# Negative Dependence Tightens Variational Bounds

**Pierre-Alexandre Mattei** [1] **Jes Frellsen** [2]

## Abstract

Importance weighted variational inference (IWVI) is a promising strategy for learning latent variable models. IWVI uses new variational bounds, known as Monte Carlo objectives (MCOs), obtained by replacing intractable integrals by Monte Carlo estimates—usually simply obtained via importance sampling. Burda et al. (2016) showed that increasing the number of importance samples provably tightens the gap between the bound and the likelihood. We show that, in a somewhat similar fashion, increasing the negative dependence of importance weights monotonically increases the bound. To this end, we use the supermodular order as a measure of dependence. Our simple result provides theoretical support to several different approaches that leveraged negative dependence to perform efficient variational inference of deep generative models.

## 1. Introduction

Often, objective functions that arise in machine learning applications involve seemingly intractable high-dimensional integrals.Variational inference constitutes a toolbox of techniques that tackle this issue by replacing the objective function to maximise by a lower bound of it (that is supposed to be more tractable or easier to optimise).

A recent and promising approach to variational inference was proposed by Burda et al. (2016), notably building on prior work by Bornschein & Bengio (2015). The idea is simply to replace the intractable integrals by Monte Carlo estimates of it, and optimise the expected value of this approximation with respect to both model parameters and the randomness induced by the Monte Carlo approximation. Following Mnih & Rezende (2016), these new bounds are called **Monte Carlo objectives (MCOs)**, and are typically obtained using importance sampling with a parametrised posterior that can be optimised. This new flavour of variational inference is usually called **importance weighted variational inference (IWVI)**.

While they were originally developed to learn unsupervised deep latent variable models similar to variational autoencoders (VAEs, Kingma & Welling, 2014; Rezende et al., 2014), MCOs have been successfully applied to a diverse family of problems, from inference for Gaussian processes (Salimbeni et al., 2019) or sequential models (Maddison et al., 2017; Naesseth et al., 2018; Le et al., 2018) to missing data imputation (Mattei & Frellsen, 2019) or causal inference (Josse et al., 2020).

These empirical successes have been calling for theoretical developments. For example, a natural question is then: **how do properties of the Monte Carlo estimate translate into properties the variational bound?** This question, which is the main topic of this paper, has until now mostly been tackled from an *asymptotic* point of view. More specifically, most results are concerned with the behaviour of MCOs when the number of Monte Carlo samples is large (or when the variance is small). This contrasts with the fact that, in practice, the number of samples rarely exceeds a few dozens (for computational reasons), and the variance is very large if not infinite. Motivated by this gap between theory and practice, our perspective here is *non-asymptotic*. One exception to the asymptotic focus is a beautifully simple theorem of Burda et al. (2016): when the weights are exchangeable, **increasing the number of samples always improves the tightness of the bound**.

Recently, using **negatively dependent weights has been empirically shown to lead to tighter variational bounds** (Klys et al., 2018; Huang et al., 2019; Ren et al., 2019; Wu et al., 2019; Domke & Sheldon, 2019). The idea is to leverage the variance-reduction effect of negative dependence, leading hereby to more accurate unbiased estimates, and hopefully tighter bounds. First, **we revisit and challenge the popular heuristic that variance-reduction generally leads to tighter bounds**. Then, we give a **non-asymptotic result that shows that negative dependence**

[1] Université Côte d'Azur, Inria, Maasai project-team, Laboratoire J.A. Dieudonné, UMR CNRS 7351 [2] Department of Applied Mathematics and Computer Science, Technical University of Denmark. Correspondence to: PAM <pierre-alexandre.mattei@inria.fr>, JF <jefr@dtu.dk>.

*ICML 2020 workshop on Negative Dependence and Submodularity for ML*, Vienna, Austria, PMLR 119, 2020. Copyright 2020 by the author(s).

**provably tightens variational bounds**. Our result is quite simple and based on a stronger notion of dependence than covariance: the supermodular order.

