






Article

Refining the Allostatic Self-Efficacy Theory of Fatigue and Depression Using Causal Inference

Alexander J. Hess ^{1,*} , Dina von Werder ^{1,2,3} , Olivia K. Harrison ^{1,4} , Jakob Heinzle ¹ 
and Klaas Enno Stephan ^{1,5} 

- ¹ Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, 8032 Zurich, Switzerland; vonwerder@biomed.ee.ethz.ch (D.v.W.); olivia.harrison@otago.ac.nz (O.K.H.); heinzle@biomed.ee.ethz.ch (J.H.); stephan@biomed.ee.ethz.ch (K.E.S.)
- ² Institute of Medical Technology, Brandenburg University of Technology Cottbus-Senftenberg, 03048 Cottbus, Germany
- ³ Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany
- ⁴ Department of Psychology, University of Otago, Dunedin 9016, New Zealand
- ⁵ Max Planck Institute for Metabolism Research, 50931 Cologne, Germany
- * Correspondence: hess@biomed.ee.ethz.ch

Abstract: Allostatic self-efficacy (ASE) represents a computational theory of fatigue and depression. In brief, it postulates that (i) fatigue is a feeling state triggered by a metacognitive diagnosis of loss of control over bodily states (persistently elevated interoceptive surprise); and that (ii) generalization of low self-efficacy beliefs beyond bodily control induces depression. Here, we converted ASE theory into a structural causal model (SCM). This allowed identification of empirically testable hypotheses regarding causal relationships between the variables of interest. Applying conditional independence tests to questionnaire data from healthy volunteers, we sought to identify contradictions to the proposed SCM. Moreover, we estimated two causal effects proposed by ASE theory using three different methods. Our analyses identified specific aspects of the proposed SCM that were inconsistent with the available data. This enabled formulation of an updated SCM that can be tested against future data. Second, we confirmed the predicted negative average causal effect from metacognition of allostatic control to fatigue across all three different methods of estimation. Our study represents an initial attempt to refine and formalize ASE theory using methods from causal inference. Our results confirm key predictions from ASE theory but also suggest revisions which require empirical verification in future studies.

Keywords: allostatic self-efficacy; fatigue; depression; causality; structural causal model; directed acyclic graph; *d*-separation; conditional independence; average causal effect



Citation: Hess, A.J.; von Werder, D.; Harrison, O.K.; Heinzle, J.; Stephan, K.E. Refining the Allostatic Self-Efficacy Theory of Fatigue and Depression Using Causal Inference. *Entropy* **2024**, *26*, 1127. <https://doi.org/10.3390/e26121127>

Academic Editor: Alessandro Giuliani

Received: 14 June 2024

Revised: 11 December 2024

Accepted: 13 December 2024

Published: 23 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fatigue is a prominent symptom of major clinical significance in numerous disorders across medical disciplines [1,2]. It is fundamentally disabling for patients and profoundly affects their quality of life [3]. Fatigue is a common feature across a wide range of immunological and endocrine disorders, cancer, and neuropsychiatric diseases. In particular, it constitutes one of the core diagnostic criteria of major depression in standard psychiatric classification schemes (ICD-10 and DSM-5; Refs. [4,5]).

The clinical concept of fatigue is a heterogeneous construct, and the subjective perception of chronic fatigue needs to be distinguished from objectively observable fatigability of cognitive and motor processes [6] as well as from tiredness, which not only differ in phenomenology but are thought to be mechanistically distinct [7]. This study focuses on subjectively perceived fatigue.

The pathophysiological mechanisms leading to fatigue are likely diverse [6]. Previous theories have focused on a variety of neurophysiological, immunological, and inflammatory

processes. Unfortunately, there are no mechanistically interpretable clinical tests available for fatigue that could be used to guide individual treatment [6].

More recently, a novel perspective on fatigue has been proposed—the ‘allostatic self-efficacy’ theory (ASE; Refs. [6–8]). The ASE theory is based on computational concepts of brain–body interactions [7,8] which, in turn, are conceptually related to and inspired by Bayesian theories of perception (predictive coding; Ref. [9]) and action (active inference; Ref. [10]). Given these Bayesian roots, ‘beliefs’—i.e., probabilistic representations of latent variables—therefore play a central role in the ASE theory. However, in its current formulation, the theory does not explicitly comment on whether these beliefs are conscious or subconscious.

The ASE theory emphasizes the role of two cognitive processes for fatigue: interoception and metacognition. Interoception corresponds to the perception of bodily states and is of major importance for understanding determinants of mental health [11,12]. Many contemporary concepts of interoception are grounded in Bayesian theories of perception and conceptualize interoception as an inference process based on the brain’s generative model of sensory inputs from the body [8,12–16]. More specifically, interoception can be conceptualized as “inferences about bodily (physiological and biochemical) states that are coupled to regulatory processes which serve to control these states” [17]. Metacognition can be summarized as ‘cognition about cognition’ [18], comprising a variety of evaluation processes by which the brain monitors its own performance. Building on a generic mathematical model of brain–body interactions, the ASE theory describes how the brain attempts to control bodily states via monitoring interoceptive surprise (as an index of the degree of dyshomeostasis; Ref. [7]).

In brief, the ASE theory proposes that the subjective experience of fatigue arises when, in a situation of persistent dyshomeostasis (and thus enduringly elevated interoceptive surprise), the brain arrives at the metacognitive diagnosis that its control over bodily states is failing; a condition also referred to as low allostatic self-efficacy. Put differently, the ASE theory views fatigue as a feeling state that arises from a specific metacognitive process—i.e., the insight that regulatory actions are not capable of reducing interoceptive surprise—and which signals the imperative need to rest because active regulatory actions fail to resolve dyshomeostasis. While rest also represents a regulatory action, here, we contrast it to targeted regulatory actions that involve active behavior of an overt or latent (autonomic nervous system related) nature. The theory predicts that if rest cannot resolve dyshomeostasis and thus reduce interoceptive surprise, fatigue becomes entrenched and chronic. Furthermore, if low self-efficacy beliefs generalize beyond the body, leading to a general sense of helplessness and perceived lack of control, this is postulated to trigger the onset of depression [7,19].

At present, several explanations of clinically relevant fatigue exist; however, all of these are restricted to specific disorders, such as diseases due to infections, inflammation, or autoimmune processes. For example, classical accounts of fatigue are mainly related to infectious disorders and the associated ‘sickness behavior’ (for a review, see [20]). Broader conceptualizations, with links to cognition, exist but are restricted to inflammatory conditions (for a review, see [21]). In addition, concepts of fatigue have been proposed for other specific disorders; for example, Heitmann et al. (2022; Ref. [22]) interpreted fatigue in multiple sclerosis as the consequence of inflammation-induced reduction in monoaminergic transmission and subsequent impairment of reward processing. By contrast, so far, the ASE theory is arguably the only general concept of fatigue that explains its ubiquitous occurrence across chronic disorders (with fundamental differences in the underlying pathophysiology) and combines biological, cognitive, and computational (algorithmic) perspectives. The theory offers testable predictions based on either (i) computational quantities (prediction error or surprise) which can be estimated from behavioral and/or neurophysiological data or on (ii) self-reported data about perceived control over bodily states (metacognition of allostatic control). In this study, we focus on the latter option.

