is trained on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ with a budget of 5 changes. The epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

Figure 16 depicts the three distinct trajectories when $\mathcal{F}_{SBF}$ is used as a standalone objective alongside the hospitalization attack rate $\mathcal{R}_{ARH}$. Interestingly, in the trajectory causing the most hospitalizations, the schools remain mostly restricted at the end. This is due to the limited budget not allowing PCN to adapt this mitigation to the current epidemiological state. Additionally, work and leisure contacts are restricted slightly in the beginning, but then loosened at around two months in. This is due to work and leisure impacting the lost contacts according to hospitalization risks considerably (Figure 5). When investigating the policy that performs best on hospitalization attack rate, we identify the impact of $\mathcal{F}_{SBF}$. In this trajectory, we observe that all areas are severely restricted in the first two months. Then the mitigation measures are gradually loosened with a preference to work and leisure. This shows the effect of PCN trying to optimize the fairness notion, favoring less lost contacts in work and leisure as a result of considering the hospitalization risks. In the middle panel, PCN restricts work and school somewhat consistently throughout the trajectory. Only in the last month of the observed episode, there is a gradual loosening of the restriction on leisure contacts and ultimately of school and work contacts.

9.1.2 $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$

Next, we compare the previous results of utilizing social burden as a metric of social distress to using *social burden fairness*. To examine the differences correctly, we align two policies that cause the maximum common peak of current ICU patients and the minimum common peak. Additionally, we compare two policies that cause a maximum spike of approximately 2,000 and

1,000 ICU patients in both settings. To facilitate the examination of the trajectories, we plot the total lost intergroup contacts for the different policies over the full trajectories. We first examine the common maximum and minimum peaks of ICU patients in Figure 17.

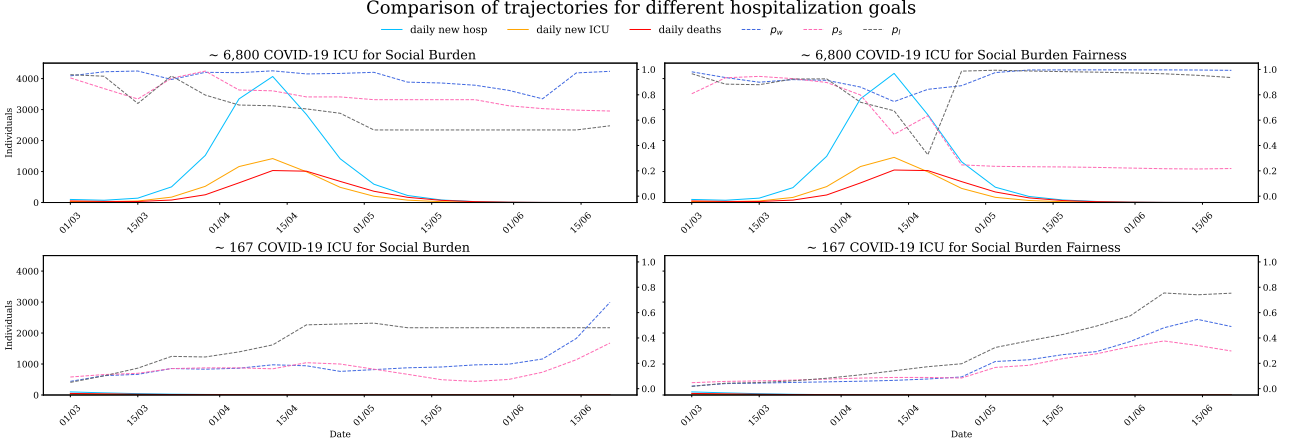


Figure 17: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings changes. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In Figure 17, we observe the impact of utilizing *social burden fairness*. In the top trajectory, we note that the right policy restricts the schools more at the end, while restricting leisure and work barely. On the left, PCN restricts leisure the most, and schools and work less. In the two bottom trajectories, we note more interesting results. Both policies achieve a minimal peak of ICU patients, but there is a difference in approach to the mitigation measures. On the left, PCN keeps a consistent heavy restriction on the schools and work until the end of the trajectory. On the right, the restrictions are loosened faster and more gradual throughout the episode, with a preference of leisure contacts. This approach could reduce the mental effort required of the population, while still keeping the hospitalizations to a minimum. Next, we compare the lost intergroup contacts for these policies.

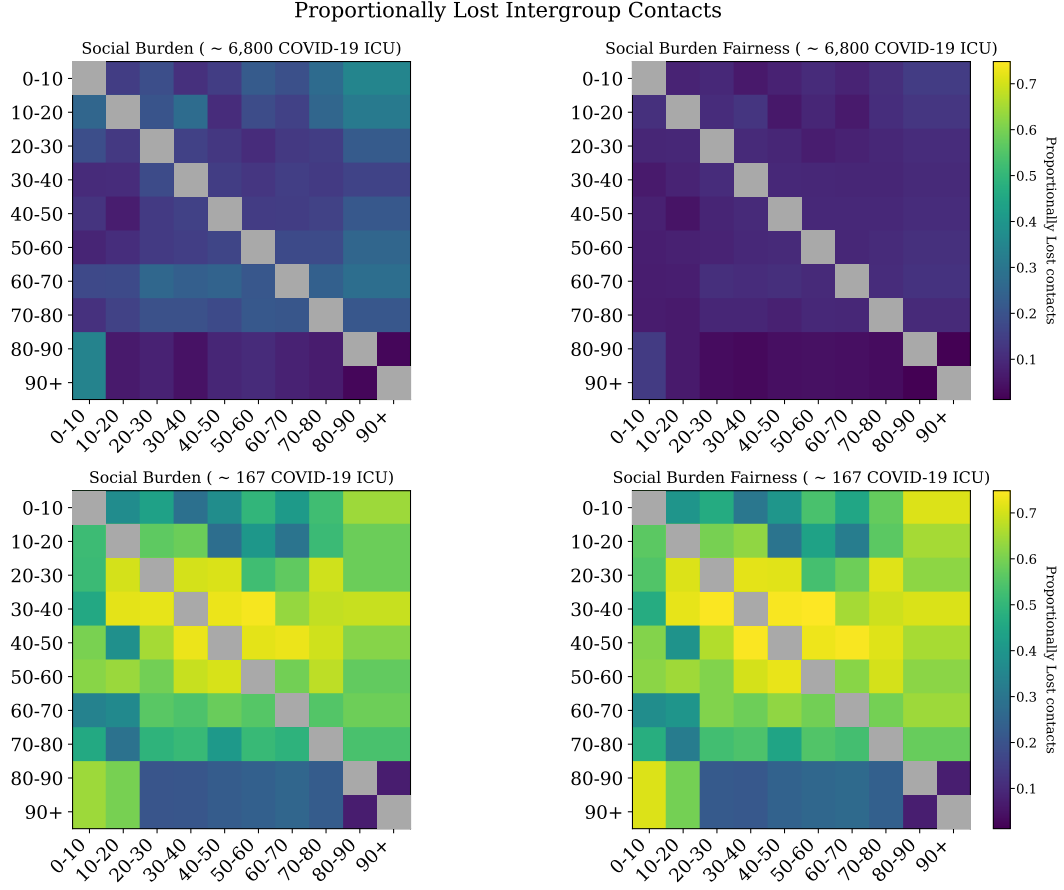


Figure 18: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings.

Figure 18 shows different results for the policies that cause a peak of approximately 6,800 ICU patients. The left heatmap shows a higher average social restriction across the age-groups, compared to the right heatmap. However, the overall social restriction remains limited. This is to be expected as there are no drastic restrictions in these policies. In the bottom policies, there are less pronounced differences. On the left, PCN restricts the different age-groups slightly more evenly compared to the right. Here, we observe a minor focus on the groups with a lower hospitalization risk. This subtle difference is due to the strictness of the policies. Since the policies ensure a low hospitalization attack rate, the number of different restriction patterns to minimize the social strain simultaneously are limited. Next, we compare trajectories that cause a peak of around 2,000 and 1,000 peak ICU patients for both settings.

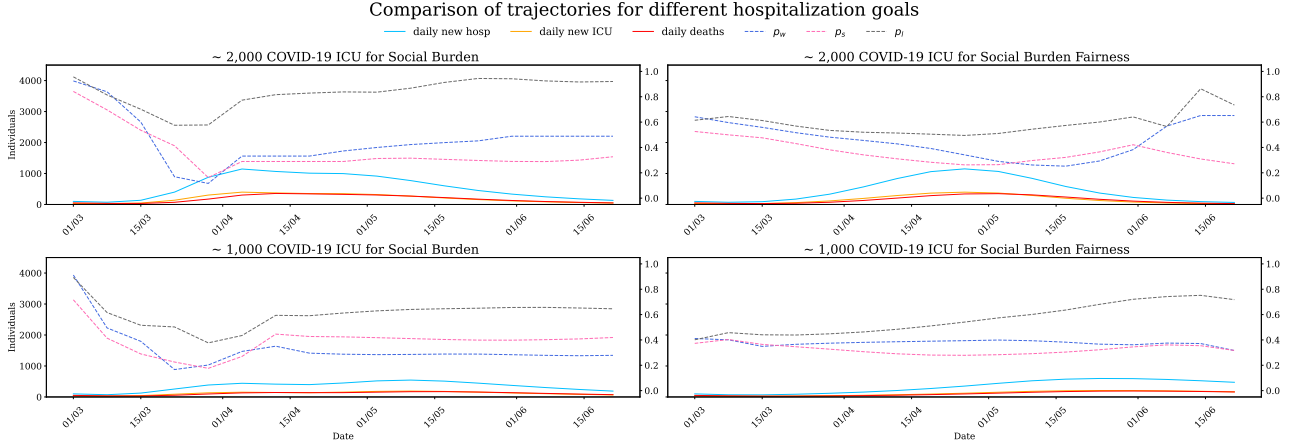


Figure 19: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In the trajectories depicted in Figure 19, we note some differences as well. In a looser policy, we observe that the policy on the left restricts work and school the most. Leisure is increasingly restricted at the beginning and gradually loosened to almost unrestricted towards the end. On the right, the restrictions are more balanced and increased gradually. Towards the end, these measures are loosened with a preference for work and leisure. On the bottom, the left panel shows three similar restriction curves. All three social restrictions are increased in the first month, to then be decreased at roughly the same pace. Afterward, the social restrictions remain consistent at their level until the end of the trajectory, with leisure contacts being restricted the least. The right panel displays a different approach. PCN restricts the schools and work somewhat consistently throughout the complete trajectory. Leisure is restricted similarly at the beginning, but then gradually loosened towards the end of the epidemiological episode. We inspect the proportionally lost intergroup contacts for these policies in Figure 20.