## 2. Variational inference using Monte Carlo objectives

### 2.1. Inference via Monte Carlo objectives

We consider some data $\mathbf{x} \in \mathcal{X}$ governed by a latent variable $\mathbf{z} \in \mathcal{Z}$ through a model with density

$$p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z}) p(\mathbf{x}|\mathbf{z}), \tag{1}$$

with respect to a dominating measure on $\mathcal{X} \times \mathcal{Z}$. Typically, the latent variable models we focus on depend on many parameters that we would like to learn via (potentially approximate) maximum likelihood. Since $\mathbf{z}$ is hidden and only $\mathbf{x}$ is observed, the log-likelihood (or log-marginal likelihood if the model is Bayesian) is equal to

$$\ell = \log p(\mathbf{x}) = \log \int_{\mathcal{Z}} p(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z}. \tag{2}$$

A fruitful idea to approach $\ell$ is to replace the typically intractable integral $p(\mathbf{x})$ inside the logarithm by a Monte Carlo estimate of it. Of particular interest are unbiased estimates, since they lead to lower bounds of the likelihood $\ell$. Indeed, if $R$ is a random variable such that $R > 0$ and $\mathbb{E}[R] = p(\mathbf{x})$, then **the quantity $\mathcal{L} = \mathbb{E}[\log R]$ is a lower bound of the likelihood** $\ell$, by virtue of Jensen's inequality and the concavity of the logarithm. Moreover, the fact that, in $\mathcal{L}$, the expectation is now located *outside* of the logarithm means that $\mathcal{L}$ is more suited for stochastic optimisation techniques. **The lower bound $\mathcal{L}$ is called a Monte Carlo objective (MCO)**, and is maximised in lieu of the likelihood.

In this paper, we will study in particular importance sampling estimates of the form

$$R_K = \frac{1}{K} \sum_{k=1}^{K} \frac{p(\mathbf{x}|\mathbf{z}_k) p(\mathbf{z}_k)}{q(\mathbf{z}_k|\mathbf{x})}, \tag{3}$$

where $\mathbf{z}_1, \ldots, \mathbf{z}_K$ follow a *proposal distribution* $q(\mathbf{z}_1, \ldots, \mathbf{z}_K|\mathbf{x})$ that usually is a function of the data $\mathbf{x}$ (e.g. via a neural network, as in VAEs). The corresponding MCO is then $\mathcal{L}_K = \mathbb{E}[\log R_K]$, which may be optimised using stochastic optimisation.

### 2.2. General setting and notations

We consider a potentially infinite sequence of positive random variables $\mathbf{w} = (w_k)_{k \in \mathcal{K}}$ with common mean $\mu > 0$. This sequence, called the **sequence of importance weights**, is indexed by $\mathcal{K} = \{1, \ldots, K_{\text{max}}\}$, where $K_{\text{max}} \in \mathbb{N}^* \cup \{\infty\}$. The joint distribution of $\mathbf{w}$ is denoted by $Q$.

The Monte Carlo estimate of $\mu > 0$ is $R_K = S_K/K$, where $S_K = w_1 + \ldots + w_K$ and $K \in \mathcal{K}$. The **sequence of Monte Carlo objectives $\boldsymbol{\mathcal{L}}(Q) = (\mathcal{L}_K(Q))_{K \in \mathcal{K}}$**, is defined by

$$\mathcal{L}_K(Q) = \mathbb{E}_Q \left[ \log \left( \frac{1}{K} \sum_{k=1}^{K} w_k \right) \right] \tag{4}$$

$$= \mathbb{E}_Q \left[ R_K \right] = \mathbb{E}_Q \left[ \log S_K \right] - \log K. \tag{5}$$

It is possible to be slightly more general by replacing the uniform coefficients $1/K, \ldots, 1/K$ by a vector $\boldsymbol{\alpha}$ in the $K$-simplex $\Delta_K$. This leads to

$$\mathcal{L}_{\boldsymbol{\alpha}}(Q) = \mathbb{E}_Q \left[ \log \left( \boldsymbol{\alpha}^T \mathbf{w} \right) \right] = \mathbb{E}_Q \left[ \log \left( \sum_{k=1}^{K} \alpha_k w_k \right) \right]. \tag{6}$$

In particular $\mathcal{L}_{(1/K, \ldots, 1/K)}(Q) = \mathcal{L}_K(Q)$. Jensen's inequality ensures that $\mathcal{L}_{\boldsymbol{\alpha}}(Q) \leq \log \mu$. Note, however, that it is possible to have $\mathcal{L}_{\boldsymbol{\alpha}}(Q) = -\infty$ (we will show an example of this in the next section).