Empirically, there is initial evidence that metacognition of allostatic control—as measured by a self-report questionnaire—is inversely associated with fatigue, as predicted by ASE theory [19]. However, a comprehensive investigation of the predictions made by the ASE theory is still lacking to date. Furthermore, as with almost all disease concepts in psychiatry, ASE theory has been formulated verbally, but not as a precise causal model.

Here, we present an initial attempt to tackle the latter issue. To this end, we identify variables and causal relations that are part of the ASE theory, namely metacognition of allostatic control (M ; specifically, the feeling of being in control over one's own bodily states), fatigue (F), general self-efficacy (S), and depression (D). We then formalize the causal structure implied by the ASE theory in the language of causal inference, more precisely, in the form of a structural causal model (SCM; Refs. [23–25]). Notably, the proposed SCM only contains variables and relations that are explicitly mentioned by the ASE theory in its current formulation [7]. In contrast to classical probabilistic models, an SCM induces not only an observational distribution but also a set of so-called interventional distributions. In other words, an SCM predicts how a system reacts under interventions [26]. We made use of a publicly available empirical dataset to test key aspects of the structure of the proposed SCM. Moreover, we used established methods for the estimation of average causal effects, focusing on central aspects of the ASE theory.

2. Materials and Methods

2.1. Empirical Dataset

In this work, we used data from a previous study conducted at the Translational Neuro modeling Unit (TNU) Zurich, the perception of breathing in the human brain (PBIHB) study; a detailed description of the dataset can be found elsewhere [27]. It comprises behavioral, questionnaire, and neuroimaging data from 60 healthy individuals. The questionnaire data used for our analysis are freely available for download from the Zenodo open data repository at <https://doi.org/10.5281/zenodo.10992529> (accessed on 12 December 2024). Participants completed a battery of psychological questionnaires assessing subjective affective measures, both general and breathing-specific subjective interoceptive beliefs, as well as measures of general positive and negative affect, resilience, self-efficacy, and fatigue.

For our analysis, we focused on the following measures as representations of the central quantities of the ASE theory:

- fatigue (F): Fatigue Severity Scale (FSS)
- general self-efficacy (S): General Self-Efficacy Scale (GSES)
- depression (D): Centre for Epidemiologic Studies Depression Scale (CES-D)
- metacognition of allostatic control (M): Sum of the subscales 3 (not worrying) and 8 (trusting) of the Multidimensional Assessment of Interoceptive Awareness (MAIA_{3,8}).

One important caveat is that, to our knowledge, there does not yet exist a measure specifically developed for the construct of M (metacognition of allostatic control, i.e., the feeling of being in control over one's own bodily states). In this study, as a proxy measure, we used the sum of the subscales 3 and 8 of the MAIA questionnaire. These subscales reflect an individual's tendency not to experience distress in response to bodily inputs signaling dyshomeostasis and to perceive the body as a safe place, respectively. The sum of these subscales was used in a previous study testing predictions from ASE theory [19] and may currently represent the best approximation to M that is easily applied in practice.

2.2. SCM of the ASE Theory

An SCM [23,25] over variables $\mathbf{X} = [X_1, \dots, X_n]$ comprises a set of structural equations and distributions of the noise variables (a formal definition of an SCM is provided in Appendix Definition A1). The structural equations, together with the noise distributions, induce the observational distribution $\mathbb{P}_{\mathbf{X}}$ as a simultaneous solution to the structural equations [26]. In addition to the observational distribution, an SCM induces interventional distributions. Each intervention denotes a scenario in which we fix a certain subset of the variables to a certain value, e.g., $\mathbb{P}_{do(X_1:=x_1)}$.

We restricted our formulation of an SCM of the ASE theory to variables and relations that the current formulation of the ASE theory [7] explicitly comments on. Additionally, we considered the effects of the variables age (A) and gender (G), which are not directly part of the ASE theory, but are known to be associated with several variables of the SCM and may act as confounders. Under assumptions of linearity and normality, the SCM of the ASE theory takes the following form:

$$A = N_a \tag{1}$$

$$G = N_g \tag{2}$$

$$M = \theta_1 A + \theta_2 G + N_m \tag{3}$$

$$F = \theta_3 M + \theta_4 A + \theta_5 G + N_f \tag{4}$$

$$S = \theta_6 A + \theta_7 G + N_s \tag{5}$$

$$D = \theta_8 F + \theta_9 S + \theta_{10} FS + \theta_{11} A + \theta_{12} G + N_d \tag{6}$$

where A stands for age, G for gender, M for metacognition of allostatic control, F for fatigue, S for general self-efficacy, and D for depression, and where N_i are jointly independent noise variables. $\forall i \neq g, N_i$ follows a normal distribution and N_g is a Bernoulli random variable.

Figure 1 displays a graphical summary of the causal structure implied by the ASE theory in the form of a directed acyclic graph (DAG) J_0 . The directed edge from metacognition of allostatic control (M) to fatigue (F) represents the prediction that fatigue arises as a consequence of a metacognitive diagnosis by the brain—i.e., the brain concludes that it has low control over its bodily states. When this low allostatic self-efficacy (for which fatigue is the accompanying feeling state) is combined with beliefs of lack of control in other domains than the body (low general self-efficacy), this is predicted to lead to the onset of depression. These effects are represented by the directed edges from fatigue (F) to depression (D) and from general self-efficacy (S) to depression (D). The DAG J_0 in Figure 1 is representative of the induced observational distribution \mathbb{P} and the interventional distributions induced by interventions on metacognition of allostatic control ($M; \mathbb{P}_{do(M:=m)}$), fatigue ($F; \mathbb{P}_{do(F:=f)}$), or general self-efficacy ($S; \mathbb{P}_{do(S:=s)}$).