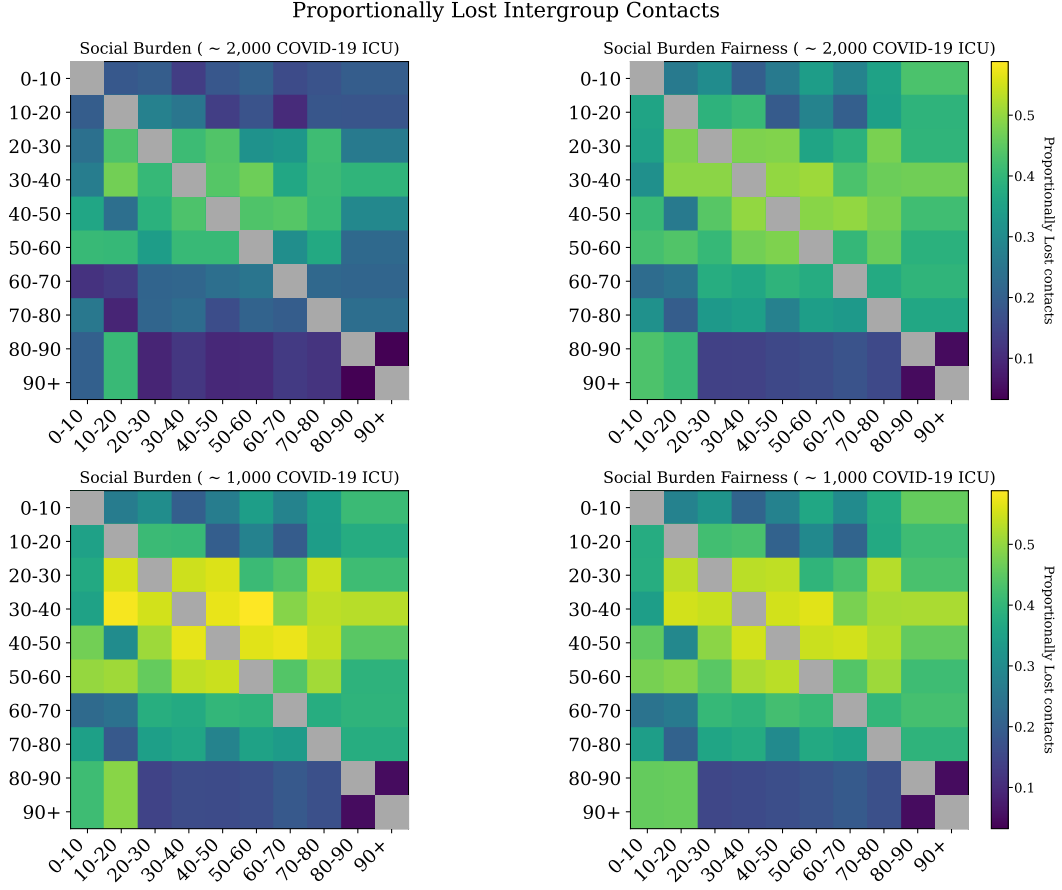


Figure 20: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings.

Figure 20 displays the effect of using *social burden fairness* more clearly. On the left, the looser policy causes the age-groups with lower hospitalization risks to be restricted more. On the right, PCN distributes the social restrictions more evenly among the groups. This suggests that the fairness notion successfully induces PCN to consider the hospitalization risks in determining the mitigation measures. In the more strict policies, we identify similar results. The different age-groups lose a comparable proportional amount of intergroup contacts in both settings. However, accounting for *social burden fairness* ensures a more even distribution of this reduction in social life.

9.2 Age-based Fairness Through Unawareness

In this section, we include a different fairness notion. This fairness notion incorporates a measure that tracks the difference in contact reductions and the difference in hospitalization risks between the age-groups. If age-groups have a similar risk of being hospitalized, they should experience similar social restrictions. First, we examine the Pareto fronts found by PCN when it optimizes on $\mathcal{R}_{ARH}$ and $\mathcal{F}_{ABFTU}$.

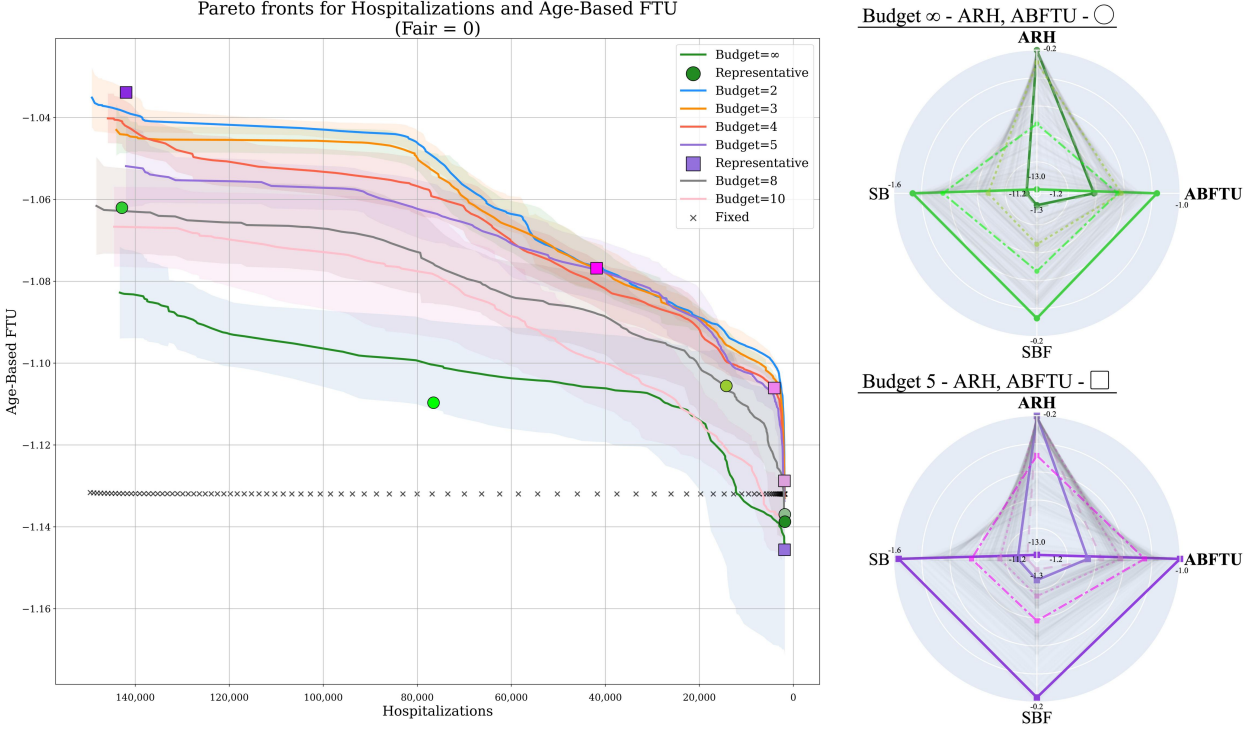


Figure 21: Interpolated Pareto fronts for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ for budgets $\infty, 2, 3, 4, 5, 8, 10$. Radar plots visualize the different trade-offs achieved by PCN for budgets ∞ and 5 while optimizing on given objectives. Each objective is shown scaled individually to a range $[-1, 0]$ along with the actual minimum and maximum values. The best performing policies for the optimized objectives are drawn solid and the supplementary policies dashed with less opacity. Each policy is marked in the corresponding color on the Pareto fronts. The x-axis visualizes the amount of new hospitalizations at the end of an epidemiological episode of 17 weeks, and the y-axis displays the unfairness measured by age-based fairness through unawareness. A result closer to zero indicates a more fair trajectory of social restrictions. The outcome for fixed policies is marked by crosses.

In Figure 21, we observe that the budgets mostly outperform the fixed policies. Interestingly, the lower budgets perform better considering the fairness notion. To highlight this characteristic, additional budgets 8 and 10 are displayed. Moreover, in the radar plots, these policies span a considerably larger area towards the different objectives compared to the policies that can adjust the social restrictions more frequently. Interestingly, we again observe how the policies performing better in terms of $\mathcal{F}_{ABFTU}$, perform increasingly on $\mathcal{F}_{SBF}$ as well. Furthermore, the fairness of the fixed policies is fairly constant. This is due to the age-groups all receiving the same social restrictions. This causes the similar groups to automatically have similar social restrictions, resulting in a fairly constant divergence between the social reduction distributions. Since the difference in hospitalization risks is constant, this causes the fairness notion to be nearly constant as well. Therefore, this baseline remains roughly the same value for different situations. We observe the worst and best policies according to hospitalization attack rate, and a trajectory that induces a peak of approximately 1,500 ICU patients in Figure 22.

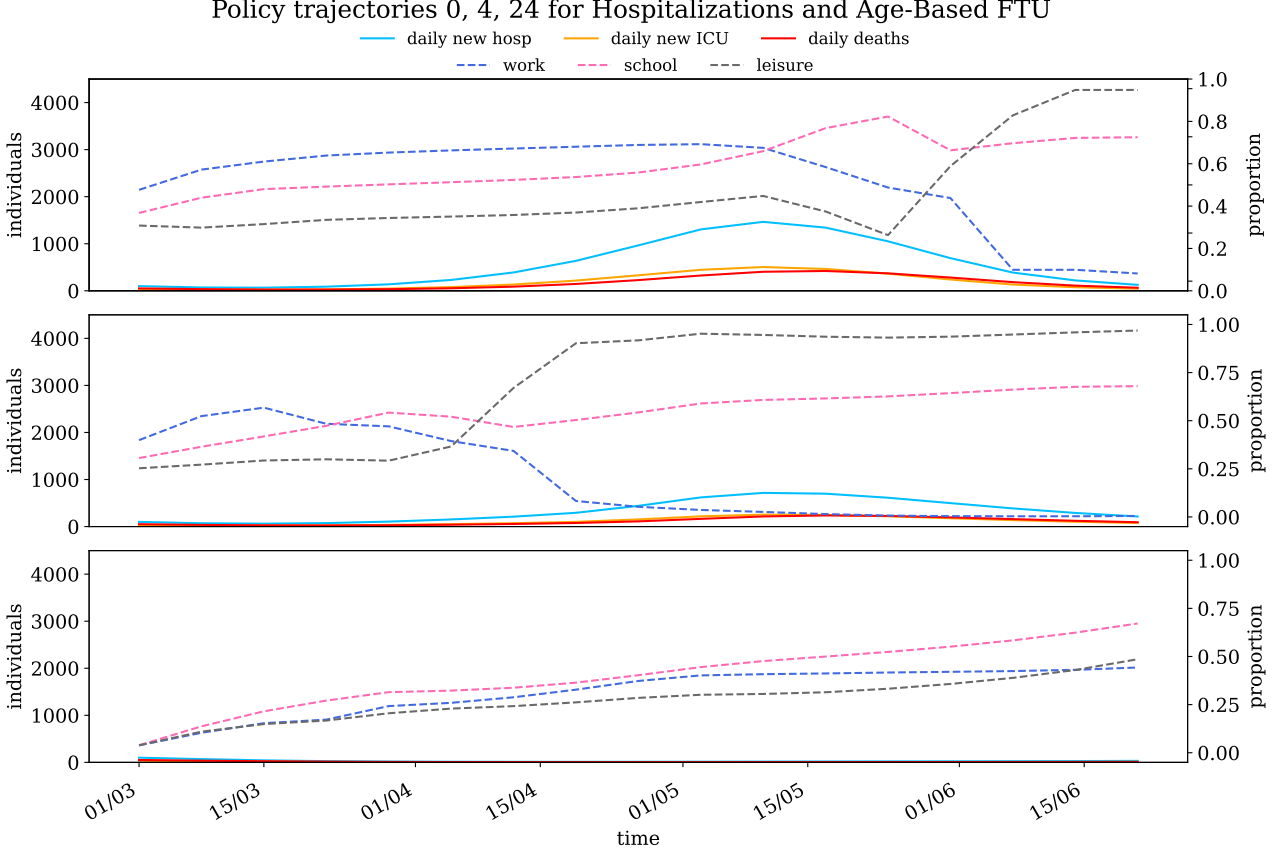


Figure 22: Epidemiological trajectories of policies achieving the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom). PCN is trained on $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ with a budget of 5 changes. The epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

Figure 22 illustrates that the worst performing policy for hospitalization attack rate still imposes moderately strict mitigation measures. Moreover, we note that the social contacts at work and school are restricted the least for the majority of the episode. This trajectory illustrates the shortcomings of only considering this particular fairness notion. Since no form of overall social burden is included in the objectives, PCN does not assess the amount of social contacts lost in the age-groups. Therefore, PCN tries to minimize the differences in social restrictions between the similar age-groups without considering how it affects the population overall. As a result, the least strict policy regarding hospitalizations still imposes heavy mitigation measures on the population. In the best performing policy considering hospitalizations, the model imposes severe restrictions at the beginning to gradually loosen these throughout the trajectory. In the center panel, the restrictions start off similar, but start to diverge after one month. At this point, the schools, and leisure are loosened while work contacts continue to be restricted heavily. This is due to $\mathcal{F}_{ABFTU}$ benefiting from restricting a large and homogeneous group like the working population rather than the heterogeneous or small groups in leisure and school. This ensures that the differences in hospitalization risks and social reductions are more in line with each other. Subsequently, the reductions cause the hospitalizations curve to be flattened.