In the context of LVMs, $\mu = p(\mathbf{x})$; $\mathbf{w}$ is the sequence of importance weights; for all $K \in \mathcal{K}$ and $R_K$ is the importance sampling estimate of the likelihood $p(\mathbf{x})$, as in Equation (3). The non-uniform version $\mathcal{L}_{\boldsymbol{\alpha}}$ corresponds to using *multiple importance sampling*.

We believe that this general simple framework covers most ways of defining importance-sampling based MCOs, from the original ones of Burda et al. (2016), corresponding to i.i.d. weights with uniform coefficients, to the more elaborated ones of Huang et al. (2019), where the weights are correlated and not identically distributed.

## 3. Variance reduction as a heuristic towards tighter bounds

### 3.1. The variance heuristic

At its simplest level, what we call the **variance heuristic** may be informally formulated like this: **in a MCO, if Var$(R)$ gets smaller, then $R$ is a more accurate estimate of $\mathbb{E}[R] = \mu$, and the variational bound $\mathbb{E}[\log R]$ gets tighter**. It is possible to be more formal by Taylor-expanding the logarithm of $R$ around $\mu$:

$$\log(R) = \log(\mu) + \frac{R - \mu}{\mu} + \frac{(R - \mu)^2}{2\mu^2} + \text{Rem}(R). \tag{7}$$

The Taylor remainder Rem$(R)$ may be for example written exactly using its integral form. Then, assuming that Var$(R)$ is finite, computing the expectation leads to

$$\mathbb{E}[\log(R)] = \log(\mu) + \frac{\text{Var}(R)}{2\mu^2} + \mathbb{E}[\text{Rem}(R)]. \tag{8}$$

The variance heuristic can then be seen as a consequence of the assumption that, in Equation (8), the variance term dominates the remainder. In other words, it can be seen as *second order heuristic*. There are good reasons to believe that this assumption is reasonable when $R$ is very concentrated around $\mu$ (e.g. when $\text{Var}(R)$ is small). This is the rationale behind the results of Maddison et al. (2017, Proposition 1), Nowozin (2018, Proposition 1), Klys et al. (2018), Domke & Sheldon (2019, Theorem 3), and Huang & Courville (2019). Similar ideas (in a setting more general than the one of MCOs) are also present in Rainforth et al. (2018, Theorem 3). Huang et al. (2019, Section 4) also suggested to look at $\text{Var}[\log R]$ as an asymptotic indication of tightness of the bound.

Let us see what might sometimes break in this line of reasoning. First, we have no guarantee that the variance is actually finite. It is even quite common to encounter infinite variance importance sampling estimates, and we will give empirical evidence that the ones commonly used in VAEs have indeed infinite variance. Even assuming that the variance is finite, there are many situations where we could expect the Taylor remainder to be non-negligible. Indeed, the radius of convergence of the logarithm as a power series is quite small (the radius of $x \mapsto \log(x)$ is $\mu$ around $\mu$). This means that even a high order heuristic will not be accurate if $R$ gets far away from its mean $\mu$.

The fact that the bound is tightnened when the number of samples grows (Burda et al., 2016) can be seen a first example of success of the variance heuristic: adding more importance weights will both reduce Monte Carlo variance and tighten the bound. Let us now look briefly at the general case where $R$ can be any unbiased estimate (not necessarily obtained via importance sampling). In this very general setting, some assumptions must be made in order to be able to prove something. For example, we may wonder what happens when $R$ belongs to simple families of distributions. Sometimes, things will go as foretold by the heuristic, as seen below.

**Example 1 (a few successes of the variance heuristic).** *Let $R$ and $R'$ be either two gamma, two inverse gamma, or two log-normal distributions with finite and equal means and finite variances. Then*

$$\text{Var}[R] < \text{Var}[R'] \iff \mathbb{E}[\log R] > \mathbb{E}[\log R']. \quad (9)$$

The proof is available in Appendix B. The fact that these are exponential families suggests that a more general result may be hidden behind Example 1. What does it take to violate the heuristic using these kinds of simple distributions? While comparing two inverse gammas or two log-normals always respects it, simply blending these two family is enough to get severe violations.