DAG J_0

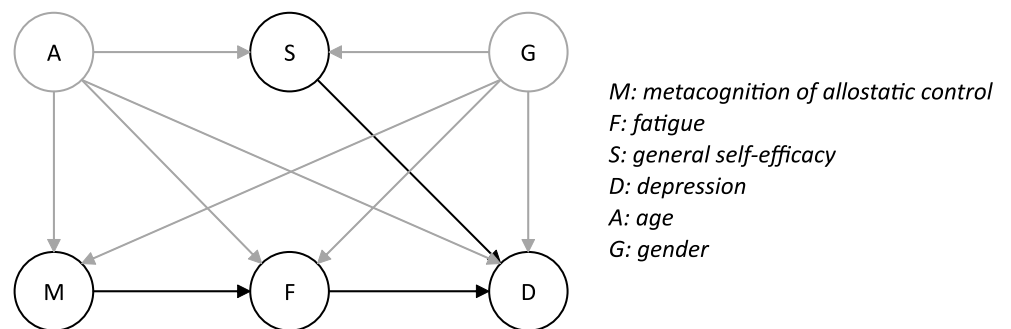


Figure 1. Directed acyclic graph (DAG) J_0 summarizing the key proposal of the allostatic self-efficacy theory (ASE; Ref. [7]). The DAG J_0 is representative of the induced observational distribution \mathbb{P} and the interventional distributions induced by interventions on metacognition of allostatic control ($M; \mathbb{P}_{do(M:=m)}$), fatigue ($F; \mathbb{P}_{do(F:=f)}$), or general self-efficacy ($S; \mathbb{P}_{do(S:=s)}$). The other variables in the graph are depression (D), age (A), and gender (G). Black edges represent causal directions as proposed explicitly by the ASE theory, grey edges represent effects from age and gender that are not explicitly part of the ASE theory but are included to account for potentially confounding effects.

When taking a closer look at the causal graph in Figure 1, there are a number of points worth highlighting. (i) There is no direct link between metacognition of allostatic control (M) and general self-efficacy (S). (ii) There is no direct link between metacognition of allostatic control (M) and depression (D). All of its influence is mediated by fatigue (F).

(iii) There is no direct link between fatigue (F) and general self-efficacy (S). While these three links are, in principle, plausible causal influences, they were not included in the original formulation of the ASE theory [7]. Whether these links should be included in a revision of the ASE theory can, in principle, be tested using methods of causal inference, given appropriate readouts of the involved quantities and relying on the assumption of the Markov condition.

2.3. Statistical Analysis

Our hypotheses, as well as the entire analysis, were pre-registered in a time-stamped analysis plan that is publicly available on the Zenodo open data repository at <https://doi.org/10.5281/zenodo.10559656> (accessed on 12 December 2024). Below, we explicitly highlight any deviations from the pre-specified analysis plan. The analysis code is available at <https://github.com/alexjhess/pbihb-ase-causality> (accessed on 12 December 2024). The analysis pipeline underwent an internal code review by a researcher not involved in the initial data analysis to identify errors and ensure the reproducibility of our results.

2.3.1. Causal Structure of ASE Theory in the PBIHB Dataset

Learning causal structure from observational data is inherently difficult. One reason for this is the existence of models that are observationally but not interventionally equivalent [25,26,28,29]. This has several implications (e.g., see [26]), one of them being that, without assumptions, it is impossible to learn causal structures from observational data.

In graphical models, the Markov condition (see, e.g., [30]) is a formalization of the following principle (sometimes referred to as Reichenbach's common cause principle): If two random variables X and Y are dependent, then there must be some cause-effect structure that explains the observed dependence. That is, either X causes Y , or Y causes X , or another unobserved variable H causes both X and Y , or some combination of the aforementioned [31]. A formal definition of the Markov condition is presented in Appendix Definition A2. The Markov condition establishes a connection from graphical separation properties (d -separation; see Appendix Definition A3 for a formal definition) to conditional independencies in the distribution. Any distribution induced by an acyclic SCM satisfies the Markov condition with respect to the corresponding graph [25,32]. Hence, the Markov condition is typically considered to be a mild assumption.

Assuming that the observational distribution \mathbb{P} induced by the SCM of the ASE theory (Equations (1)–(6)) is Markov with respect to the DAG J_0 , we tested whether we found any contradictions to the structure of the DAG J_0 in the PBIHB dataset. More precisely, we examined the three predictions described in the last paragraph of Section 2.2 and formalized them as part of our pre-registered Hypothesis 1: Data from the PBIHB study satisfy the following conditional independence statements:

- (i) $M \perp\!\!\!\perp S \mid A, G$
- (ii) $M \perp\!\!\!\perp D \mid F, A, G$ and $M \perp\!\!\!\perp D \mid F, A, G, S$
- (iii) $F \perp\!\!\!\perp S \mid A, G$ and $F \perp\!\!\!\perp S \mid A, G, M$

As a statistical test for conditional independence, we used an asymptotic χ^2 test on the mutual information for conditional Gaussians (MI_{cg}) for mixed discrete and normal variables, as implemented in the R package `bnlearn` [33], using a significance level $\alpha = 0.01$ (Bonferroni corrected).

Since conditional independence testing is a difficult statistical problem [34], we validated our results using two alternative methods: a kernel conditional independence test (KCI; Ref. [35]) as implemented in the R package `CondIndTests`, and a test based on the generalized covariance measure (GCM; Ref. [34]), as implemented in the R package `GeneralisedCovarianceMeasure`. These additional tests of conditional independence were not part of our pre-specified analysis. We decided to conduct these additional tests to evaluate the robustness of our results across different methods of conditional independence testing (i.e., a sensitivity analysis). We used the same significance level $\alpha = 0.01$ for the KCI- and the GCM-based tests to ensure compatibility with the pre-specified tests.

2.3.2. Estimating the Average Causal Effect from M to F

ASE theory predicts that fatigue is a feeling state that is triggered by a metacognitive diagnosis of loss of control over bodily states. We aimed to test this prediction as part of our Hypothesis 2: There is a negative average causal effect from metacognition of allostatic control (M) to fatigue (F)

$$\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F] = \theta_3. \quad (7)$$

Adjusting for covariates is one of the various methods for estimating causal effects from observational data. Suppose we are interested in finding the effect of M on F and assume the factors deemed relevant to the problem are structured as in Figure 1. In other words, we are interested in calculating the intervention distribution $\mathbb{P}_{do(M:=m)}(f)$. Given a valid adjustment set (VAS) \mathbf{Z} , here e.g., $\mathbf{Z} = (A, G)$, the intervention distribution can be calculated (see [36–38]) as $\mathbb{P}_{do(M:=m)}(f) = \sum_{\mathbf{z}} \mathbb{P}(f | m, \mathbf{z}) \mathbb{P}(\mathbf{z})$, since

$$\mathbb{P}_{do(M:=m)}(f) = \sum_{\mathbf{z}} \mathbb{P}_{do(M:=m)}(f, m, \mathbf{z}) \quad (8)$$

$$= \sum_{\mathbf{z}} \mathbb{P}_{do(M:=m)}(f | m, \mathbf{z}) \mathbb{P}_{do(M:=m)}(m, \mathbf{z}) \quad (9)$$

$$= \sum_{\mathbf{z}} \mathbb{P}_{do(M:=m)}(f | m, \mathbf{z}) \mathbb{P}_{do(M:=m)}(\mathbf{z}) \quad (10)$$

$$= \sum_{\mathbf{z}} \mathbb{P}(f | m, \mathbf{z}) \mathbb{P}(\mathbf{z}) \quad (11)$$

where, in the last step, one can use the fact that causal relationships are autonomous under interventions (this property is sometimes referred to as ‘autonomy’) [28].