Proportionally Lost Intergroup Contacts

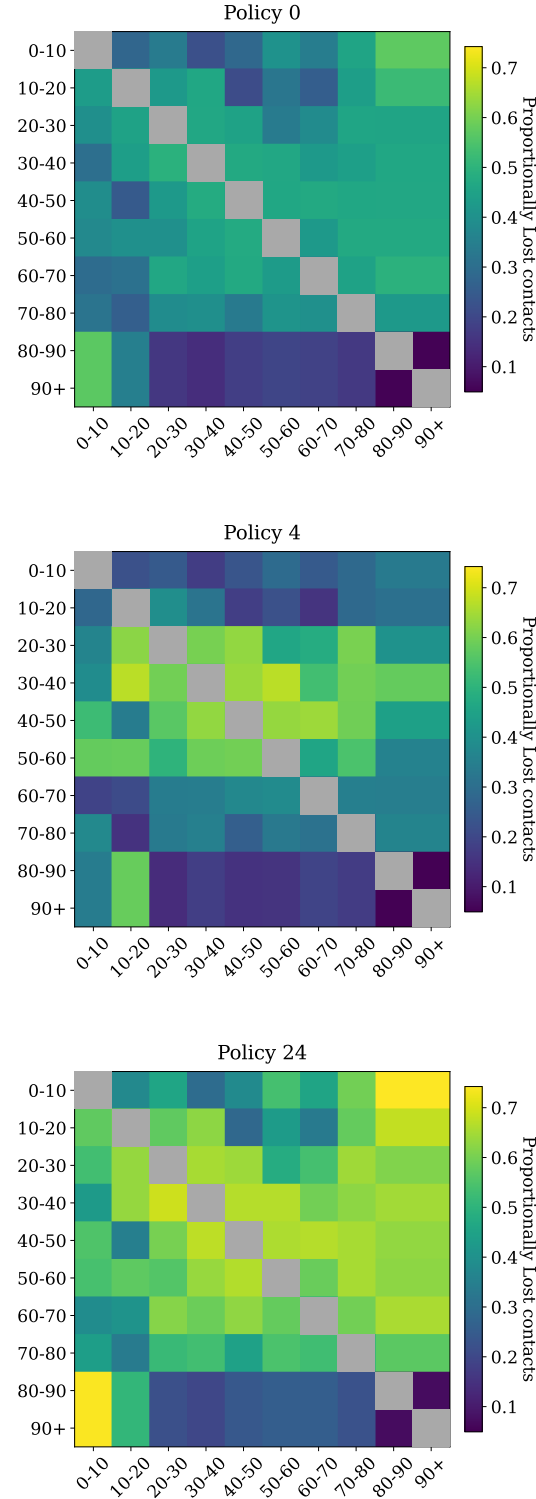


Figure 23: The average proportionally lost intergroup contacts for the displayed policies over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ with a budget of 5 changes. The policies achieve the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom).

In Figure 23, we observe additional insightful results. In the policy that causes the highest hospitalization attack rate, the reduced contacts are distributed evenly among the age-groups. The social contact between the oldest and youngest age-groups is affected the most. In the middle heatmap, the average lost intergroup contacts are concentrated in the groups that have a lower hospitalization risk. These groups are targeted when work life is restricted (Figure

5,22). Restricting this area is convenient as it targets a large group that is similar in their hospitalization risk. Therefore, this induces a positive effect on the fairness. When the policy gets increasingly strict, we observe that the restrictions are more severe and spread out over the different age-groups. This causes the unfairness to increase, as more different groups experience similar losses in social contacts. However, this induces a positive effect on the disease outbreak, as the hospitalizations will be limited. This effect can be observed in Figure 21 as well. The policies that cause little new hospitalizations, perform worse in terms of fairness.

9.2.1 $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$

Next, we compare versions of PCN trained solely to optimize on $\mathcal{R}_{ARH}$ and one fairness notion. First, we compare the common worst and best policies in terms of peak ICU patients for both of these versions.

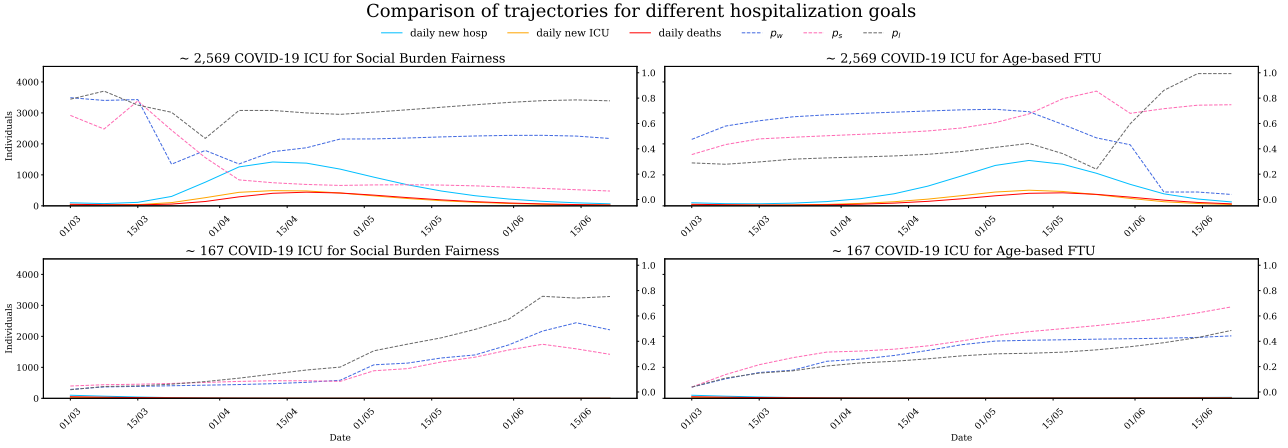


Figure 24: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

Figure 24 highlights multiple distinct patterns. Especially in the top trajectories, the networks approach the epidemic differently. On the left, PCN starts by restricting work, school, and leisure increasingly for one month. Afterward, it gradually allows more work and leisure contacts, while keeping social life at school to a minimum. On the right, we note a different pattern. PCN imposes moderate initial restrictions on all three domains. This delays the peak of hospitalizations compared to the trajectory on the left. When the hospitalization attack rate starts decreasing, PCN starts to loosen school and leisure restrictions, while increasingly restricting work life. The bottom panels show more similar results between the two settings. In both trajectories, PCN imposes heavy mitigation measures at the beginning, which are slowly lifted towards the end of the episode. On the left, there is a slight preference for loosening leisure restrictions, while the trajectory on the right leans towards allowing more school contacts. To compare the differences on a more pragmatic level, we compare the heat maps of these different policies in Figure 25.

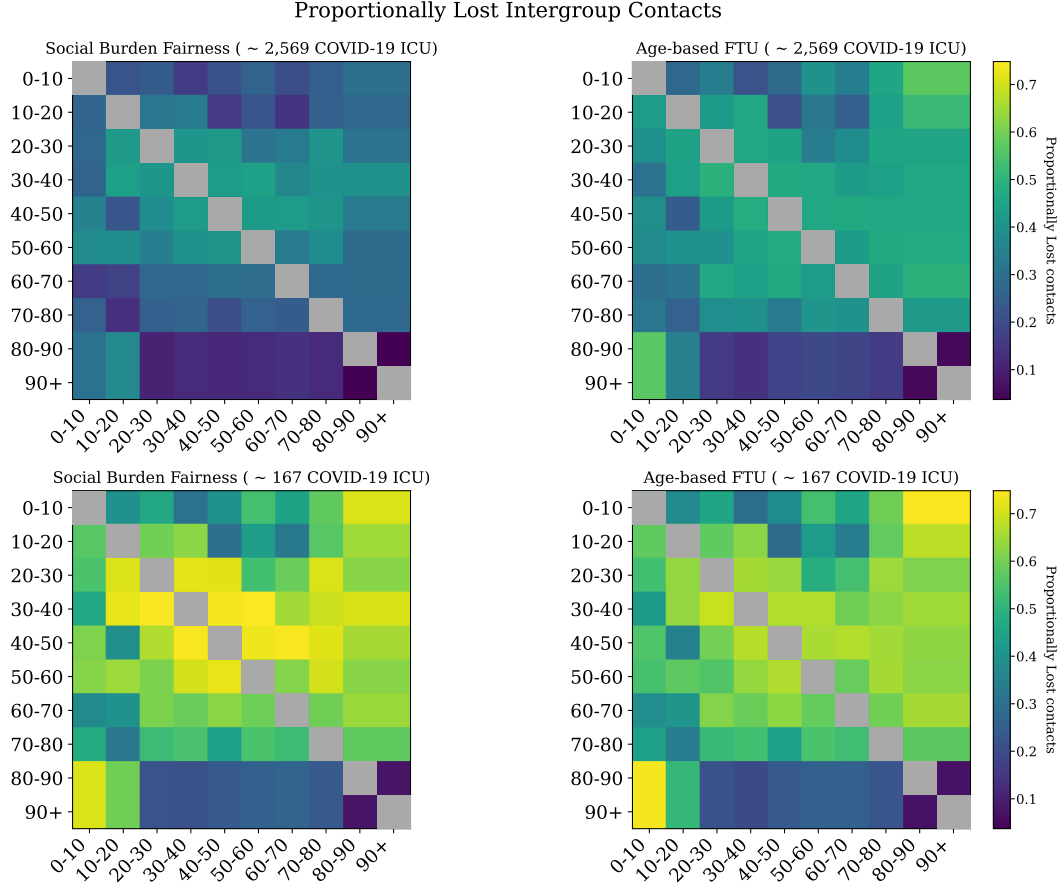


Figure 25: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings.

The heatmaps in Figure 25 highlight the effects of *age-based fairness through unawareness*. In the common worst policy for hospitalizations, the left heatmap shows a limited reduction in social contacts. However, these restrictions are slightly concentrated on the age-groups with lower hospitalization-risks. On the right, PCN restricts more, but the reduction in social contacts is more comparable for a larger portion of the population. In the bottom heatmaps, we identify a difference as well. To achieve the low hospitalization attack rate, the policy on the left still focuses the social restrictions mostly on the less vulnerable groups. On the right, we observe a more even distribution of these social restrictions. Next, we investigate two policies for both settings that limit the peak of ICU patients to 2,000 and 1,000.

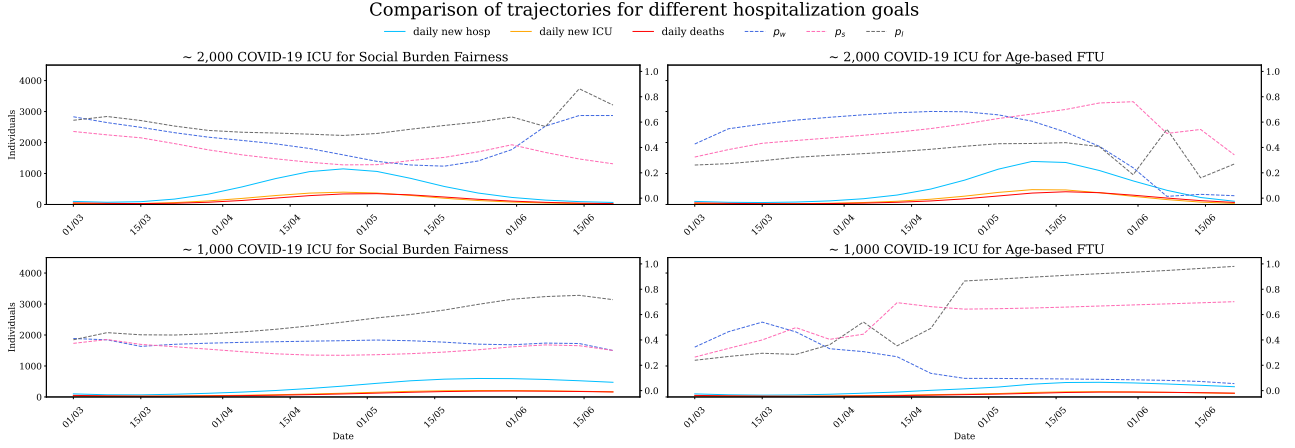


Figure 26: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In Figure 26, we note multiple differences. In the right top panel, PCN follows a more atypical approach. When the daily hospitalizations start to decrease, the model overall increases all restrictions. On the left, PCN decreases the allowed social contacts gradually until the peak of hospitalizations has passed. Afterward, it loosens the restrictions with a preference for leisure and work contacts. In the more strict trajectories, we identify differences as well. On the left, PCN restricts social behavior at work and school somewhat consistently, with a gradual loosening of the leisure restrictions. On the right, there is a more drastic shift. The mitigation measures start off similarly, but diverge at around one month into the epidemiological episode. At this point, leisure and schools are loosened considerably, while social contacts at work are restricted more. In both settings, we observe a delay of the peak of hospitalizations until the end of the trajectory. Next, we discuss the average proportionally lost intergroup contacts over the episodes.

