A Causal Perspective on Meaningful and Robust Algorithmic Recourse

Gunnar König [1 2] **Timo Freiesleben** [3] **Moritz Grosse-Wentrup** [2]

Abstract

Algorithmic recourse explanations inform stakeholders on how to act to revert unfavorable predictions. However, in general ML models do not predict well in interventional distributions. Thus, an action that changes the prediction in the desired way may not lead to an improvement of the underlying target. Such recourse is neither meaningful nor robust to model refits. Extending the work of Karimi et al. (2021), we propose meaningful algorithmic recourse (MAR) that only recommends actions that improve both prediction and target. We justify this selection constraint by highlighting the differences between model audit and meaningful, actionable recourse explanations. Additionally, we introduce a relaxation of MAR called effective algorithmic recourse (EAR), which, under certain assumptions, yields meaningful recourse by only allowing interventions on causes of the target.

1. Introduction

Predictive systems are increasingly deployed in high-stakes environments such as hiring (Raghavan et al., 2020), recidivism prediction (Zeng et al., 2015) or loan approval (Van Liebergen et al., 2017). To enable individuals to revert unfavorable decisions, a range of work develops tools that offer individuals possibilities for algorithmic recourse (Wachter et al., 2018; Dandl et al., 2020; Karimi et al., 2020b; 2021). When suggesting actions for recourse, it is desirable that recommendations are *robust* to shifts, meaning that they can be honored if acted upon (Venkatasubramanian & Alfano, 2020). We argue that they should also be *meaningful*, such that not only the prediction, but also the underlying target is improved.

We take a causal perspective on the issue at hand and argue that robustness and meaningfulness are related problems since non-meaningful recourse itself leads to distribution shift. Let us consider a simple motivational example[1] illustrated in Figure 1. The goal is to predict the insurance risk of a car. In addition to the direct cause of whether the car is driven by the *car owner* (green), the confounded variable *minivan* is observed (blue). The latent confounder driver *defensiveness* cannot be observed. The ML model learns to exploit not only the direct cause (green) but also the associated variable (blue). Algorithmic recourse actions on the model may therefore suggest explainees to game the predictor by intervening on the non-causal variable minivan, thereby affecting the prediction without actually improving the insurance risk. In the distribution of agents that have gamed the prediction model, the association of *minivan* with the prediction target is broken. A refitted model would adapt accordingly and the recourse would not be honored.

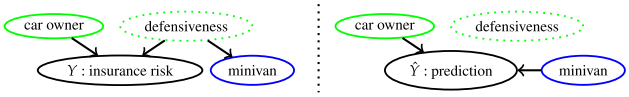

Figure 1. Left: Directed Acyclic Graph (DAG) illustrating the data generating process. Right: Directed Acyclic Graph (DAG) illustrating prediction model.

Miller et al. (2020) suggest to adapt the model such that gaming is not incentivized, which may come at the cost of predictive performance (Shavit et al., 2020). Instead we suggest to tackle the problem in the explanation domain. We propose Meaningful Algorithmic Recourse (MAR) (Section 4), which restricts recourse recommendations to meaningful actions that alter both the prediction $\hat{Y}$ and the target Y coherently. We justify the restriction by separating two goals: contestability and recourse (Section 5). While stakeholders that seek explanations for model audit (contestability) should be given full access to the model, agents that seek to revert an unfavorable outcome (recourse) should only be offered recourse options that are meaningful. A relaxation of the meaningfulness restriction is introduced in Section 6.

[1]Institute for Statistics, LMU Munich [2]Research Group Neuroinformatics, University of Vienna [3]Munich Center for Mathematical Philosophy, LMU Munich. Correspondence to: Gunnar König <g.koenig.edu@pm.me>.

ICML (International Conference on Machine Learning) Workshop on Algorithmic Recourse.

[1]The example is inspired by (Shavit et al., 2020).

2. Related Work and Contributions

Causal Perspective on Strategic Modeling: The related field of strategic modeling investigates how the prediction mechanism incentivizes rational agents. Miller et al. (2020) thereby distinguish models that incentivize *gaming*, i.e., interventions that affect the prediction $\hat{Y}$ but not the underlying target Y in the desired way, and *improvement*, actions that yield the desired change both in $\hat{Y}$ and Y. Building on this distinction, (Shavit et al., 2020) elaborate that except for cases where all causes can be measured, the following three goals are in conflict: incentivizing improvement, predictive accuracy and retrieving the true underlying mechanism.