In linear Gaussian systems, a causal effect from M to F can be approximated by $\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F]$ (see e.g., [28]). Assuming that \mathbf{Z} is a VAS for $\{M, F\}$ and $\{M, F\}, \mathbf{Z}$ follow a Gaussian distribution, then the conditional $F | M = m, \mathbf{Z} = \mathbf{z}$ follows a Gaussian distribution as well. Hence, the mean of the distribution is given by

$$\mathbb{E}[F | M = m, \mathbf{Z} = \mathbf{z}] = \theta_3 m + \mathbf{b}^t \mathbf{z} \quad (12)$$

for some θ_3 and \mathbf{b} . Equation (7) then follows from Equation (11).

One can estimate the conditional mean (Equation (12)) by regressing F on M and \mathbf{Z} and subsequently reading off the regression coefficients for M . Alternatively, more sophisticated techniques for estimation of the average causal effect can be used, such as the propensity score method [39] and double/debiased machine learning (DML; Ref. [40]). In Appendix B, the two methods are described in more detail.

As pre-specified in our analysis plan, we conducted linear regression in combination with a one-sided t-test on the regression coefficient of M to evaluate Hypothesis 2. We compared our estimate of the causal effect from M to F obtained via linear regression with the results obtained from using more sophisticated estimation techniques, i.e., the propensity score method [39] and DML [40], following our pre-registered analysis plan.

2.3.3. Estimating the Average Causal Effect from F^*S on D

Another prediction of ASE theory is that fatigue, in combination with a generalization of low self-efficacy beliefs beyond bodily control, induces depression. We formalized this prediction as part of our Hypothesis 3: There is a negative average causal effect of the interaction term between fatigue and general self-efficacy (F^*S) on depression (D)

$$\frac{\partial}{\partial f \partial s} \mathbb{E}_{do(F:=f, S:=s)}[D] = \theta_{10}. \quad (13)$$

Evaluation of Hypothesis 3 followed the same line of reasoning as for Hypothesis 2. We used linear regression in combination with a one-sided t-test on the regression

coefficient of F^*S . Subsequently, we compared the resulting estimate to the results obtained using the propensity score method and DML.

3. Results

3.1. Raw Data

Figure 2 shows a scatter plot matrix of the raw data. Displayed are the measures for all variables A, G, M, F, S, D used in the analysis.

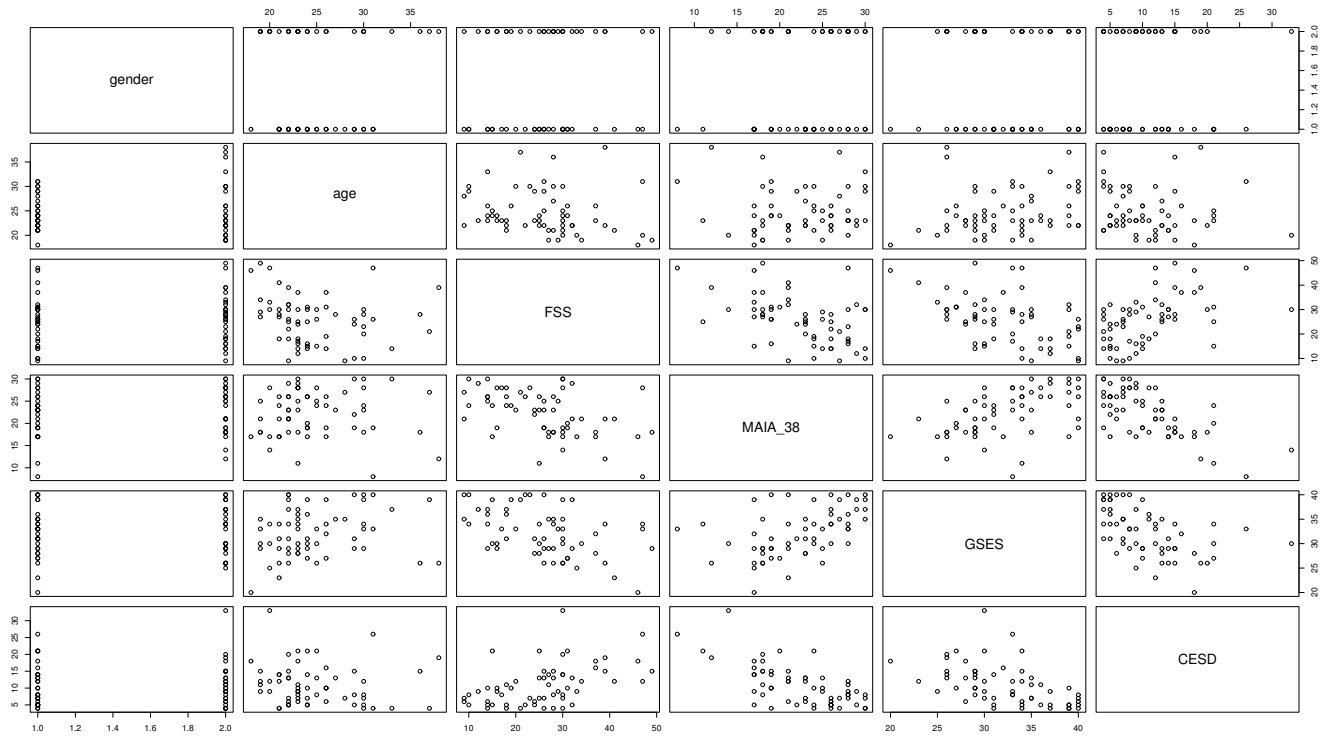


Figure 2. Scatter plot matrix of raw data used in the analysis. Displayed are all the pairwise scatter plots of the variables used for the analysis in a matrix format. For example, the scatter plot located at the intersection of row 3 and column 2 is a plot of the variables age versus fatigue (as measured by the FSS). The variables displayed are gender, age, fatigue (assessed by the FSS), metacognition of allostatic control (assessed by the MAIA_{3,8}), self-efficacy (assessed by the GSES), and depression (assessed by the CES-D).

3.2. Results from the Statistical Analysis

3.2.1. Causal Structure of ASE Theory in the PBIHB Dataset

Table 1 displays the results from conditional independence testing to evaluate the three predictions formulated as part of Hypothesis 1. The results can be summarized as follows:

Table 1. Results from different conditional independence test methods (MI_{cg} , GCM, KCI) for the three predictions formulated as part of Hypothesis 1. Results are presented for three different test methods. An asterisk indicates statistically significant evidence against the null hypothesis (H_0 : variables are conditionally independent) using the pre-specified level $\alpha = 0.01$, which corresponds to a threshold of $p < 0.05$ Bonferroni corrected for the multiple comparisons of the five tests, p -values are shown in parentheses.

Hypothesis 1	d -Separation Statement	MI_{cg} (p -Value)	GCM (p -Value)	KCI (p -Value)
(i)	$M \perp\!\!\!\perp_{J_0} S \mid A, G$	22.044* (1.634×10^{-5})	4.254* (2.104×10^{-5})	26.451* (5.194×10^{-6})

Table 1. Cont.