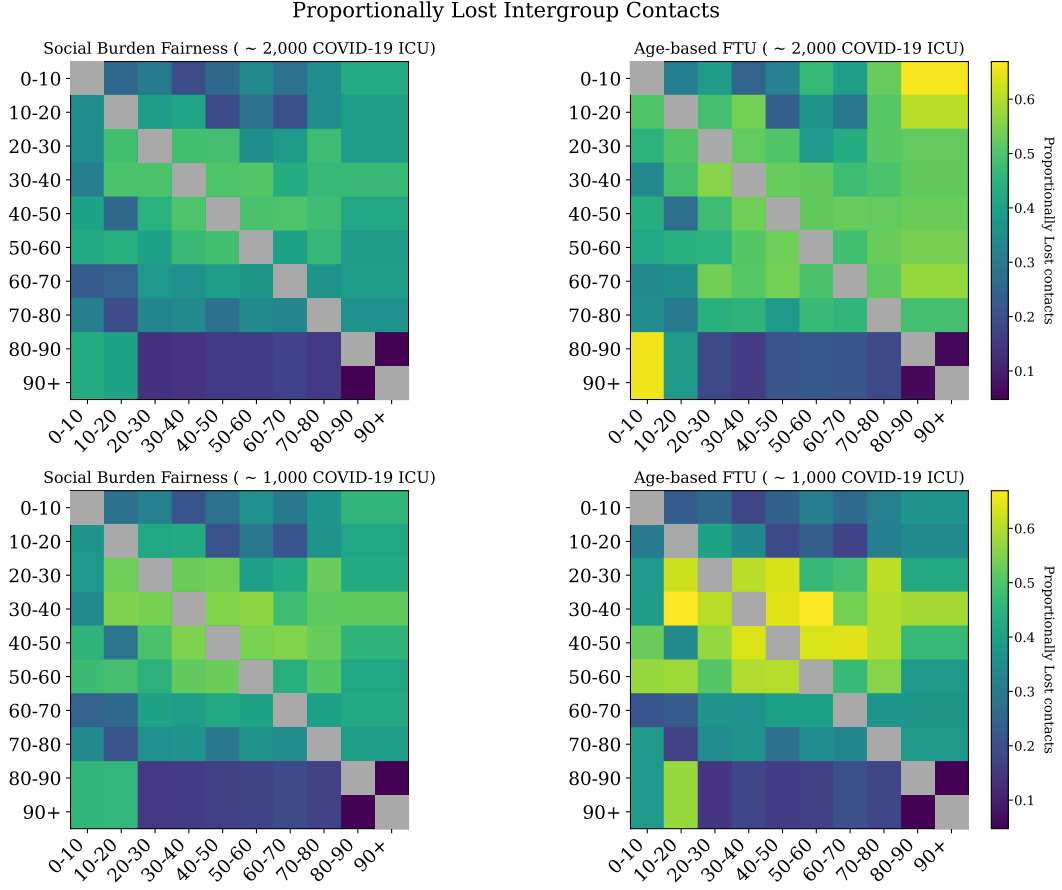


Figure 27: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings.

In Figure 27, the heatmaps show distinct patterns. In the top trajectories, PCN restricts the population more when accounting for $\mathcal{F}_{ABFTU}$, but restricts more evenly. On the left, the restrictions have a less severe impact, but are slightly concentrated on the less vulnerable groups. In the more strict policies, we note a diverging result compared to Figure 25. These trajectories limit the peak ICU patients more. On the left, PCN restricts more evenly compared to the right. In the right bottom policy, we note a more pronounced reduction in social contacts for the intergroup contacts of the working or less vulnerable population. This aligns with the observations of the trajectory in Figure 26.

9.3 Social Burden Fairness & Age-based FTU

In the last setting, we include both fairness notions along the hospitalization attack rate as objectives. First, we observe the Pareto fronts and the accompanying radar plots.

Pareto fronts for Hospitalizations, Social Burden Fairness and Age-Based FTU
(Fair = 0)

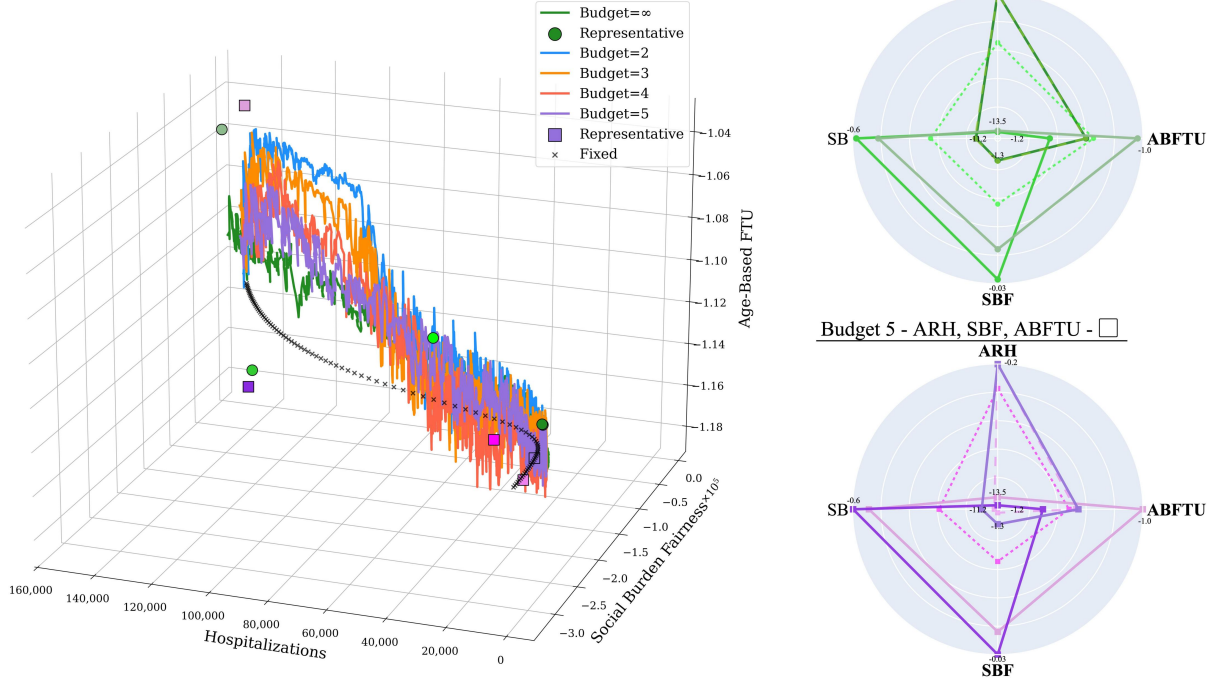


Figure 28: Interpolated Pareto fronts for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ for budgets $\infty, 2, 3, 4, 5$. Radar plots visualize the different trade-offs achieved by PCN for budgets ∞ and 5 while optimizing on given objectives. Each objective is shown scaled individually to a range $[-1, 0]$ along with the actual minimum and maximum values. The best performing policies for the optimized objectives are drawn solid and the supplementary policies dashed with less opacity. Each policy is marked in the corresponding color on the Pareto fronts. The x-axis visualizes the amount of new hospitalizations at the end of an epidemiological episode of 17 weeks, and the y-axis shows the scaled down unfairness, measured by social burden fairness. The z-axis shows the unfairness, measured by age-based fairness through unawareness. If the fairness notion measures a result closer to zero, this indicates a more fair trajectory for the population. The outcome for fixed policies is marked by crosses. Displaying the standard deviation was omitted for visibility purposes.

Figure 28 shows a fluctuating result for *age-based fairness through unawareness*. The Pareto fronts illustrate that most policies can outperform the fixed policies for both fairness notions. Only in the more strict policies, we observe various policies performing worse. When a low hospitalization attack rate is required, optimizing both fairness notions simultaneously becomes increasingly difficult. However, as stated before, the fixed policies have a nearly constant value for this fairness notion. In the corresponding radar plots, we observe the various trade-offs made. These display an interesting pattern. The policies that perform well on $\mathcal{F}_{ABFTU}$ also perform well on $\mathcal{F}_{SBF}$, but not vice versa. Additionally, for a higher budget, the policy that performs best in terms of hospitalizations, displays a better overall span across the other objectives. However, when focusing on fairness, the higher budget does not necessarily perform better. Next, we observe three trajectories for this setting.

Policy trajectories 0, 180, 469 for Hospitalizations, Social Burden Fairness and Age-Based FTU