**Example 2 (severe failure of the variance heuristic).** *Let $R$ be an inverse-gamma variable with finite mean. It is possible to find a log-normal random variable $R'$ such that*

- $\mathbb{E}[R] = \mathbb{E}[R']$, $\text{Var}[R] = \infty$, $\text{Var}[R'] < \infty$,
- $\mathbb{E}[\log R] > \mathbb{E}[\log R']$.

Again, the proof is available in Appendix B. In particular, we show that the gap $\mathbb{E}[\log R] - \mathbb{E}[\log R']$ can be made arbitrarily large by choosing the log-normal parameters (im)properly. This means that, when comparing MCOs, it is possible to be in a situation where **infinitely worse variance leads to an arbitrarily better bound**. It is also possible to be in a situation that is somehow the opposite of the previous example: the variance is finite, but the bound is not (e.g. if $R$ follows the finite moment log-stable distribution of Carr & Wu, 2003).

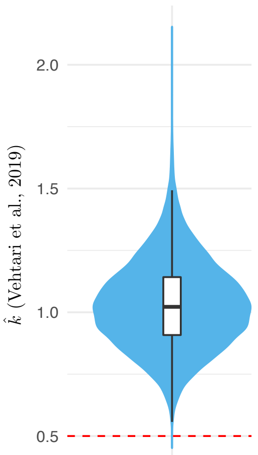

*Figure 1.* Importance sampling diagnostics for a VAE trained on MNIST. To each training digit corresponds a value of $\hat{k}$. Values of $\hat{k}$ above the dashed line correspond to digits whose weights have potentially infinite variance.

While it is not very surprising to find counter examples of these sorts, it is interesting to see that such severe failures may be observed using quite simple distributions. This phenomenon is reminiscent of the line of thought of Chatterjee & Diaconis (2018), who argued that the variance is not a very good metric for devising good importance sampling estimates.

**Is the variance finite in practice?** It is often the case that importance weights have infinite variance. We provide empirical evidence that this is the case in the simple case of a VAE trained on MNIST (Figure 1). After training, we compute 10,000 weights for each digits that we use to compute the $\hat{k}$ diagnostic of Vehtari et al. (2019). Most digits have a $\hat{k} > 0.5$, and are therefore suspect of having infinite variance. This illustrates again the shortcomings of the variance. More details on this experiment are provided in Appendix C.

## 4. Negative dependence and tighter bounds

A popular branch of variance reduction techniques is based on **negative dependence**. In its simplest form, this idea is based on the fact that

$$\mathrm{Var}(\boldsymbol{\alpha}^T \mathbf{w}) = \sum_{k=1}^{K} \alpha_k^2 \mathrm{Var}(w_k) + 2 \sum_{1 \le k < k' \le K} \alpha_k \alpha_{k'} \mathrm{Cov}(w_k, w_{k'}), \quad (10)$$

which means that negative covariances in the right hand side of Equation (10) will lead to a smaller variance of $\boldsymbol{\alpha}^T \mathbf{w}$. **Anthitetic sampling** is for example a famous variance reduction technique based on this idea (see e.g. Owen, 2013, Section 8.2).

Variants of this rationale were used successfully in the MCO context by Klys et al. (2018), Huang et al. (2019), Ren et al. (2019), Wu et al. (2019), and Domke & Sheldon (2019). Their motivations were essentially based on variants of the variance heuristic: since negative dependence can reduce the variance, it might also improve the bound. Our goal here is to prove that negative dependence can indeed tighten the bound, giving hereby a non-asymptotic theoretical justification for the works aforementioned.