Hypothesis 1	<i>d</i> -Separation Statement	MI _{cg} (<i>p</i> -Value)	GCM (<i>p</i> -Value)	KCI (<i>p</i> -Value)
(ii)	$M \perp_{J_0} D \mid F, A, G$	24.167* (5.652×10^{-6})	−3.131* (0.001743)	8.513* (0.001346)
	$M \perp_{J_0} D \mid F, A, G, S$	16.883* (0.000216)	−2.574 (0.010064)	2.992 (0.022626)
(iii)	$F \perp_{J_0} S \mid A, G$	13.010* (0.001496)	−3.390* (0.000700)	13.613* (0.001279)
	$F \perp_{J_0} S \mid A, G, M$	4.057 (0.131500)	−2.088 (0.036799)	2.013 (0.118908)

* *p* < 0.01.

(i) $M \not\perp S \mid A, G$. We found significant evidence that metacognition of allostatic control (*M*) and general self-efficacy (*S*) are not independent conditional on age (*A*) and gender (*G*) across all three different conditional independence test methods. In other words, we found a contradiction between *d*-separation within the DAG J_0 and conditional independence of *M* and *S* given *A, G*.

(ii) $M \not\perp D \mid F, A, G$ and $M \perp D \mid F, A, G, S$. We found significant evidence that *M* and depression (*D*) are not independent conditional on fatigue (*F*), *A, G* across all three methods for conditional independence testing. This result was consistent with our findings for (i), in the sense that if we add a directed edge from *M* to *S* in the DAG J_0 (Figure 1), the only set of variables that *d*-separates *M* and *D* is the set *F, A, G, S* (and not *F, A, G*). However, the results for conditional independence tests of *M* and *D* conditional on *F, A, G, S* were mixed, with 2 out of 3 tests (GCM and KCI) not reaching the pre-specified significance level $\alpha = 0.01$. Hence, further evidence is needed to draw conclusions regarding the statement $M \perp D \mid F, A, G, S$.

(iii) $F \not\perp S \mid A, G$ and $F \perp S \mid A, G, M$. When looking at the conditional independence between *F* and *S*, the results depended on the set of variables that we conditioned on. We found significant evidence that *F* and *S* are not independent conditional on *A, G* across all three different test methods. However, we failed to reject the null hypothesis that *F* and *S* are independent conditional on the set *M, A, G* consistently across all three different test methods. This result is also in line with our findings for (i), in the sense that if we add a directed edge from *M* to *S* in the DAG J_0 (Figure 1), the only set of variables that *d*-separates *F* and *S* is the set *M, A, G*.

3.2.2. Estimating the Average Causal Effect from *M* to *F*

As predicted by the ASE theory, we found significant evidence for a negative average causal effect from metacognition of allostatic control (*M*) to fatigue (*F*) $\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F] = \theta_3$ across all three different estimation methods. The resulting estimates $\hat{\theta}_3$ for the VAS $\mathbf{Z} = (A, G)$ are displayed in Table 2, alongside lower and upper bounds of a 95% confidence interval for $\hat{\theta}_3$, the corresponding value of the *t*-statistic, as well as the *p*-value for the one-sided *t*-test.

Table 2. Average causal effect from *M* to *F* using $\mathbf{Z} = (A, G)$. Displayed are estimates of the average causal effect from *M* to *F* $\hat{\theta}_3$ across three different methods to adjust for the covariates $\mathbf{Z} = (A, G)$. We report a point estimate $\hat{\theta}_3$, the lower and upper bounds of a 95% confidence interval for $\hat{\theta}_3$, the value of the *t*-statistic, as well as the *p*-value for the one-sided *t*-test. An asterisk indicates statistical significance using the pre-specified level $\alpha = 0.017$ (Bonferroni-corrected).

Estimation Method	$\hat{\theta}_3$	Confidence Interval		<i>t</i> Value	<i>p</i> -Value
linear regression	−0.4845*	−0.712	−0.257	−4.259	3.968×10^{-5}
propensity score	−0.4816*	−0.717	−0.246	−4.092	6.689×10^{-5}
DML	−0.3872*	−0.6481	−0.1262	−2.9082	0.0018

* *p* < 0.017.

The results from our sensitivity analysis, i.e., estimating θ_3 using a different VAS $\mathbf{Z} = (A, G, S)$, are listed in Table 3. They confirmed the finding of a negative average causal effect from M to F when using $\mathbf{Z} = (A, G)$ as a VAS. The main difference between the results of the two analyses was that the second analysis using $\mathbf{Z} = (A, G, S)$ yielded a slightly lower absolute value for $\hat{\theta}_3$, as well as a non-significant p -value using the DML method.

Table 3. Average causal effect from M to F using $\mathbf{Z} = (A, G, S)$. Displayed are estimates of the average causal effect from M to F $\hat{\theta}_3$ across three different methods to adjust for the covariates $\mathbf{Z} = (A, G, S)$. We report a point estimate of $\hat{\theta}_3$, the lower and upper bounds of a 95% confidence interval for $\hat{\theta}_3$, the value of the t -statistic, as well as the p -value for the one-sided t -test. An asterisk indicates statistical significance using the pre-specified level $\alpha = 0.017$ (Bonferroni-corrected).

Estimation Method	$\hat{\theta}_3$	Confidence Interval		t Value	p -Value
linear regression	−0.3545*	−0.610	−0.099	−2.785	0.0037
propensity score	−0.3775*	−0.692	−0.063	−2.400	0.0098
DML	−0.2049	−0.563	0.153	−1.122	0.1309

* $p < 0.017$.

3.2.3. Estimating the Average Causal Effect from F^*S to D

We did not find evidence for the predicted negative average causal effect of the interaction term between fatigue and general self-efficacy (F^*S) on depression (D) $\frac{\partial}{\partial f \partial s} \mathbb{E}_{do(F:=f, S:=s)}[D] = \theta_{10}$ across all three different estimation methods for either VAS $\mathbf{Z} = (A, G)$ or $\mathbf{Z} = (A, G, M)$. Tables containing the resulting estimates for $\hat{\theta}_{10}$ including a 95% confidence interval and the value of the t -statistic, as well as the p -value for the one-sided t -test, are listed in Appendix C.