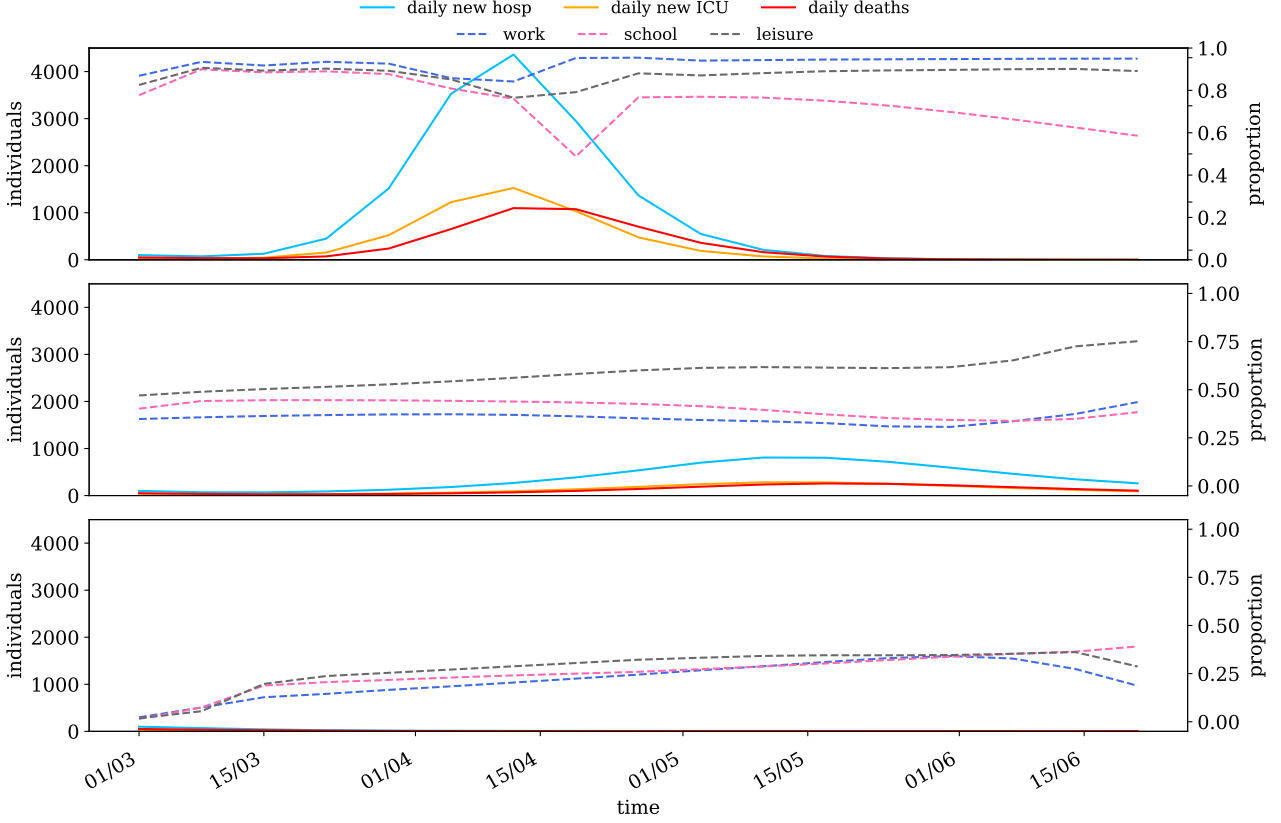


Figure 29: Epidemiological trajectories of policies achieving the worst hospitalization attack rate (top), 1,500 peak ICU patients (middle) and the best hospitalization attack rate (bottom). PCN is trained on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ with a budget of 5 changes. The epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In Figure 29, the top trajectory displays minimal social restrictions on leisure and work. The restrictions on school contacts are increased when the hospitalizations increase, but loosened as they decrease again. Towards the end of this episode, school contacts remain limited due to the limited budget. In the center panel, we note moderate and consistent restrictions. The social restrictions of work and school remain roughly the same level throughout the whole trajectory. Additionally, we note a gradual increase in the allowed leisure contacts. In the bottom and most strict trajectory, we identify considerably consistent and comparable restrictions on all domains. All restrictions are gradually loosened towards the end of the trajectory. At the very end, the social restrictions on work and leisure are increased. Overall, we note that there is a limited fluctuation of restriction levels in the trajectories.

9.3.1 $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ Vs. $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$

In this section, we examine how the incorporation of *age-based fairness through unawareness* impacts the trajectories generated by PCN. To this end, we compare policies in settings that solely include $\mathcal{F}_{SBF}$ as a metric of social strain to policies that consider both $\mathcal{F}_{SBF}$ and $\mathcal{F}_{ABFTU}$. We first explore the trajectories that cause the highest and lowest common peaks in ICU patients.

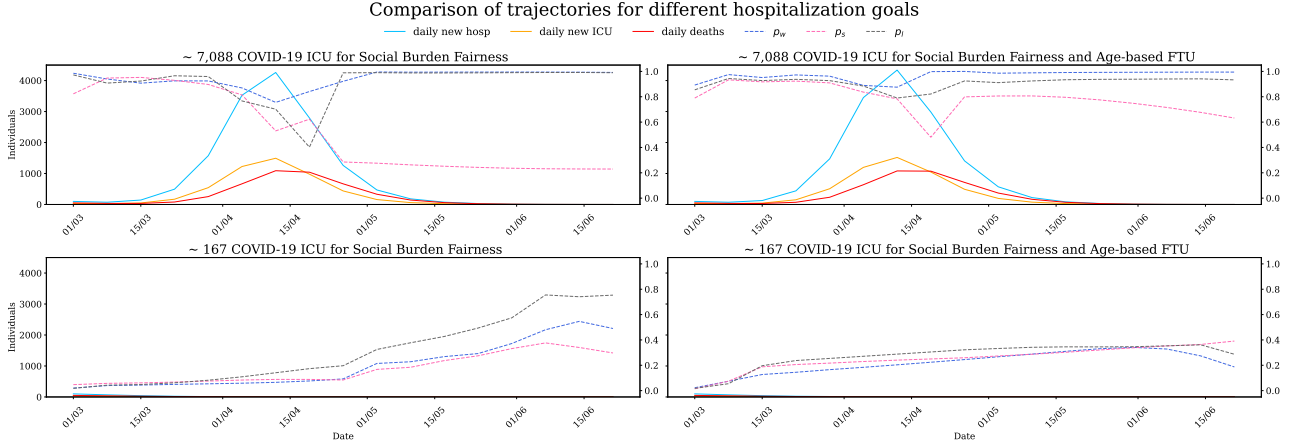


Figure 30: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

Figure 30 displays how including both fairness notions affects the trajectories. In the top panel, we identify severe restrictions on school contacts at the end of the epidemiological trajectory on the left. On the right, we note that the school contacts are restricted less. PCN increases the social restrictions less quickly and to a lesser extent in the right trajectory, while achieving a similar peak in ICU patients. In the bottom trajectories we identify a different comparison. On the left, the restrictions start off more severe, but are loosened faster towards the end of the trajectory. In the right trajectory, PCN loosens the restrictions quite slowly and keeps the levels of social restriction similar. Next, we observe how the social restrictions impacted the age-groups on average over the trajectories.

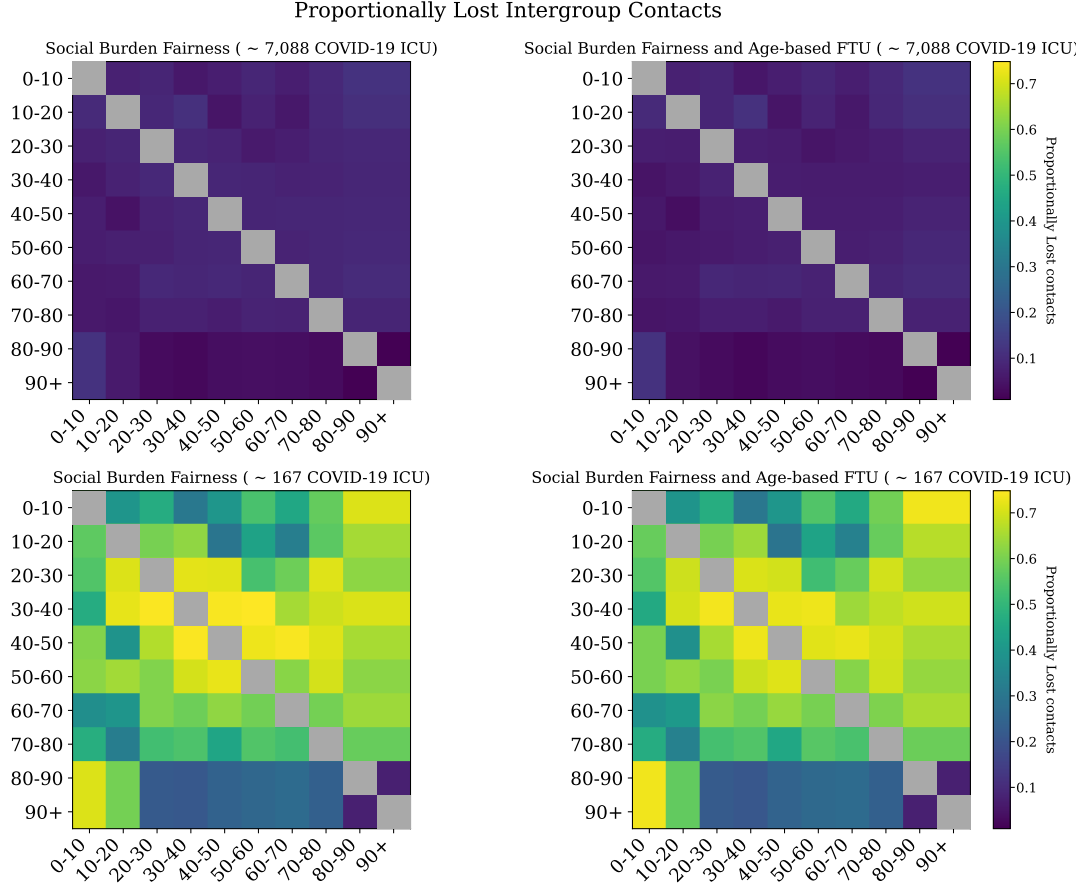


Figure 31: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause the maximum common peak of current ICU patients and the minimum common peak in their respective settings.

Figure 31 displays how the top trajectories restrict the different age-groups minimally. On the left, PCN causes slightly more contacts to be lost due to the remaining restrictions of school life. In the bottom policies, we again identify a subtle difference. In the right heatmap, the restrictions are slightly less targeted at the working population and more towards the contacts between elderly people and children. Next, we observe two trajectories that achieve a realistic peak of ICU patients over the trajectory.

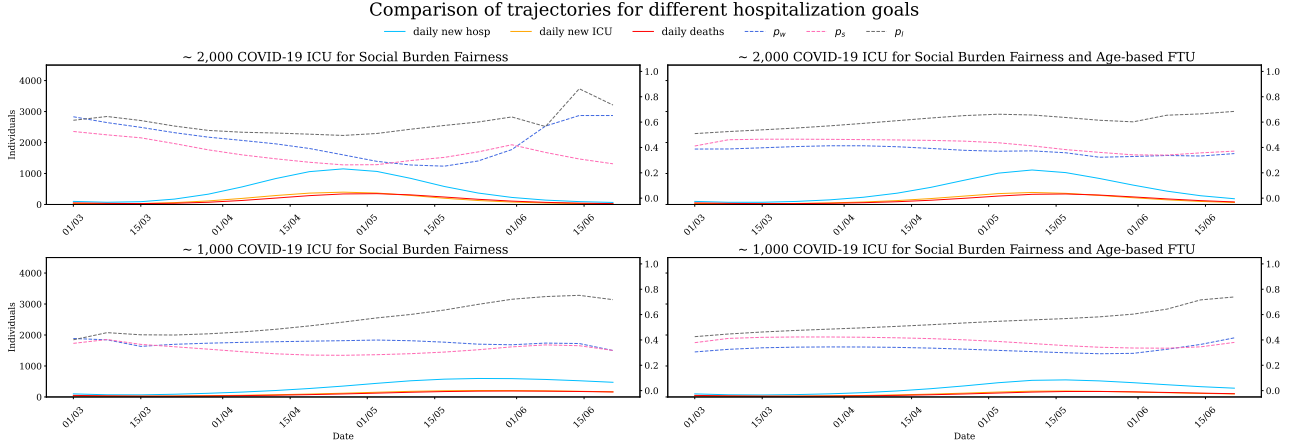


Figure 32: Comparing two epidemiological trajectories for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. Top and bottom trajectories cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings. In each panel, the epidemiological state is illustrated by daily new hospitalizations, ICU patients, and deaths. The dashed lines depict the proportional amount of contacts allowed in the corresponding domain.

In Figure 32, we identify a different result. In the top panel, the left trajectory shows more fluctuating social restrictions. Social contacts are increasingly restricted over the first two months of the trajectory. In the last month, PCN loosens the social restrictions on leisure and work, while maintaining the social restrictions at school. In the right trajectory, PCN imposes similar restrictions in the beginning. These mitigation measures remain moderately constant for work and schools, while leisure is slowly loosened towards the end. In the bottom panels, we observe two similar approaches. In both settings, PCN chooses to restrict schools and work similarly, while increasingly allowing more leisure contacts throughout the trajectory. Additionally, we observe different hospitalization curves in these settings. On the left, the curve is more spread out as a result of less severe restrictions on the working population. This causes a more prolonged strain on the hospitals. We investigate the heatmaps of the trajectories to illustrate how the age-groups are impacted overall.