Robust Algorithmic Recourse: Barocas et al. (2020); Venkatasubramanian & Alfano (2020) argue that counterfactual explanations (CE) assume the model to be stable over time, but that recourse should be guaranteed even if the model changes. In a similar vein, Wachter et al. (2018) suggest guaranteeing recourse based on a counterfactual within a pre-specified period of time. The robustness of CE and recourse has been investigated before (Rawal et al., 2020; Upadhyay et al., 2021; Pawelczyk et al., 2020), yet only with respect to generic shifts. To the best of our knowledge, we are the first to suggest recommendations that are robust to the shift induced by recourse itself.

Contributions: We suggest to restrict algorithmic recourse to recommend meaningful actions that improve prediction and outcome coherently. In contrast to work in the strategic modeling literature (Miller et al., 2020), our approach does not require to adapt the model at the cost of predictive performance. We justify the restriction by distinguishing two explanation goals: model audit and algorithmic recourse. Furthermore, we suggest a relaxation of MAR that does not require computing the structural counterfactual for Y. We derive assumptions given which the relaxation guarantees meaningful recourse.

3. Background and Notation

3.1. Causal Data Model

We model the data generating process using a structural causal model (SCM) $\mathcal{M} \in \Pi$ (Pearl, 2009; Peters et al., 2018). The model $\mathcal{M} = \langle \mathbb{F}, X, U \rangle$ consists of the endogenous variables $X \in \mathcal{X}$, the mutually independent exogenous variables $U \in \mathcal{U}$ and a sequence of structural equations $\mathbb{F} : \mathcal{U} \to \mathcal{X}$. The SCM entails a directed graph $\mathcal{G}$. The structural equations specify how X is determined from U. The index set of endogenous variables is denoted as D, the set of observed variables as $O \subseteq D$. The Markov blanket $MB_O(Y)$ is the minimal subset of O, such that $X_O \perp\!\!\!\perp Y | MB_O(Y)$.[2]

[2]Sometimes the MB is defined as minimal d-separating set.

3.2. Actionable Recourse

Following (Karimi et al., 2021), we model actions as structural interventions $a : \Pi \to \Pi$. In this framework, an action can be constructed as $a = do(\{X_i := a_i\}_{i \in I})$ where I is the index set of features to be intervened upon.

Assuming invertability of $\mathbb{F}$, the effect of an intervention for an individual can be determined using structural counterfactuals that are computed in three steps (Pearl, 2009): First, the factual distribution of exogenous variables given the endogenous variables is computed (abduction), i.e., $U^F = \mathbb{F}^{-1}(x^F)$. Second, the structural interventions corresponding to a, yielding $\mathbb{F}_a$ are performed (action). Finally, the counterfactuals are predicted $x^{SCF} = \mathbb{F}_a(u^F)$.

The optimization problem for recourse through minimal interventions (Karimi et al., 2021) is given by

$$a^* \in \underset{a \in \mathbb{A}}{\operatorname{argmin}} \operatorname{cost}(a, x^F) \text{ subject to } f(x^{SCF,a}) \geq t$$

where $\mathbb{A}$ is the action space, $x^{SCF,a} = \mathbb{F}_a(\mathbb{F}^{-1}(x^F))$, f is the predictor and t a threshold. Further constraints have been suggested, e.g., $x^{SCF,a} \in \mathcal{P}$lausible or $a \in \mathcal{F}$easible.

3.3. Generalizability and Intervention Stability

We leverage necessary conditions for invariant conditional distributions as derived in (Pfister et al., 2019). The authors introduce a d-separation based intervention stability criterion that is applied to modified version of $\mathcal{G}$. For every intervened upon variable X_l an auxiliary intervention variable, denoted as $\mathcal{I}_l$, is added as direct cause of X_l, yielding $\mathcal{G}^*$. The intervention variable can be seen as a switch between different mechanisms. A set $S \subseteq \{1, \dots, d\}$ is called *intervention stable* with respect to all actions if for all intervened upon variables X_l (where $l \in I^{\text{total}}$) the d-separation[3] $\mathcal{I}^l \perp\!\!\!\perp_{\mathcal{G}^*} Y | X_S$ holds in $\mathcal{G}^*$. The authors show that intervention stability implies an invariant condition distribution, i.e., for all actions $a, b \in \mathbb{A}$ with $I_a, I_b \subseteq I^{\text{total}}$ it holds that $P(Y^a | X_S) = P(Y^b | X_S)$ (Pfister et al. (2019), Appendix A).