4. Discussion

In this paper, we proposed a formulation of the allostatic self-efficacy (ASE) theory of fatigue and depression in the language of causal inference. Specifically, we identified the variables of central interest to the ASE theory and formulated a structural causal model (SCM) under assumptions of linearity and normality. The SCM, as well as the induced directed acyclic graph (DAG), describe the direction of causality among these variables. The proposed SCM only contains variables and relations that were explicitly suggested by the ASE theory (together with age and gender as potential confounders) in its current formulation [7]. This is not meant to imply that other variables or connections cannot be present or may not be included in future revisions of the theory. However, in this preregistered analysis, the focus was entirely on the ASE theory in its current form. Using the data of 60 healthy individuals from a previous study on interoception of breathing and its relation with several psychopathological constructs [27], we tested the proposed causal model empirically. Relying on the assumption of the Markov condition, we used the dataset to search for contradictions to conditional independence statements (Hypothesis 1) that are implied by the graph structure (d -separation). In a second and third step, we estimated the value of two causal effects that are predicted by the ASE theory using methods of covariate adjustment, propensity scores, and double/debiased machine learning. As predicted by the ASE theory, we found a statistically significant negative average causal effect from metacognition of allostatic control (M) to fatigue (F) $\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F] = \theta_3$ across all three methods of estimation. Our sensitivity analysis using a different valid adjustment set largely confirmed this finding with two out of three estimation methods yielding a significant result.

The assumption of the Markov condition establishes a connection from d -separation statements in a causal graph to conditional independence statements in the distribution. In the analysis of Hypothesis 1, we tested concrete predictions implied by the DAG J_0

(Figure 1). (i) Using the data from the PBIHB study, we were able to reject the null hypothesis of $M \perp\!\!\!\perp S \mid A, G$ at the pre-specified level $\alpha = 0.01$. (ii) We found significant evidence against $M \perp\!\!\!\perp D \mid F, A, G$ in the empirical dataset. However, in line with the graph structure J_0 implied by the ASE theory, we did not find clear evidence against $M \perp\!\!\!\perp D \mid F, A, G, S$. That is, only one out of three conditional independence tests rejected the null hypothesis of metacognition of allostatic control (M) being independent from depression (D) conditional on the set F, A, G, S . (iii) We also found significant evidence against $F \perp\!\!\!\perp S \mid A, G$ in the empirical data. Yet, we did not find any evidence against $F \perp\!\!\!\perp S \mid A, G, M$. All three conditional independence test methods consistently failed to reject the null hypothesis of fatigue (F) and general self-efficacy (S) being independent given the set A, G, D, M .

There are a number of potential explanations for the results related to Hypothesis 1. The most straightforward explanation is that the proposed causal model is incorrect. This could include the presence of additional edges between nodes, as well as variables that were not considered, acting as mediators or confounds or a combination of all of the aforementioned. For example, although ASE theory does not make an explicit statement about a direct link between metacognition of allostatic control (M) and general self-efficacy (S), it is plausible to assume the existence of a directed edge from M (the feeling of control over bodily states) to S (an individual's general expectation of personal mastery and control [41]). The construct of S is closely related to concepts of metacognition (see, e.g., [42]) and represents a 'global' construct of self-beliefs about one's capacity to achieve goals and overcome adversity; this can be understood as including more 'local' domain-specific forms of self-efficacy, such as metacognition of allostatic control. From this viewpoint, the idea that metacognition of allostatic control (M) may contribute to (and thus influence) general self-efficacy (S) beliefs is therefore not entirely unreasonable and would be a potential explanation for the results of (i) and (iii). More precisely, a directed edge from M to S would render M and S d -connected, since there would always exist a path between M and S that is not blocked by any set of variables. This cause–effect structure would explain the observed dependence between the two variables in the empirical dataset according to Reichenbach's common cause principle [31]. Another consequence of introducing an edge from M to S would be that the set of variables that d -separates F and S would consist of variables A, G, M and not only A, G , which corresponds to our findings for (iii). The same is true for the set of variables d -separating M and D , which would consist of variables F, A, G, S and not F, A, G in this case, potentially explaining our findings for (ii). However, since the evidence for $M \perp\!\!\!\perp D \mid F, A, G, S$ was mixed, further research is needed to bring clarity to the question of the (conditional) independence of M and D .

The revised DAG J_1 (Figure 3) provides a graphical summary of the above considerations regarding the results related to Hypothesis 1. From DAG J_0 to J_1 , we added a directed edge from M to S . However, there are several other potential explanations for the observed results, so this example should by no means be taken as 'the correct model'. If anything, this should be regarded as an updated hypothesis to be tested in future investigations.

One of our reviewers wondered why we did not compare different graph structures against each other by evaluating how well they fit the empirical data. In causal inference, this approach would belong to the class of score-based methods for causal discovery [28]. This is certainly an interesting alternative strategy, provided a suitable generative model exists, and may well be adopted in future work on this topic. By contrast, the approach in this work belongs to the category of independence-based methods for causal structure learning. It consists of specifying a theory in the form of an SCM and subsequently testing which parts of the associated DAG J_0 are inconsistent with empirical data using conditional independence tests, which subsequently leads to an amended DAG J_1 . This was particularly well suited for our problem, since the DAG J_0 was based only on the variables and relations explicitly suggested by ASE theory [7] that we wanted to examine and refine.

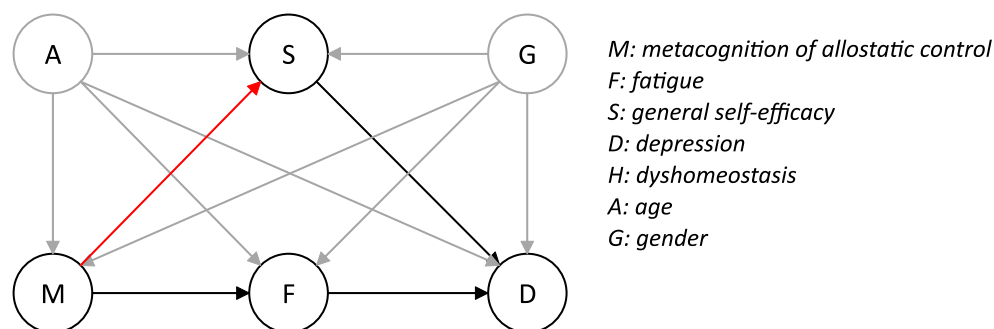
DAG J_1 

Figure 3. Updated directed acyclic graph (DAG) J_1 of the allostatic self-efficacy theory (ASE; Ref. [7]) providing a potential explanation for the observed results from analysis of Hypothesis 1. Modifications from DAG J_0 to J_1 are shown in red.

Concerning Hypothesis 2, we found evidence for a negative average causal effect from metacognition of allostatic control (M) to fatigue (F) $\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F] = \theta_3$ across all three estimation methods (covariate adjustment, propensity scores, DML) for two different VAS. This is in line with the prediction by ASE theory that the subjective experience of fatigue arises as a consequence of a metacognitive diagnosis that the brain's control over bodily states is failing (low allostatic control). This also confirms findings from previous research, which identified metacognition of allostatic control (M) (operationalized by the sum of the subscales 3 and 8 of the MAIA questionnaire) as associated with fatigue (F) scores [19]. Our new results go beyond this previous finding, in the sense that the current study suggests a direction of the effect, as opposed to purely associative statements. It is worth highlighting that the estimation of the causal effect from M to F would not be affected by the proposed additional link between M to S as suggested by the analysis results concerning Hypothesis 1 (see Figure 3), since the set A, G would still be a valid adjustment set (VAS).