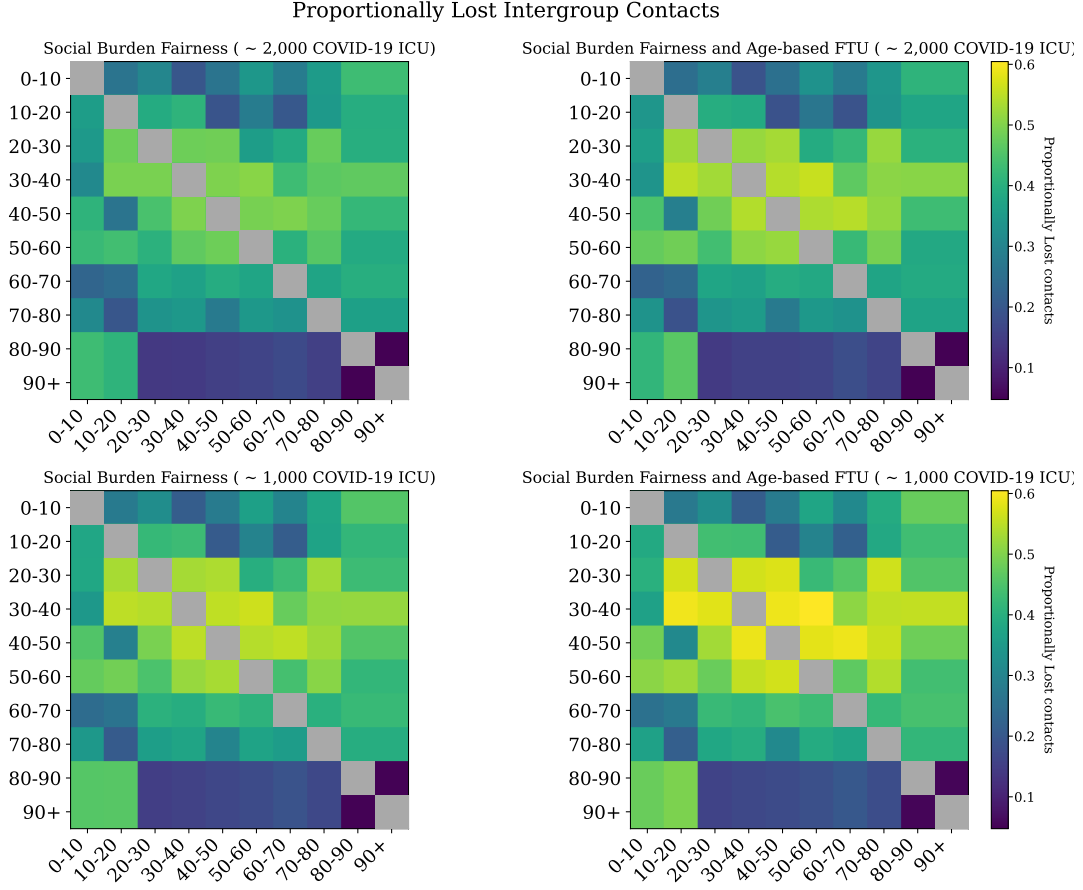


Figure 33: Comparing two heatmaps of the average proportionally lost intergroup contacts over the trajectory. PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ (left) and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}, \mathcal{F}_{ABFTU}]$ (right) with a budget of 5 changes. The top and bottom heatmaps cause a peak of 2,000 current ICU patients and a peak of 1,000 current ICU patients in their respective settings.

Figure 33 denotes how including both fairness notions impacts the age-groups socially. In the top heatmaps, we note that incorporating both fairness notions, impacts the working population slightly more. This aligns with the trajectories in Figure 32, where the social restrictions at work are maintained until the end of the trajectory. On the bottom, we identify a similar pattern. In the left heatmap, we observe a slightly less severe social impact on the less vulnerable working population compared to the right heatmap. We identify a trade-off between satisfying both fairness notions, as the heatmaps on the right show similarities to the corresponding heatmaps of both settings in Figure 27.

9.4 Overall Impact of Social Burden Fairness

To elucidate further on the impact of *social burden fairness*, we examine the overall impact of the metric. To this end, we visualize three heatmaps. The left heatmap denotes the average lost intergroup contacts for the age-groups across all policies when PCN is optimized on hospitalization attack rate and social burden. The middle heatmap depicts the average lost intergroup contacts when PCN is optimized on hospitalization attack rate and *social burden fairness*. On the right, the difference between these two heatmaps is shown.

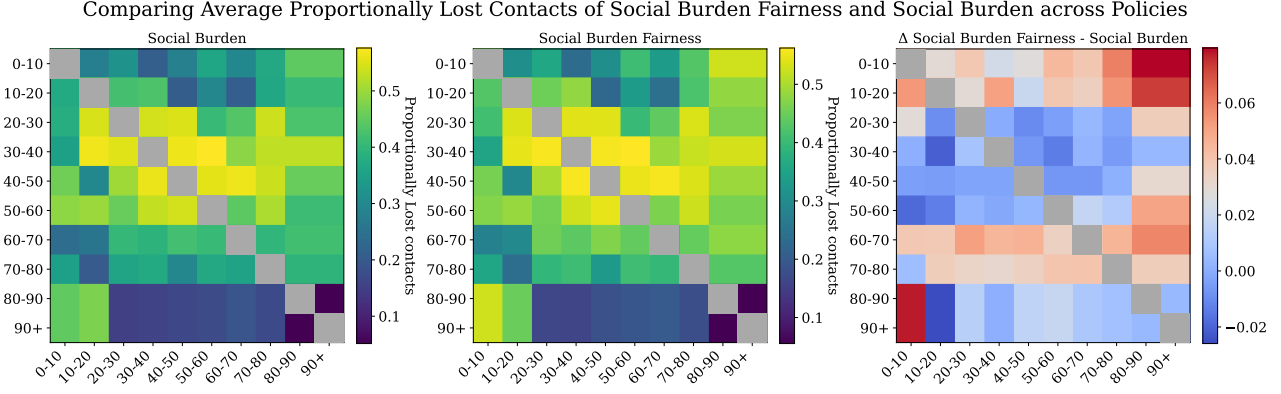


Figure 34: Comparing average proportionally lost contacts across all policies for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ with a budget of 5 changes. The rightmost heatmap denotes the difference between the heatmaps. Blue indicates a lower overall restriction by *social burden fairness*, and red a higher overall restriction.

Figure 34 illustrates the overall difference between the learned policies by both versions of PCN. In the right heatmap, we note an overall lower restriction for multiple age-groups with lower hospitalization risks. Additionally, we identify an increase for some groups with higher hospitalization risks. Despite the learned policies not aligning perfectly, the heatmaps provide some insight. The visualization highlights an overall difference in social restrictions of certain age-groups across all the learned policies. Next, we investigate how the average proportionally lost contacts are distributed among all policies for each age-group in the different settings.

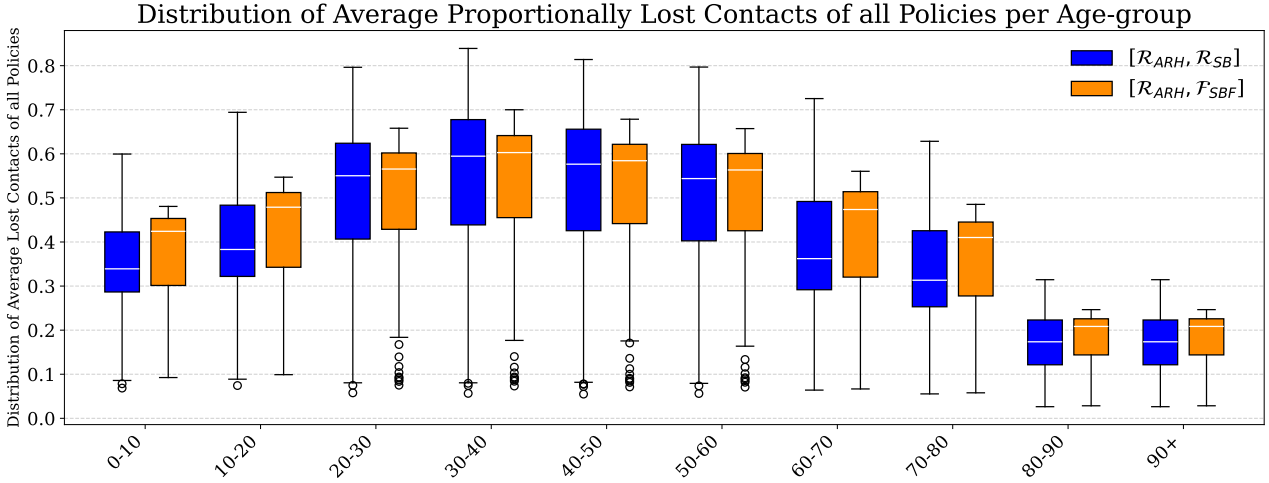


Figure 35: Comparing the distribution of the average proportionally lost contacts of learned policies for PCN optimized on $[\mathcal{R}_{ARH}, \mathcal{R}_{SB}]$ and $[\mathcal{R}_{ARH}, \mathcal{F}_{SBF}]$ with a budget of 5 changes for the different age-groups.

Figure 35 visualizes the spread of social reductions per age-group for the learned policies per setting. The box plots illustrate how *social burden fairness* impacts the social restrictions across the different age-groups. We identify that the spread for PCN optimized on social burden is consistently larger across all age-groups. When optimizing the fairness notion, PCN restricts the age-groups in a more uniform way and on average similarly to the different setting. However, the box plots highlight that *social burden fairness* encourages PCN to keep the social restrictions limited. Particularly in the less vulnerable age-groups, we note that PCN mostly restricts them less overall. Consequently, most age-groups with higher hospitalization risks display a spread that indicates overall higher restrictions when optimizing for fairness. As the epidemic control requires the population to be overall restricted to a certain extent, the restrictions are allocated slightly towards these age-groups.

10 Discussion

In this work, we explored how epidemic control is an issue that covers multiple, potentially conflicting objectives [Kao et al., 2024, Hamatani et al., 2022, Mazrekaj and Witte, 2024, Ceylan et al., 2020]. Since lockdowns are a useful tool, but can have undesirable implications, more sophisticated ways of disrupting epidemiological progression should be explored [Meo et al., 2020, Onyeaka et al., 2021, Auray and Eyquem, 2020]. We observed how MORL techniques such as Pareto Conditioned Networks can be used to choose mitigation measures for optimizing epidemic control [Reymond et al., 2024]. We extended the existing approach by Reymond et al., including aspects of fair treatment. To capture fair treatment in a robust way, we proposed two fairness notions applicable in RL settings. These fairness notions would require PCN to consider the fair treatment of the different age-groups based on their hospitalization risks.

For each variation of requested objectives, we identified the different trade-offs made by PCN by visualizing the Pareto fronts for different budgets. Accompanying this, we illustrated a selection of policies on radar plots to observe how the focus of PCN shifts on the Pareto front. These figures allowed us to increase our understanding of the effects of the different budgets, while also examining potentially conflicting objectives. Here, we observed how measures of social strain on the population could be compatible with each other or rather antagonistic.

To determine the measures of social distress, we proposed two novel fairness notions. The fairness notion *social burden fairness* was constructed by adapting the predefined social burden by Reymond et al.. It maintains the focus of minimizing the overall social restrictions of the population, while considering the different hospitalization risks. We investigated the overall impact of using *social burden fairness* compared to utilizing social burden in Section 9.4. As the used methods do not provide sufficient grounds for formal statistical examinations, explicit conclusions cannot be drawn. Nevertheless, these visualizations offer valuable insights and enable exploratory analysis to some extent.

To capture fairness on a more fine-grained level, we defined *age-based fairness through unawareness*. We proposed a simple way of expressing the outcome distributions of the social restrictions on the different age-groups. To quantify these distributions, we considered how the different age-groups are restricted for the different areas of reduction. However, this approach could potentially be improved as it provides a limited expressiveness of social stress. Inspecting the pair-wise differences in social restrictions of the age-groups, would highlight better whether the fairness notion induces PCN to choose actions according to differences in hospitalization risks. Additionally, the fairness notion displays a counterintuitive pattern. The higher budgets seem to worsen the ability of PCN to find policies that optimize fairness while considering the hospitalization attack rate. This could be due to the sensitivity of this fairness notion to performing more actions. Changing the policy more frequently, to decrease the hospitalization attack rate, causes the fairness notion to diverge more quickly to unfair conditions due to its small range.