4. Meaningful Algorithmic Recourse

Algorithmic recourse searches optimal actions by assessing the prediction over a range of interventions. However, since predictive models are designed to be employed in a static observational distribution, they often fail to generalize to interventional environments: For example, predictive models exploit all useful associations with the target, irrespective of the type of causal relationship between feature and target. As a consequence, variables that are not causal for the target can be causal for the prediction (Molnar et al., 2020). Interventions on such non-causal features may flip the prediction, but do not affect the underlying target. Thus, the model's

[3]Background on d-separation in Appendix B.1.

performance is not stable under such interventions.
This lack of intervention stability is problematic for two reasons: Firstly, following a recourse recommendations might not lead to improvement but rather game the predictor. Secondly, a refit with access to the post-recourse distribution in which the exploited associations are weakened will not honor the original recourse recommendation.
The related field of strategic prediction aims to adapt the model to strategic agent behavior (Miller et al., 2020). However, the approach has a catch: As Shavit et al. (2020) argue, designing the model to incentivize agent outcome (improvement) is often in conflict with achieving optimal predictive accuracy. For instance, in the example in Section 1 the model's reliance on *minivan* would need to be to be reduced to incentivize improvement.
We propose an alternative: Instead of altering the model such that gaming is not lucrative, we allow the model to use gameable associations but constrain algorithmic recourse recommendations to those that are meaningful (Definition 1). Therefore we require knowledge of the full SCM generating (observed) feature and target variables.

Assumption 1 *We assume knowledge of the SCM that generates X and Y. Furthermore, we assume the existence of $\mathbb{F}^{-1}(x^F)$ such that $\mathbb{F}^{-1}(x^F) = u^F$ (with $\mathbb{F}(u^F) = (x^F, y^F)$).*

Definition 1 *Meaningful actionable recourse (MAR) is algorithmic recourse (Karimi et al., 2021), with the additional constraint that the underlying target Y is improved coherently, i.e., $y^{SCF,a} \geq t$ where $(y^{SCF,a}, x^{SCF,a}) = \mathbb{F}_a(\mathbb{F}^{-1}(x^F))$.* [4]

Naturally, such a restriction reduces the insight that we gain into the model. However, in our view, depending on the goal of the explainee full model insight is not required. Therefore, we distinguish two tales of algorithmic recourse in Section 5. Since we may not have access to the causal model required to compute MAR we propose an alternative formulation of MAR that only relies on the predictor (Section 6).

5. The two tales of algorithmic recourse

Machine learning explanations as suggested by (Karimi et al., 2020b; 2021) may be used for two distinct purposes—for *model audit* and for *meaningful, actionable recourse*.
Model auditors aim to make sure that models meet desired standards (e.g., fairness) and extrapolate well to unseen regions. Model audit explanations can allow inspectors to contest model decisions, suggest model-debugging strategies or give insight into model behavior within and outside of the data distribution (Wachter et al., 2018; Freiesleben, 2020). Hence, these explanations must be maximally faithful to the prediction model.
Recourse recommendations on the other side need to satisfy various side constraints that are not related to the model. Even the causal dependencies between variables that are taken into account in algorithmic recourse are not reflected in the prediction model (Karimi et al., 2021). Recourse recommendations must also be actionable for the explainee, thus, changes in non-actionable features like age, ethnicity, or height are commonly prohibited (Ustun et al., 2019; Karimi et al., 2021). Moreover, recourse recommendation must be plausible, i.e., make realistic suggestions that are jointly satisfiable and prefer sparse over widespread action recommendations (Karimi et al., 2020a; Dandl et al., 2020).
In conclusion, model audit explanations are more complete and faithful to the model while recourse recommendations are more selective, faithful to the underlying process and account for the limitations of the data-subject. We believe that the selectivity and reliance of recourse recommendations on factors beside the model itself is not a limitation but essential to make explanations more relevant to the data-subject. In the same vein, we see MAR as another step towards making recourse recommendations more meaningful.
If however a data-subject is more interested in contesting or auditing the algorithmic decision, recourse recommendations are not suitable. Instead, we suggest that data-subjects should additionally receive model audit explanations upon request.

