Algorithm 1 Automated Blocking

$i \leftarrow 0$
$\Psi_0 \leftarrow \Psi_{\text{scalar}}$
loop:
    $i \leftarrow i + 1$
    Generate samples of Θ under Ψ_{i-1}, where X_j represents the sample chain of θ_j
    Discard initial 50% of each chain X_j
    $\rho_{j,k} \leftarrow \text{cor}(X_j, X_k)$
    $d_{j,k} \leftarrow 1 - |\rho_{j,k}|$
    Construct distance matrix D from elements $d_{j,k}$
    Construct hierarchical clustering tree T from D
    $\boldsymbol{\Psi}_{\text{cand}} \leftarrow \{\Psi(T_{\text{cut}=0}), \Psi(T_{\text{cut}=0.1}), \Psi(T_{\text{cut}=0.2}), \ldots, \Psi(T_{\text{cut}=1})\}$
    $\Psi_i \leftarrow \text{argmax}_{\Psi \in \boldsymbol{\Psi}_{\text{cand}}} E(\Psi)$
if $E(\Psi_i) > E(\Psi_{i-1})$ **and** $\Psi_i \neq \Psi_{i-1}$ **then goto** *loop*
$\Psi_{\text{AutoBlock}} \leftarrow \Psi_i$
return $\Psi_{\text{AutoBlock}}$

somewhat arbitrary initial 50% of all samples. This should not be confused with a traditional "burn-in," whose purpose is to "forget" the initial state and ensure convergence to the target distribution. Instead, discarding these initial samples allows all adaptive scalar and block samplers ample time to self-tune, and thereby achieve their theoretically optimal algorithmic efficiency. The choice of 50% is largely arbitrary, and excessive in most cases, and could almost certainly be relaxed without affecting algorithm performance.

Empirical correlations are transformed into distances using the transformation $d_{j,k} = 1 - |\rho_{j,k}|$. The form of this transformation is selected to induce several properties for elements of the distance matrix D: the main diagonal consists of zeros; strong correlation results in $d \approx 0$; weak or zero correlation results in $d \approx 1$; and correlations of ρ and $-\rho$ result in the same distance.

We use the R function `hclust` to create the hierarchical tree T from the distance matrix D. The default "complete linkage" clustering (Everitt, 2011, chapter 4) is appropriate, since this ensures that all parameters within each cluster have a minimum absolute pairwise correlation. At height $h \in [0, 1]$ in T, the absolute correlation between parameter pairs (within clusters) is at least $1 - h$.

We use the R function `cutree` for cutting the hierarchical clustering tree T at a specified height $h \in [0, 1]$ to produce disjoint parameter groupings, which may be used to define parameter blocks for the purpose of MCMC sampling. We justify this means of generating parameter sampling blocks, insofar as to increase algorithmic efficiency we strive to group *correlated* parameters into sampling