With regard to Hypothesis 3, we did not find evidence for a negative average causal effect of the interaction term between fatigue and general self-efficacy ($F*S$) on depression (D) $\frac{\partial}{\partial f \partial s} \mathbb{E}_{do(F:=f, S:=s)}[D] = \theta_{10}$. The present work is, to the best of our knowledge, the first attempt to investigate the predicted influence of the interaction between fatigue and general self-efficacy on depression. Across all three different estimation methods and using different VAS, we found, if anything, very small effects. However, one may rightfully question whether the sample in this study was adequate for testing Hypothesis 3, at least in the context of the ASE theory. This is because our participants were drawn from the general population and, not surprisingly, did not show pronounced levels of depression (compare Figure 2). By contrast, the predictions of ASE theory concerning depression assume a clinically relevant state of depression [7]. Therefore, the potential interaction effect $F*S$ on D remains an open question that should be addressed in the future using samples with clinically relevant levels of depression. In the context of depression, it is worth mentioning that the current formulation of the ASE theory does not make any explicit reference to emotional states. This could potentially be addressed in future developments by extending the generic Bayesian model of brain–body interactions that represents the foundation of the ASE theory [7] with developments from active inference which introduce explicit representations of valence (e.g., [43]).

One might ask why we did not use a more complex causal model and examine any potential bidirectional causal effects between the variables of interest, e.g., fatigue and depression. Indeed, the lack of bidirectional causal influences in our model can be regarded as a limitation, in that it prevents identification of an optimal causal structure from data without constraints. However, the objective of the current study was to examine whether the causal structure implied by a particular theory (i.e., ASE) is compatible with available cross-sectional questionnaire data. Notably, the complexity of the proposed SCM was commensurate with both the level of complexity of the ASE theory and the

cross-sectional dataset, i.e., our model only contains those variables and relations that are explicitly suggested by the ASE theory in its current formulation [7] and were measured using available questionnaires. Nevertheless, should future formulations of the ASE theory include bidirectional effects, the current methodological approach (i.e., conditional independence tests in the context of DAGs) could be extended. Specifically, if longitudinal timeseries data are available, DAGs can be extended in time, which allows testing cyclic and bidirectional causal effects (for examples, see [44,45]), thus addressing the limitation discussed above. Furthermore, if required by the theory to be tested, longitudinal data would also enable one to consider the role of time. While not required for the purpose of the current study (since the current formulation of ASE theory does not make any statements about the temporal dynamics of variables and their relations), a more exploratory approach could examine, for example, whether causal influences operate at particular time scales. Such putative extensions could employ other methods of causal inference, e.g., Granger causality [46] convergent cross mapping [47], or dynamic causal modeling (e.g., [48]), once longitudinal timeseries data for the variables of interest become available.

The current study has strengths and limitations. Its strengths include the first concrete formulation of the ASE theory in the language of causal inference: our proposal of an SCM brings the content of a verbally formulated theory into the realm of concrete mathematical equations. Together with the induced DAG, this provided a formal basis for analysis and allowed us to identify a set of empirically testable hypotheses which may guide future research. A second notable strength is that we used multiple independent methods, both for conditional independence testing (Hypothesis 1) and for the estimation of causal effects (Hypotheses 2 and 3). This enhances the robustness of our conclusions, since they do not depend on assumptions and properties of any single method. Last but not least, all of our hypotheses and statistical analysis procedures were pre-registered and specified in detail in an ex ante analysis plan (<https://doi.org/10.5281/zenodo.10559656> (accessed on 12 December 2024)). Preregistration is an important and effective protection for the robustness of research, given the many degrees of freedom and the numerous cognitive biases that scientists may inadvertently be affected by [49].

Turning to limitations, first, as in any observational study, one may question whether the results may have been affected by (unknown or non-measured) confounding variables (compare our discussion of the results from Hypothesis 1 above). Specifically, one limitation is that our current analysis did not take sleep into account. While sleep is not an explicit component of the ASE theory, previous work has repeatedly demonstrated that sleep quality generally influences the level of fatigue (e.g., [19,50]). In the present study, we did not examine the potential influence of sleep, since the available dataset did not include any measures of sleep quality. This will be rectified in follow-up studies.

Second, our statistical approach made a number of assumptions, as explained in the Methods. Most of these were standard assumptions and not specific to our study, but are still worth keeping in mind, since violation of these assumptions may affect the results. For example, our structural model assumed that effects are linear and random variables (except gender) follow a normal distribution. Furthermore, as already discussed above, our statistical approach did not take into account bidirectional causal interactions. This is because our causal model is constrained by the explicit assumptions of the ASE theory (which does not include bidirectional causal effects). Finally, conditional independence testing, which is at the heart of causal discovery [38], is generally challenging [51]; we, therefore, employed multiple independent tests to ensure our results were robust.

Finally, our dataset has several properties that deserve consideration. To begin with, our dataset is cross-sectional and observational, which is the norm for studies in neuroscience but challenging for causal inference. To facilitate causal inference, follow-up studies could include interventions (e.g., cognitive interventions targeting M) and/or acquire longitudinal data. Another issue is the question of sample size. While the size of our sample ($N = 60$) was large for the neuroscientific focus of the original study, it was not determined in advance for the questions addressed in the current study. Furthermore, the

question of what constitutes adequate sample sizes for causal inference from observational data is complex and controversial [52]. It is therefore important to replicate our findings in independent (and larger) samples. A third issue concerns the validity of the available measurements. While validated self-report measures for fatigue, depression, and general self-efficacy are available and were used in this study, a measurement tool specifically developed for the construct of metacognition of allostatic control (M) is lacking so far. Here, we followed previous research [19] and used a plausible proxy measure of M , i.e., the sum of two pre-specified subscales from a validated questionnaire on interoception, MAIA. An important goal for future research is the development of readouts for M that are fully validated and easily applied in practice. Ideally, such tests would go beyond self-report measures and be implicit in nature, e.g., neurophysiological readouts similar to the mismatch negativity.

5. Conclusions

In summary, our work provides a formal basis for testing predictions by the ASE theory of fatigue and depression in the context of causal inference. We evaluated central aspects of our proposed SCM using a publicly available dataset and provided an updated version of the SCM that accounts for our empirical findings. In addition, we were able to confirm previous findings regarding the association between metacognition of allostatic control (M) and fatigue (F). Our analysis enabled us to quantify the direction, as well as the sign, of the causal effect, i.e., we found a negative average causal effect from M to F $\frac{\partial}{\partial m} \mathbb{E}_{do(M:=m)}[F] = \theta_3$, as predicted by the ASE theory. Finally, we identified a number of open questions that remain to be addressed in future research and that may help unravel the mechanisms behind fatigue and depression.