6. Formulation based on Effective Intervention Constraint

Even if we have access to the SCM modeling X, we may not know the structural equation generating Y. Over the course of this Section we therefore introduce *Effective actionable recourse* (EAR) as an alternative. Instead of computing whether the counterfacutal $y^{SCF,a}$ is flipped as desired (MAR), EAR restricts actions to exclusively intervene on causes of Y. Since any meaningful recourse recommendation intervenes on causes exclusively, the constraint does not exclude any *MAR* recommendation. And, as we demonstrate in this Section, access to the full causal graph $\mathcal{G}$ can suffice to ensure robustness of the model f to interventions on causes of Y and therefore meaningful recourse.

Definition 2 *Effective actionable recourse (EAR) is algorithmic recourse (Karimi et al., 2021) with the further constraint that only effective actions $a \in \mathcal{E}$ffective are allowed, i.e. that $I_a \subseteq Causes(Y)$*

We leverage research on invariant prediction (Pfister et al., 2019) (Section 3.3) to formalize under which assumptions EAR provides meaningful recourse. For the recommendation to lead to improvement, the model has to predict accurately in the respective action distribution. Since in-

[4] The assumption implies that an f exists such that $Y = f(X)$.

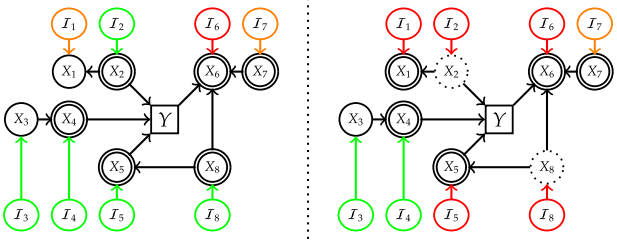

Figure 2. A schematic drawing illustrating under which interventions $\mathcal{I}_1, \ldots, \mathcal{I}_8$ the markov blanket (double circle) is intervention stable. In this setting, we consider the intervention variables to be independent treatment variables: We would like to know how the different actions influence the conditional distribution, irrespective of how likely they are to be applied. Therefore, they are modeled as parent-less variables. Green indicates intervention stability, red indicates no intervention stability. Orange indicates intervention stability of non-causal variables. Dotted variables are not observed (X_3). *Left:* Since all endogenous variables are observed, $MB(Y)$ is stable w.r.t. interventions on every endogenous cause of Y (Proposition 1). *Right:* Unobserved variables open paths between interventions on causes and Y.

tervention stability implies invariant conditionals (Section 3.3), for predictors that rely on intervention stable sets, EAR recommendations are as likely to lead to improvement as the predictor is able to correctly predict in the pre-recourse distribution.

In Figure 2 we illustrate under which interventions the minimal optimal set, the so-called Markov blanket, is intervention stable. Thereby, the set of observed variables O plays a crucial role. For example, if the minimal set of variables in $\mathcal{G}$ that d-separates all remaining variables from Y is observed, the markov blanket is intervention stable (Proposition 1). In contrast, an intervention on unobserved direct causes (e.g., X_2) (as well as interventions on non-causal variables) may alter the conditional $P(Y|MB(Y))$.

Proposition 1 *If all endogenous direct causes, direct children and spouses are observed, the markov blanket is stable with respect to interventions on all endogenous causes of Y.*

If the predictor is perfect and intervention stable, meaning that $f(x) = y$, EAR affects prediction and target coherently.

Proposition 2 *Assuming a perfect predictor f (i.e., $f(x) = y$ in the pre-recourse distribution) and that the predictor relies on a variable set S that is stable with respect to interventions on actionable causes of Y, for EAR recommendations with $P(X = x_S^{SCF}) > 0$ it holds that $\hat{y}^{SCF} = y^{SCF}$.*

Even if the conditional distribution is invariant, the joint distribution of the variables is most certainly affected if individuals act on recourse recommendations. If the interventional joint distribution extends the support of the observation distribution, the model would be forced to predict outside of the training distribution. EAR recommendations that extrapolate should therefore be handled with care.

If the conditional is stable and the predictor is able to perfectly predict Y under any action that may be recommended by recourse, ceteris paribus the recourse will be honored by an optimal refit of the model on pre- and post-recourse data.

7. Limitations and Discussion

In order to generate more robust and meaningful recourse recommendations affecting not only $\hat{Y}$ but also Y, we introduce meaningful algorithmic recourse (MAR) and a relaxation called effective algorithmic recourse (EAR).