We examined how incorporating the fairness notions in the objectives affects the chosen social restrictions. We visualized multiple trajectories to illustrate the different approaches PCN employed to achieve various goals. These goals were determined by reasoning about the driving factors of interventions during the COVID-19 pandemic [Faes et al., 2021]. Moreover, we considered realistic expectations that would induce manageable disease outbreaks regarding available ICU beds in Belgium [Belgian Health Care Knowledge Centre (KCE), 2024]. This enabled us to align the different policies on these goals across the different versions of PCN.

We displayed how the network effectively restricts the different age-groups over the weeks in epidemiological trajectories. Here, depending on the magnitude of the hospitalization goals, we identified different patterns of social restrictions.

Due to PCN’s complex architecture, explaining why the model enforces certain mitigation measures is difficult. To this end, we attempted to investigate the different imposed restrictions chosen by PCN by utilizing varying visualizations. However, to fully comprehend the system’s decision-making, more fine-grained examinations would be needed. We illustrated the advantages and strengths of the network, while also indicating the limitations at hand. The effectiveness of the network illustrated how PCN is a useful tool to support decision-making in epidemic control. It can provide useful insights for policy-makers, facilitating epidemic interventions. Moreover, the dimensions of epidemic control reach further than hospitalizations and effects on social lives [Deb et al., 2022, Kao et al., 2024, Mazrekaj and Witte, 2024, Ceylan et al., 2020]. Restricting work environments consequently has economic implications [Deb et al., 2022]. Unfortunately, increasing the objectives in MORL increases the solution space, resulting in an exponentially increasing number of policies. Currently, this impedes the ability of policy-makers to reason about the different trade-offs of distinct policies. Therefore, it is crucial that new strategies and techniques are continuously developed to address this issue.

The various radar plots, along with including both fairness notions, illustrated how both fairness notions might be non-conflicting in nature. We noted that policies performing well on one fairness notion, seemed to perform well on the other. However, including both fairness notions displayed a more unusual pattern. In the radar plots, we identified that the policies optimizing *social burden fairness*, did not necessarily imply a good performance on *age-based fairness through unawareness*. This highlights the need for a more extensive study on the proposed metric, as the understanding of its behavior could be improved.

11 Future Work

In this work, we focused on optimizing the momentary fairness. Each week, the fairness is assessed over the imposed restrictions. An important objective for future research would be to examine this evaluation over a longer time horizon. Reviewing fairness over multiple weeks or months could lead to different results. This would potentially involve incorporating different behaviors of the population such as resistance or fatigue as a result of imposed restrictions over a long period of time. Despite this introducing more difficulties due to the complexity of human behavior, this direction could provide additional insight and improve the robustness of handling epidemics for different time durations.

Furthermore, diseases do not necessarily involve higher risk factors for the older population. The influenza pandemic of 1918 had a different age-risk distribution compared to COVID-19. Here, the mortality rate of the younger population was remarkably higher [Gagnon et al., 2013]. Exploring this historical, or a hypothetical setting, could test the effectiveness of using PCN incorporated with fairness notions. Therefore, this research would reinforce the insights obtained in this work and broaden it across varying epidemiological settings.

The inclusion of economical objectives in RL settings has been explored [Kao et al., 2024, Ohi et al., 2020]. However, it has not been studied alongside incorporating fair treatment of the population. Future research can explore this direction by including economical effects of restrictions on work alongside the proposed fairness notions.

12 Use of Generative AI

Generative AI was utilized in a supporting role in multiple parts of the research process. Primarily, ChatGPT was applied to generate source code for data visualizations. All generated code was carefully revised before execution. Additionally, LanguageTool was occasionally employed for spelling errors and language enhancement. Lastly, ChatGPT was used for targeted information retrieval in academic research, which was consistently followed by the manual examination of the results.

References

- Steven Abrams, James Wambua, Eva Santermans, Lander Willem, Elise Kuylen, Pietro Coletti, Pieter Libin, Christel Faes, Oana Petrof, Sereina A. Herzog, Philippe Beutels, and Niel Hens. Modelling the early phase of the belgian covid-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics*, 35, 6 2021. ISSN 18780067. doi: 10.1016/j.epidem.2021.100449.
- Stéphane Auray and Aurélien Eyquem. The macroeconomic effects of lockdown policies. *Journal of Public Economics*, 190:104260, 2020. ISSN 0047-2727. doi: <https://doi.org/10.1016/j.jpubeco.2020.104260>. URL <https://www.sciencedirect.com/science/article/pii/S0047272720301249>.
- Norman TJ Bailey. *The mathematical theory of infectious diseases and its applications*. /, 1975.
- Belgian Health Care Knowledge Centre (KCE). Performance of the Belgian health system – report 2024. Kce report, Belgian Health Care Knowledge Centre (KCE), 2024. URL <https://www.healthybelgium.be/metadata/hspa/2024/R2.pdf>.
- Samantha K Brooks, Rebecca K Webster, Louise E Smith, Lisa Woodland, Simon Wessely, Neil Greenberg, and Gideon James Rubin. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The Lancet*, 395(10227):912–920, 2025/05/30 2020. doi: 10.1016/S0140-6736(20)30460-8. URL [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8).
- Rahmiye Figen Ceylan, Burhan Ozkan, and Esra Mulazimogullari. Historical evidence for economic effects of covid-19, 2020.
- Xinwei Chen, Tong Wang, Barrett W. Thomas, and Marlin W. Ulmer. Same-day delivery with fair customer service. *European Journal of Operational Research*, 308(2):738–751, 2023. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2022.12.009>. URL <https://www.sciencedirect.com/science/article/pii/S0377221722009365>.
- Alexandra Cimpean, Catholijn Jonker, Pieter Libin, and Ann Nowé. A group and individual aware framework for fair reinforcement learning, 2025.
- Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M Eggo. Age-dependent effects in the transmission and control of covid-19 epidemics. *Nature medicine*, 26(8):1205–1211, 2020.
- Pragyan Deb, Davide Furceri, Jonathan D Ostry, and Nour Tawk. The economic effects of covid-19 containment measures. *Open Economies Review*, 33(1):1–32, 2022.
- Salvatore Di Martino and Isaac Prilleltensky. Happiness as fairness: The relationship between national life satisfaction and social justice in eu countries. *Journal of Community Psychology*, 48(6):1997–2012, 2020. doi: <https://doi.org/10.1002/jcop.22398>.
- Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *3rd Innovations in Theoretical Computer Science Conference*, ITCS ’12, page 214–226, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450311151. doi: 10.1145/2090236.2090255. URL <https://doi.org/10.1145/2090236.2090255>.

- Cynthia Dwork, Christina Ilvento, Guy N. Rothblum, and Pragma Sur. Abstracting Fairness: Oracles, Metrics, and Interpretability. In *1st Symposium on Foundations of Responsible Computing*, 2020.
- Christel Faes, Niel Hens, and Marius Gilbert. On the timing of interventions to preserve hospital capacity: lessons to be learned from the belgian sars-cov-2 pandemic in 2020. *Archives of Public Health*, 79(1):164, 2021. doi: 10.1186/s13690-021-00685-2. URL <https://doi.org/10.1186/s13690-021-00685-2>.
- Zimeng Fan, Nianli Peng, Muhang Tian, and Brandon Fain. Welfare and Fairness in Multi-objective Reinforcement Learning, October 2023. URL <http://arxiv.org/abs/2212.01382>. doi: <https://doi.org/10.48550/arXiv.2212.01382>.
- Alain Gagnon, Matthew S. Miller, Stacey A. Hallman, Robert Bourbeau, D. Ann Herring, David JD. Earn, and Joaquín Madrenas. Age-specific mortality during the 1918 influenza pandemic: Unravelling the mystery of high young adult mortality. *PLOS ONE*, 8(8):1–9, 08 2013. doi: 10.1371/journal.pone.0069586. URL <https://doi.org/10.1371/journal.pone.0069586>.
- Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 51–60, 2018.
- Sayo Hamatani, Daiki Hiraoka, Kai Makita, Akemi Tomoda, and Yoshifumi Mizuno. Longitudinal impact of covid-19 pandemic on mental health of children in the abcd study cohort. *Scientific Reports*, 12(1):19601, 2022. doi: 10.1038/s41598-022-22694-z. URL <https://doi.org/10.1038/s41598-022-22694-z>.
- Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36(1):26, 2022. doi: 10.1007/s10458-022-09552-y. URL <https://doi.org/10.1007/s10458-022-09552-y>.
- Thang Van Hoang, Pietro Coletti, Yimer Wasihun Kifle, Kim Van Kerckhove, Sarah Vercruysse, Lander Willem, Philippe Beutels, and Niel Hens. Close contact infection dynamics over time: insights from a second large-scale social contact survey in flanders, belgium, in 2010-2011. *BMC infectious diseases*, 21:1–15, 2021.
- Mamoru Kaneko and Kenjiro Nakamura. The nash social welfare function. *Econometrica: Journal of the Econometric Society*, pages 423–435, 1979. doi: <https://doi.org/10.2307/1914191>.
- Yamin Kao, Po-Jui Chu, Pai-Chien Chou, and Chien-Chang Chen. A dynamic approach to support outbreak management using reinforcement learning and semi-connected seiqr models. *BMC Public Health*, 24(1):751, 2024. doi: 10.1186/s12889-024-18251-0. URL <https://doi.org/10.1186/s12889-024-18251-0>.
- Yimer Wasihun Kifle, Nele Goeyvaerts, Kim Van Kerckhove, Lander Willem, Christel Faes, Herwig Leirs, Niel Hens, and Philippe Beutels. Animal ownership and touching enrich the context of social contacts relevant to the spread of human infectious diseases. *PLOS ONE*,

- 10(7):1–13, 07 2015. doi: 10.1371/journal.pone.0133461. URL <https://doi.org/10.1371/journal.pone.0133461>.
- S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951. doi: 10.1214/aoms/1177729694. URL <https://doi.org/10.1214/aoms/1177729694>.
- Pieter Libin, Arno Moonens, Timothy Verstraeten, Fabian Perez-Sanjines, Niel Hens, Philippe Lemey, and Ann Nowé. Deep reinforcement learning for large-scale epidemic control. ..., 3 2020. URL <http://arxiv.org/abs/2003.13676>.
- Pieter J.K. Libin, Lander Willem, Timothy Verstraeten, Andrea Torneri, Joris Vanderlocht, and Niel Hens. Assessing the feasibility and effectiveness of household-pooled universal testing to control covid-19 epidemics. *PLoS Computational Biology*, 17, 3 2021. ISSN 15537358. doi: 10.1371/journal.pcbi.1008688.
- Karima Makhoul, Sami Zhioua, and Catuscia Palamidessi. Machine learning fairness notions: Bridging the gap with real-world applications. ..., 6 2020. doi: 10.1016/j.ipm.2021.102642. URL <http://arxiv.org/abs/2006.16745><http://dx.doi.org/10.1016/j.ipm.2021.102642>.
- Harry Markowitz. Portfolio Selection. *Journal of Finance*, 7(1):77–91, March 1952. doi: j.1540-6261.1952.tb01525.x. URL <https://ideas.repec.org/a/bla/jfinan/v7y1952i1p77-91.html>.
- Deni Mazrekaj and Kristof De Witte. The impact of school closures on learning and mental health of children: Lessons from the covid-19 pandemic. *Perspectives on Psychological Science*, 19(4):686–693, 2024. doi: 10.1177/17456916231181108. URL <https://doi.org/10.1177/17456916231181108>. PMID: 37427676.
- Sultan Ayoub Meo, Abdulelah Adnan Abukhalaf, Ali Abdullah Alomar, Faris Jamal AlMutairi, Adnan Mahmood Usmani, and David C Klonoff. Impact of lockdown on covid-19 prevalence and mortality during 2020 pandemic: observational analysis of 27 countries. *European journal of medical research*, 25:1–7, 2020.
- Mafalda N. S. Miranda, Marta Pingarilho, Victor Pimentel, Andrea Torneri, Sofia G. Seabra, Pieter J. K. Libin, and Ana B. Abecasis. A tale of three recent pandemics: Influenza, hiv and sars-cov-2. *Frontiers in Microbiology*, Volume 13 - 2022, 2022. ISSN 1664-302X. doi: 10.3389/fmicb.2022.889643. URL <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.889643>.
- Włodzimierz Ogryczak, Hanan Luss, Michał Pióro, Dritan Nace, Artur Tomaszewski, et al. Fair optimization and networks: A survey. *Journal of Applied Mathematics*, 2014, 2014. doi: <https://doi.org/10.1155/2014/612018>.
- Abu Quwsar Ohi, M. F. Mridha, Muhammad Mostafa Monowar, and Md. Abdul Hamid. Exploring optimal control of epidemic spread using reinforcement learning. *Scientific Reports*, 10(1):22106, 2020. doi: 10.1038/s41598-020-79147-8. URL <https://doi.org/10.1038/s41598-020-79147-8>.
- Helen Onyeaka, Christian K Anumudu, Zainab T Al-Sharify, Esther Egele-Godswill, and Paul Mbaegbu. Covid-19 pandemic: A review of the global lockdown and its far-reaching effects. *Science Progress*, 104(2):00368504211019854, 2021. doi: 10.1177/00368504211019854. URL <https://doi.org/10.1177/00368504211019854>. PMID: 34061685.