Let $\mathbf{w} \sim Q_1$ and $\mathbf{v} \sim Q_2$ be two $K$-dimensional random variables with identical marginals, i.e. $w_k \overset{d}{=} v_k$ for all $k \in \{1, \dots, K\}$. What mathematical sense could we give to the sentence **"the coordinates of w are more negatively dependent than those of v"**? Again, stochastic orders provide good tools for assessing this. Indeed, the idea of **dependence orders** is to define binary relations $\preceq$ between distributions such that $Q_1 \preceq Q_2$ means that, in some sense, the coordinates of $\mathbf{w} \sim Q_1$ are more negatively dependent that those of $\mathbf{v} \sim Q_2$. A detailed overview of these dependence-based stochastic orders may be found in Shaked & Shanthikumar (2007, Chapter 9). A prominent example is the supermodular order.

**Definition 1.** *A function $\phi : \mathbb{R}^K \to \mathbb{R}$ is **supermodular** if, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^K$,*

$$\phi(\min(\mathbf{x}, \mathbf{y})) + \phi(\max(\mathbf{x}, \mathbf{y})) \ge \phi(\mathbf{x}) + \phi(\mathbf{y}), \quad (11)$$

*with min and max functions applied elementwise.*

*Let $Q_1$ and $Q_2$ be two probability distributions over $\mathbb{R}^K$. We say that $Q_1$ **is smaller than** $Q_2$ **in the supermodular order** when*

$$\mathbb{E}_{Q_1}[\phi(\mathbf{w})] \le \mathbb{E}_{Q_2}[\phi(\mathbf{v})]$$

*for all supermodular functions $\phi$ such that the involved expectations exist. We denote $Q_1 \preceq_{\mathrm{SM}} Q_2$.*

The supermodular order is one of the most popular stochastic orders when it comes to quantify dependence (see e.g. Müller & Scarsini, 2000; Shaked & Shanthikumar, 2007, Chapter 9), notably in the economics and insurance literature (see e.g. Müller, 1997; Meyer & Strulovici, 2012; 2015). Joe (1997, Section 2.2.3) proposed a set of nine axioms that would characterise good dependence orders. A few years later, Müller & Scarsini (2000) proved that the supermodular order satisfied all of these desirable properties. Here is a simple example of supermodular ordering: for two distributions $Q_1, Q_2$ with identical marginals, if the coordinates of $\mathbf{w} \sim Q_1$ are negatively associated, and those of $\mathbf{v} \sim Q_2$ are independent, then $Q_1 \preceq_{\mathrm{SM}} Q_2$ (Christofides & Vaggelatou, 2004).

An important example of supermodular function is the following: let $\phi$ be a convex function and $\boldsymbol{\alpha}$ a vector with non-negative coefficients, then $\mathbf{w} \mapsto \phi\left(\boldsymbol{\alpha}^T \mathbf{w}\right)$ is supermodular. Using this fact with $\phi = -\log$ immediately leads to the following monotonicity theorem.

**Theorem 1 (negative dependence tightens the bound).** *For all pairs $Q_1, Q_2$ of probability distributions over $\mathbb{R}^K$,*

$$Q_1 \preceq_{\mathrm{SM}} Q_2 \implies \mathcal{L}_{\boldsymbol{\alpha}}(Q_1) \ge \mathcal{L}_{\boldsymbol{\alpha}}(Q_2). \quad (12)$$

In other words, **the lower bound gets tighter when the weights get more negatively dependent** (in the supermodular sense). This gives a theoretical support to the successful recent applications of negative dependence to tighten variational bounds.

## 5. Conclusion

The main limitation of our result is that it is difficult to control the supermodular order in practice. A silver lining to this is the central role played by the supermodular order among dependence measures. In particular, the popular notion of **negative association** is in a sense stronger than the supermodular order (for a more general result than the simple one from Christofides & Vaggelatou, 2004 cited above, see Shaked & Shanthikumar, 2007, Theorem 9.E.8).

An interesting question is whether or not these sorts of investigations could provide a guide to design proposal distributions with the "right amount of correlation" required to tighten bounds.

## References

Bornschein, J. and Bengio, Y. Reweighted wake-sleep. In *International Conference on Learning Representations*, 2015.

Burda, Y., Grosse, R., and Salakhutdinov, R. Importance weighted autoencoders. In *International Conference on Learning Representations*, 2016.

Carr, P. and Wu, L. The finite moment log stable process and option pricing. *The Journal of Finance*, 58(2):753–777, 2003.