Our approach is based on strong assumptions: We require that Y can be perfectly predicted from features and assume existence and knowledge of an underlying SCM with invertible structural equations. EAR requires that the model is stable w.r.t. interventions on causes. Evaluating the intervention stability requires knowledge of the causal graph.

Furthermore, one may argue that explanations should be maximally faithful to the model. Then, the gameability should be exposed. However, we argue that recommending gaming is problematic for both model authorities *and* individuals who seek robust recommendations that help them to improve. To reconcile both positions, we recommend offering additional model audit explanations upon request, but to only offer guarantees for MAR. In order to provide a variety of options, model authorities should aim to observe causes rather than non-causal variables. If the model is able to predict accurately based on causes alone, model audit and meaningful, actionable recourse converge.

It could be insisted that even if the conditional is invariant to MAR, the actions induce a shift in the distribution of Y. Indeed, if the threshold acts as a gatekeeper towards a limited good, the threshold itself may shift as a result. Consequently, even if the underlying target Y improves as desired, recourse may not be honored.

Further research is required to transfer the results into a probabilistic setting with imperfect causal knowledge. In such settings the robustness of recourse with respect to refits is particularly challenging since recourse may be applied selectively and thereby can open additional dependence paths that allow for an improved but different predictor.

We see our work as a first step towards more meaningful and robust recourse.

Acknowledgments

This project is funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A and the Graduate School of Systemic Neurosciences (GSN) Munich.

References

Barocas, S., Selbst, A. D., and Raghavan, M. The hidden assumptions behind counterfactual explanations and principal reasons. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 80–89, 2020.

Dandl, S., Molnar, C., Binder, M., and Bischl, B. Multi-objective counterfactual explanations. In *International Conference on Parallel Problem Solving from Nature*, pp. 448–469. Springer, 2020.

Freiesleben, T. Counterfactual explanations & adversarial examples–common grounds, essential differences, and potential transfers. *arXiv preprint arXiv:2009.05487*, 2020.

Geiger, D., Verma, T., and Pearl, J. Identifying independence in bayesian networks. *Networks*, 20(5):507–534, 1990.

Karimi, A.-H., Barthe, G., Balle, B., and Valera, I. Model-agnostic counterfactual explanations for consequential decisions. In *International Conference on Artificial Intelligence and Statistics*, pp. 895–905. PMLR, 2020a.

Karimi, A.-H., von Kügelgen, J., Schölkopf, B., and Valera, I. Algorithmic recourse under imperfect causal knowledge: a probabilistic approach. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 265–277. Curran Associates, Inc., 2020b.

Karimi, A.-H., Schölkopf, B., and Valera, I. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 353–362, 2021.

Miller, J., Milli, S., and Hardt, M. Strategic classification is causal modeling in disguise. In *International Conference on Machine Learning*, pp. 6917–6926. PMLR, 2020.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., and Bischl, B. Pitfalls to Avoid when Interpreting Machine Learning Models. *ICML workshop on XAI*, 2020.

Pawelczyk, M., Broelemann, K., and Kasneci, G. On Counterfactual Explanations under Predictive Multiplicity. In *Conference on Uncertainty in Artificial Intelligence*, pp. 809–818. PMLR, 2020.

Pearl, J. Causality : models, reasoning, and inference. pp. 487, 2009. ISBN: 9780521895606.

Peters, J., Janzing, D., and Scholkopf, B. *Elements of Causal Inference - Foundations and Learning Algorithms*. 2018. ISBN 978-0-262-03731-0. Issue: December.

Pfister, N., Williams, E. G., Peters, J., Aebersold, R., and Bühlmann, P. Stabilizing variable selection and regression. *arXiv preprint arXiv:1911.01850*, 2019.

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pp. 469–481, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367.

Rawal, K., Kamar, E., and Lakkaraju, H. Can I Still Trust You?: Understanding the Impact of Distribution Shifts on Algorithmic Recourses. *arXiv preprint arXiv:2012.11788*, 2020.

Shavit, Y., Edelman, B., and Axelrod, B. Causal Strategic Linear Regression. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8676–8686. PMLR, July 2020.

Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction and Search*. 2001. ISBN 0-262-19440-6.

