We thank the reviewers for their detailed and helpful feedback. We are glad that *all* reviewers agree on the novelty of the paper and recognise its original and significant contribution to the machine learning community. The reviewers acknowledge that the ideas presented in the paper are compelling, sound and appear to be effective (R3), offering a great add to the GP literature (R1) which is also supported by a solid and an interesting theoretical foundation (R2, R4). In the following discussion, we address three main points: comparison to related work, further intuition behind the DAG-GP model and additional experimental details. We respond point-by-point to the comments raised by each reviewer. We hope our detailed response below will further highlight the paper's quality and originality.

**R#1:** *(1) Knowledge of the causal graph.* In this paper we indeed assume the DAG to be known and we mention the integration of the DAG-GP model with a causal discovery algorithm as a future direction. Different causal discovery algorithms have been proposed which could be used as a pre-processing step. In addition, one could use the interventional data to discriminate across graphs within the Markov equivalence class. Analysing what happens when the DAG is unknown goes beyond the scope of this paper and remains an area of future research. We have clarified this point. *(2) Complexity analysis.* The time complexity of the algorithm is $\mathcal{O}(N^3)$ with $N$ denoting the size of $\mathcal{D}^I$. This complexity can be reduced by resorting to sparse GP approximations e.g. inducing points approximations.

**R#2:** *(1) Comparison to related work.* As highlighted by both R#1 and R#4, this approach is completely new in the GP literature. Existing multi-output GP models are not applicable to our setting (see line 79-83) and are thus not comparable to the DAG-GP model. In the causality literature, studies have focused on observational causal inference and transferability problems where the goal is to transfer the causal effects of *one* given variable across environments. On the contrary, our goal is to transfer information *across all* causal effects in a single environment thus existing studies are not comparable. We have further clarified this point in Section 1.2. We have also mentioned the causal prior in the related work and highlighted its definition in Section 3.2. See R#3 (1) for more details on relationship to existing works. *(2) Intuition behind the DAG-GP and experimental results.* We have used the additional page to provide a deeper analysis of the DAG-GP model and additional implementation details to aid understanding of the results. In particular, as suggested by R#3, we have shown how different covariance structures are linked to different DAGs and capture different levels of transfer within tasks. We have also added a brief intro to BO and AL and clarified how the DAG-GP can be used as a surrogate model within these two frameworks. *(3) Additional comments.* We have clarified Fig 4 and Fig 2. We have added an extended version of Fig 2 in the appendix where we have plotted the prior mean functions used by DAG-GP+, the posterior distributions obtained by the alternative models and the $Z$ interventional data.

**R#3:** *(1) Comparison to related work.* We have added [1] to the related work and discussed how its goal and setting differ from our paper making it not comparable. In our paper we assume full identifiability of the causal effects (line 108). When this is the case, [1] reduces to [2] and $p(\boldsymbol{Y}|do(\boldsymbol{X}=\boldsymbol{x}))$ is computed from observational data via do-calculus $\forall \boldsymbol{X}$. This option is included as a benchmark. When $\exists \boldsymbol{X}$ s.t. $p(\boldsymbol{Y}|do(\boldsymbol{X}=\boldsymbol{x}))$ is not identifiable and interventional data for $\boldsymbol{X}$ is not available, one need to resort to gID[1] or zID[3] to find an expression in terms of the available data/distributions. Our approach cannot deal with these settings as the integrating measures wouldn't be identifiable thus preventing the propagation of $p(f)$. Finally, notice that [1] aims at expressing the causal effects of *one* given variable in terms of the available distributions. While one can repeat the procedure for all possible intervention sets in $\mathcal{G}$, [1] is not gonna necessarily express all causal effects via a shared interventional distribution. The goal of our analysis is exactly that: identifying a *shared interventional representation* while providing a probabilistic model to transfer information in practice. Thanks for highlighting this connection. We will repeat the experiments using a kernel density estimator for the observational distributions so as to show consistency of the results across different non-parametric methods. See R#2 (1) above for more details on relationship to existing works. *(2) Intuition behind the DAG-GP and experimental results.* As suggested, we have shown how different covariance structures are linked to different DAGs and capture different levels of transfer within tasks. We have used the additional page to provide more experimental details and a brief introduction to BO and AL clarifying how the DAG-GP model can be used as a surrogate model. *(3) Additional comments.* Line 18: We have softened this statement. Notice however, that observational causal inference methods (and the do-calculus) provide accurate estimates of the causal effects for interventional values that are observed in the data and for which the observational distributions are accurately estimated. For interventional values not observed in the data one needs to resort to experiments. Line 49: While our results are indeed based on the do-calculus, applying these rules is not enough to transfer information. The do-calculus does not provide a way to obtain expressions in terms of the same *shared* base function which is our main contribution.

**R#4:** We have used the additional page to expand the discussion on single-task models in Section 4. *(1) Experimental details.* As mentioned above (see R#2 (2) and R#3 (2)) we will add further implementation details and we will provide the code base to ensure full reproducibility. We have expanded Section 5 in the supplement so as to further clarify the data generating mechanism for the synthetic data.

[1] Lee, Sanghack, Juan D. Correa, and Elias Bareinboim. "General Identifiability with Arbitrary Surrogate Experiments." UAI 2019. [2] Tian, Jin, and Judea Pearl. "A general identification condition for causal effects." Aaai/iaai 2002. [3] Bareinboim, Elias, and Judea Pearl. "Causal inference by surrogate experiments: z-identifiability." UAI 2012.