- Aniruddh Raghu, Matthieu Komorowski, Imran Ahmed, Leo Celi, Peter Szolovits, and Marzyeh Ghassemi. Deep reinforcement learning for sepsis treatment, 2017. URL <https://arxiv.org/abs/1711.09602>.
- Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. Pareto conditioned networks, 2022. URL <https://arxiv.org/abs/2204.05036>.
- Mathieu Reymond, Conor F. Hayes, Lander Willem, Roxana Rădulescu, Steven Abrams, Diederik M. Roijers, Enda Howley, Patrick Mannion, Niel Hens, Ann Nowé, and Pieter Libin. Exploring the pareto front of multi-objective covid-19 mitigation policies using reinforcement learning. *Expert Systems with Applications*, 249:123686, 2024. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2024.123686>. URL <https://www.sciencedirect.com/science/article/pii/S0957417424005529>.
- Umer Siddique, Paul Weng, and Matthieu Zimmer. Learning Fair Policies in Multiobjective (Deep) Reinforcement Learning with Average and Discounted Rewards, August 2020. URL <http://arxiv.org/abs/2008.07773>. doi: <https://doi.org/110.48550/arXiv.2008.07773>.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. ISBN 0262039249.
- Kristof Van Moffaert and Ann Nowé. Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research*, 15(1):3483–3512, 2014.
- Sam Vanspringel, Ioana Alexandra Cimpéan, Pieter Libin, and Ann Nowe. A general purpose framework for fairness in job hiring. *Proceedings of the Benelux Conference on Artificial Intelligence*, <https://bnaic2023.tudelft.nl/pre-proceedings:1-3>, 2023. ISSN 1568-7805.
- Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164(10):936–944, 09 2006. ISSN 0002-9262. doi: [10.1093/aje/kwj317](https://doi.org/10.1093/aje/kwj317). URL <https://doi.org/10.1093/aje/kwj317>.
- John A. Weymark. Generalized gini inequality indices. *Mathematical Social Sciences*, 1(4): 409–430, 1981. ISSN 0165-4896. doi: [https://doi.org/10.1016/0165-4896\(81\)90018-4](https://doi.org/10.1016/0165-4896(81)90018-4). URL <https://www.sciencedirect.com/science/article/pii/0165489681900184>.
- Lander Willem, Kim Van Kerckhove, Dennis L. Chao, Niel Hens, and Philippe Beutels. A nice day for an infection? weather conditions and social contact patterns relevant to influenza transmission. *PLOS ONE*, 7(11):1–7, 11 2012. doi: [10.1371/journal.pone.0048695](https://doi.org/10.1371/journal.pone.0048695). URL <https://doi.org/10.1371/journal.pone.0048695>.
- Lander Willem, Thang Van Hoang, Sebastian Funk, Pietro Coletti, Philippe Beutels, and Niel Hens. Socrates: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for covid-19. *BMC research notes*, 13(1):293, 2020.
- Anke B Witteveen, Susanne Y Young, Pim Cuijpers, José Luis Ayuso-Mateos, Corrado Barbui, Federico Bertolini, Maria Cabello, Camilla Cadorin, Naomi Downes, Daniele Franzoi, et al. Covid-19 and common mental health symptoms in the early phase of the pandemic: An umbrella review of the evidence. *PLoS medicine*, 20(4):e1004206, 2023.
- Chao Yu, Jiming Liu, and Shamim Nemati. Reinforcement learning in healthcare: A survey, 2020. URL <https://arxiv.org/abs/1908.08796>.

A Appendices

A.1 Social Burden Fairness for different action patterns

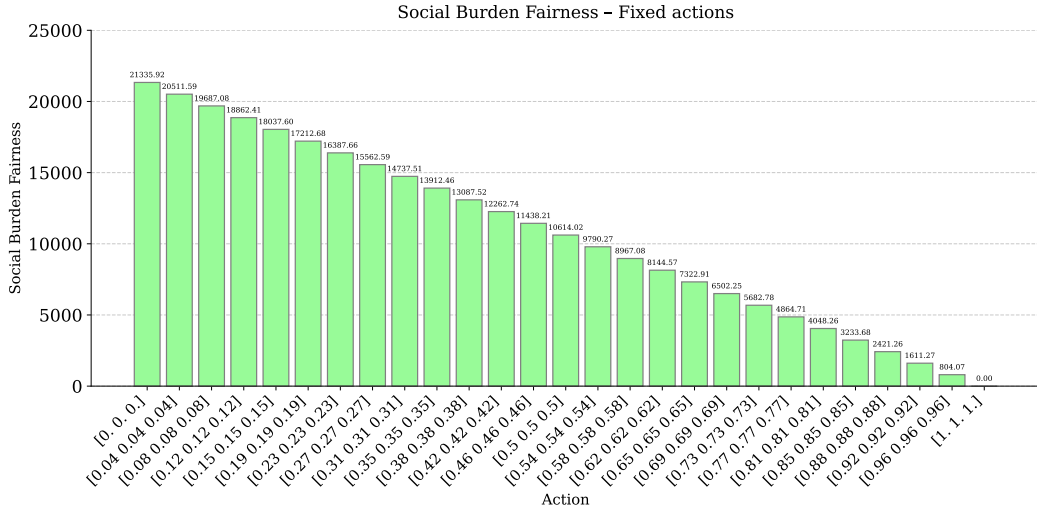


Figure 36: The absolute values of $\mathcal{F}_{SBF}$ for different fixed actions.

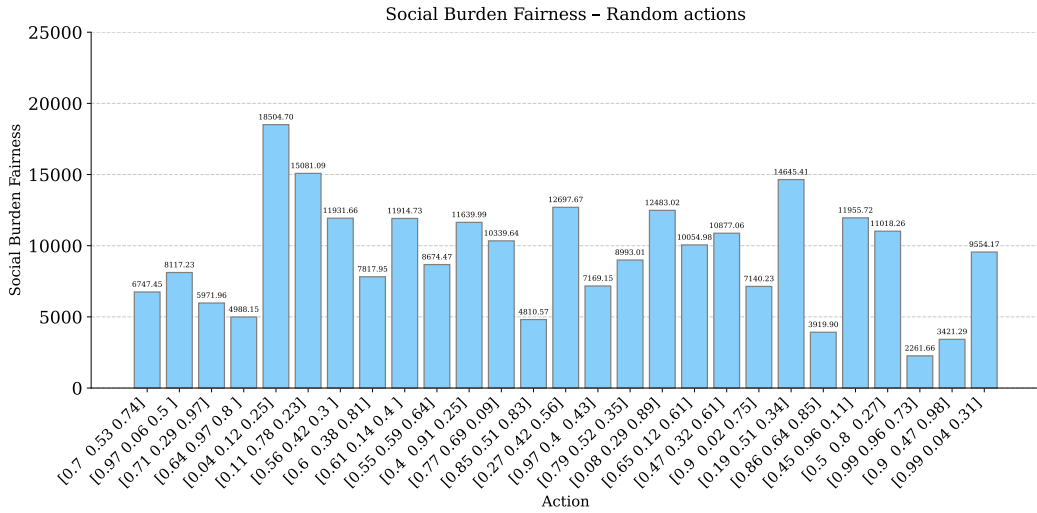


Figure 37: The absolute values of $\mathcal{F}_{SBF}$ for random actions.

A.2 Age-Based FTU for different action patterns

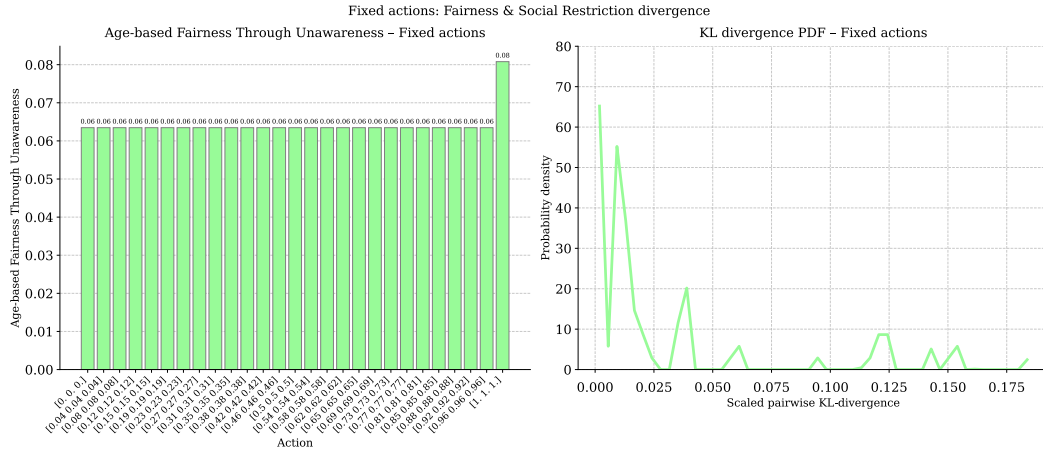


Figure 38: The absolute values of $\mathcal{F}_{ABFTU}$ for different fixed actions.

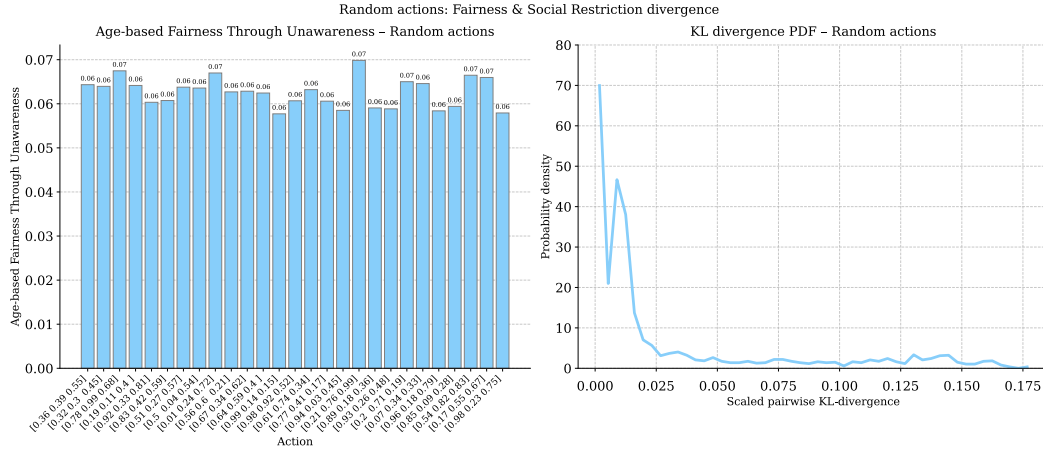


Figure 39: The absolute values of $\mathcal{F}_{ABFTU}$ for random actions.