Upadhyay, S., Joshi, S., and Lakkaraju, H. Towards Robust and Reliable Algorithmic Recourse. *arXiv preprint arXiv:2102.13620*, 2021.

Ustun, B., Spangher, A., and Liu, Y. Actionable recourse in linear classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 10–19, 2019.

Van Liebergen, B. et al. Machine learning: a revolution in risk management and compliance? *Journal of Financial Transformation*, 45:60–67, 2017.

Venkatasubramanian, S. and Alfano, M. The philosophical basis of algorithmic recourse. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 284–293, 2020.

Wachter, S., Mittelstadt, B., and Russell, C. Counterfactual Explanations without opening the black box: automated decisions and the GDPR. pp. 1–52, 2018.

Zeng, J., Ustun, B., and Rudin, C. Interpretable classification models for recidivism prediction. *arXiv preprint arXiv:1503.07810*, 2015.

A. Proofs

A.1. Proof of Proposition 1

If all endogenous direct causes, direct children and spouses are observed, the conditional $P(Y|MB(Y))$ is stable with respect to interventions on any set of endogenous causes of Y.
We prove the statement in four steps.

Given a graph $\mathcal{G}$ and an endogenous Y, the set of endogeneous direct parents, direct effects and direct parents of effects are the minimal d-separating set $S_{\mathcal{G}}$: Standard result, see e.g. Peters et al. (2018), Proposition 6.27.
The set $S_{\mathcal{G}^*}$ in the augmented graph $\mathcal{G}^*$ coincides with $S_{\mathcal{G}}$: The minimal d-separating set contains direct causes, direct effects and direct parents of direct effects. $\mathcal{I}_l$ is never a direct cause of X_l. Also, since $\mathcal{I}_l$ has no endogenous causes, it cannot be a direct effect. Furthermore, since we restrict interventions to be performed on causes, $\mathcal{I}_l$ cannot be a direct parent of a direct effect.
$S_{\mathcal{G}}$ is intervention stable: As follows, all intervention variables are d-separated from Y in $\mathcal{G}^*$ by $S_{\mathcal{G}}$. Therefore $S_{\mathcal{G}}$ is intervention stable.
Then also the markov blanket is intervention stable: Since d-separation implies independence $MB(Y) \subseteq S_{\mathcal{G}}$. Therefore, if $X_T \perp\!\!\!\perp Y|MB(Y)$ then also $X_T \perp\!\!\!\perp Y|S_{\mathcal{G}}$. If any element $s \in S_{\mathcal{G}}$ it holds that $s \notin MB(Y)$, then it must hold that $X_s \perp\!\!\!\perp Y|MB(Y)$. Therefore, if $X_T \perp\!\!\!\perp Y|MB(Y), X_s$ then also $X_T \perp\!\!\!\perp Y|MB(Y)$ and therefore any independence entailed by $S_{\mathcal{G}}$ also holds for $MB(Y)$. Since (Pfister et al., 2019) only require the independence that is implied by d-separation in their invariant conditional proof, the same implication holds for the $MB(Y)$. As follows, $P(Y|MB(Y))$ is invariant with respect to interventions on any set of endogenous causes.

A.2. Proof of Proposition 2

Assuming a perfect predictor f (i.e. $f(X) = Y$ in the pre-recourse distribution) and that the predictor relies on a variable set S that is stable with respect to interventions on actionable causes of Y, for EAR recommendations with $P(X = x_S^{SCF}) > 0$ it holds that $\hat{y}^{SCF} = y^{SCF}$.
Since S is intervention stable with respect to all actions a, for any two actions a, b it holds that $P(Y^a|X_S) = P(Y^b|X_S)$[5] and therefore if $y^a \neq y^b \Rightarrow x_S^a \neq x_S^b$. Since the predictor works perfectly under the null-action c (no interventions), $f(X_S) = y^c = y^a$ if $P^c(x^S) > 0$, which proves the statement.

[5]For a proof please refer to (Pfister et al., 2019), Appendix A.

B. Background

B.1. d-separation

Two variable sets X, Y are called d-separated (Geiger et al., 1990; Spirtes et al., 2001) by the variable set Z in a graph $\mathcal{G}$ ($X \perp\!\!\!\perp_{\mathcal{G}} Y|Z$), if, and only if, for every path p holds either (i) p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ where $m \in Z$ or (ii) p contains a collider $i \to m \leftarrow j$ such that m and all of its descendants n it holds that $m, n \notin Z$.