Chatterjee, S. and Diaconis, P. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

Christofides, T. C. and Vaggelatou, E. A connection between supermodular ordering and positive/negative association. *Journal of Multivariate analysis*, 88(1):138–151, 2004.

Domke, J. and Sheldon, D. R. Divide and couple: Using monte carlo variational objectives for posterior approximation. In *Advances in neural information processing systems*, pp. 338–347, 2019.

Huang, C.-W. and Courville, A. Note on the bias and variance of variational inference. *arXiv preprint arXiv:1906.03708*, 2019.

Huang, C.-W., Sankaran, K., Dhekane, E., Lacoste, A., and Courville, A. Hierarchical importance weighted autoencoders. In *International Conference on Machine Learning*, pp. 2869–2878, 2019.

Joe, H. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

Josse, J., Mayer, I., and Vert, J.-P. MissDeepCausal: causal inference from incomplete data using deep latent variable models. *Openreview preprint*, 2020.

Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014.

Klys, J., Bettencourt, J., and Duvenaud, D. Joint importance sampling for variational inference. *Openreview preprint*, 2018.

Le, T. A., Igl, M., Rainforth, T., Jin, T., and Wood, F. Auto-encoding sequential Monte Carlo. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=BJ8c3f-0b.

Maddison, C. J., Lawson, J., Tucker, G., Heess, N., Norouzi, M., Mnih, A., Doucet, A., and Teh, Y. Filtering variational objectives. In *Advances in Neural Information Processing Systems*, pp. 6573–6583, 2017.

Mattei, P.-A. and Frellsen, J. MIWAE: Deep generative modelling and imputation of incomplete data sets. In *International Conference on Machine Learning*, pp. 4413–4423, 2019.

Meyer, M. and Strulovici, B. Increasing interdependence of multivariate distributions. *Journal of Economic Theory*, 147(4):1460–1489, 2012.

Meyer, M. and Strulovici, B. Beyond correlation: Measuring interdependence through complementarities. Economics Series Working Papers 655, University of Oxford, Department of Economics, 2015.

Mnih, A. and Rezende, D. Variational inference for Monte Carlo objectives. In *International Conference on Machine Learning*, pp. 2188–2196, 2016.

Müller, A. Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics*, 21(3):219–223, 1997.

Müller, A. and Scarsini, M. Some remarks on the supermodular order. *Journal of Multivariate Analysis*, 73(1): 107–119, 2000.

Naesseth, C., Linderman, S., Ranganath, R., and Blei, D. Variational sequential Monte Carlo. In *International Conference on Artificial Intelligence and Statistics*, pp. 968–977, 2018.

Nowozin, S. Debiasing evidence approximations: On importance-weighted autoencoders and jackknife variational inference. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=HyZoi-WRb.

Owen, A. B. *Monte Carlo theory, methods and examples*. 2013.

Rainforth, T., Cornish, R., Yang, H., Warrington, A., and Wood, F. On nesting Monte Carlo estimators. In *International Conference on Machine Learning*, pp. 4267–4276, 2018.

Ren, H., Zhao, S., and Ermon, S. Adaptive antithetic sampling for variance reduction. In *International Conference on Machine Learning*, pp. 5420–5428, 2019.

Rezende, D., Mohamed, S., and Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 1278–1286, 2014.

Salimbeni, H., Dutordoir, V., Hensman, J., and Deisenroth, M. P. Deep gaussian processes with importance-weighted variational inference. *arXiv preprint arXiv:1905.05435*, 2019.

Shaked, M. and Shanthikumar, J. G. *Stochastic orders*. Springer Science & Business Media, 2007.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., and Gabry, J. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2019.

Wu, M., Goodman, N., and Ermon, S. Differentiable antithetic sampling for variance reduction in stochastic variational inference. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2877–2886, 2019.

## A. Do *not* have an appendix here

***Do not put content after the references.*** Put anything that you might normally include after the references in a separate supplementary file.

We recommend that you build supplementary material in a separate document. If you must create one PDF and cut it up, please be careful to use a tool that doesn't alter the margins, and that doesn't aggressively rewrite the PDF file. pdftk usually works fine.

**Please do not use Apple's preview to cut off supplementary material.** In previous years it has altered margins, and created headaches at the camera-ready stage.