Author Contributions: Conceptualization, A.J.H., D.v.W., J.H., and K.E.S.; methodology, A.J.H., J.H., and K.E.S.; software, A.J.H. and D.v.W.; validation, D.v.W.; formal analysis, A.J.H.; investigation, A.J.H. and O.K.H.; resources, O.K.H. and K.E.S.; data curation, O.K.H.; writing—original draft preparation, A.J.H.; writing—review and editing, A.J.H., D.v.W., O.K.H., J.H., and K.E.S.; visualization, A.J.H.; supervision, K.E.S.; project administration, A.J.H.; funding acquisition, O.K.H. and K.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the René and Susanne Braginsky Foundation, the ETH Foundation and the University of Zurich. O.K.H. (née Faull) was supported by a Marie Skłodowska-Curie Postdoctoral Fellowship from the European Union’s Horizon 2020 research and innovation program under the grant agreement 793580, and a Rutherford Discovery Fellowship from the Royal Society Te Apārangi.

Institutional Review Board Statement: Ethical review and approval were not required for this project according to the Swiss Human Research Ordinance, because it comprised a reanalysis of a publicly available anonymized dataset.

Data Availability Statement: Data used in this study are publicly available for download from the Zenodo open data repository <https://doi.org/10.5281/zenodo.10992529> (accessed on 12 December 2024). The statistical analysis plan detailing the analysis ex ante is available at <https://doi.org/10.5281/zenodo.10559656> (accessed on 12 December 2024). The analysis code is available at <https://github.com/alexjhess/pbihb-ase-causality> (accessed on 12 December 2024).

Acknowledgments: We wish to thank Jonas Peters for helpful discussions. And, most importantly, Happy Birthday, Karl!

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
ASE	allostatic self-efficacy
SCM	structural causal model

DAG	directed acyclic graph
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th revision
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
A	age
G	gender
M	metacognition of allostatic control
F	fatigue
S	general self-efficacy
D	depression
N	noise variable
F*S	the interaction term between fatigue and general self-efficacy
PBIHB	perception of breathing in the human brain
TNU	Translational Neuromodeling Unit
FSS	Fatigue Severity Scale
GSES	General Self-Efficacy Scale
CES-D	Centre for Epidemiologic Studies Depression Scale
MAIA	Multidimensional Assessment of Interoceptive Awareness
MI_{CG}	mutual information for conditional Gaussians
KCI	kernel conditional independence test
GCM	generalized covariance measure
DML	double/debiased machine learning
VAS	valid adjustment set

Appendix A. Definitions

Appendix A.1. Structural Causal Model

We adopt the definition of SCMs according to [26]:

Definition A1. An SCM over variables $\mathbf{X} = [X_1, \dots, X_n]$ comprises

- structural equations which relate each variable X_k to its parents $\mathbf{PA}(X_k) \subseteq \{X_1, \dots, X_n\}$ and a noise variable N_k via a function f_k , such that $X_k := f_k(\mathbf{PA}(X_k), N_k)$, as well as a
- noise distribution \mathbb{P}_N of the noise variables $\mathbf{N} = [N_1, \dots, N_n]^T$.

In a directed causal graph associated with an SCM, the nodes correspond to the variables X_1, \dots, X_n and there is an edge from X_i to X_j whenever X_i appears on the right-hand side of the equation $X_j := f_j(\mathbf{PA}(X_j), N_j)$. In other words, if $X_i \in \mathbf{PA}(X_j)$ the graph contains the edge $X_i \rightarrow X_j$. For this work, we assume that the graph does not contain any cycles. The structural equations together with the noise distributions induce the observational distribution \mathbb{P}_X of X_1, \dots, X_n as a simultaneous solution to the equations.

Appendix A.2. Markov Condition

Definition A2. Given a DAG G over nodes \mathbf{X} , we say that the distribution \mathbb{P}_X satisfies

- the **global Markov property** (MP) with respect to G if \forall disjoint $A, B, C \subseteq \mathbf{X}$ A d -sep $B \mid C \implies A \perp\!\!\!\perp B \mid C$
- the **local Markov property** (MP) if $\forall j$ $X_j \perp\!\!\!\perp ND_j \mid PA_j$
- the **factorization property** if \mathbb{P}_X is absolutely constant with respect to a product measure and $\forall x \forall j$, $p(x_{PA_j}) > 0 : p(\mathbf{x}) = p(x_1, \dots, x_d) = \prod_j^d p(x_j \mid x_{PA_j})$

In the above definition, we used the following notation: ND_j represents the non-descendants of node X_j and PA_j denotes all nodes that have a directed edge to node X_j .

Appendix A.3. *d*-Separation

Definition A3. *d*-separation is a graphical criterion of whether two nodes are connected or not. Let X, Y, Z disjoint.

- (i) A path $X = i_1, \dots, i_m = Y$ is blocked by $Z \iff \exists$ node i_k with $i_{k-1} \rightarrow i_k \rightarrow i_{k+1}$ and $i_k \in Z$
 OR \exists node i_k with $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$ and $i_k \in Z$
 OR \exists node i_k with $i_{k-1} \leftarrow i_k \rightarrow i_{k+1}$ and $i_k \in Z$
 OR \exists node i_k with $i_{k-1} \rightarrow i_k \leftarrow i_{k+1}$ and $i_k \notin Z$ and $DE(i_k) \cap Z = \emptyset$
- (ii) X, Y are *d*-connected given $Z \iff \exists X \in \mathbf{X}, Y \in \mathbf{Y}$ s.t. \exists path between X and Y that is not blocked
- (iii) if X, Y are not *d*-connected, then they are *d*-separated. We sometimes write X *d*-sep $Y \mid Z$ or $X \perp_G Y \mid Z$

Appendix B. Estimating Causal Effects Using Covariate Adjustment

Appendix B.1. The ‘Propensity Score’ Method

In a point treatment situation, one can adjust for a set of confounders $Z = (A, G)$ when estimating the effect of exposure M by weighting observations i by the inverse probability weights

$$w_i = \frac{1}{P(M_i = m_i \mid Z_i = z_i)} \tag{A1}$$

To increase statistical efficiency, one can use stabilized weights, e.g.,

$$sw_i = \frac{P(M_i = m_i)}{P(M_i = m_i \mid Z_i = z_i)} \tag{A2}$$

When dealing with a continuous exposure variable M , one can use stabilized weights

$$sw_i = \frac{f(m_i)}{f(m_i \mid z_i)} \tag{A3}$$

where $f(m_i)$ is the marginal density function of M , evaluated at the observed value in unit i , m_i , and $f(m_i \mid z_i)$ conditional density function of M given Z , evaluated at the observed values in unit i , $\{m_i, z_i\}$ [53]. Weighting observations i by sw_i , one can fit a causal model, for instance a marginal structural model (MSM)

$$\mathbb{E}[F_m] = \beta_0 + \theta_3 m \tag{A4}$$

with continuous outcome fatigue F . The response variable F_m is the potential outcome that could have been observed in a unit under study if that unit received a specific treatment level m [54]. The expectation $\mathbb{E}[F_m]$ is the mean outcome if all units under study received a specific treatment level m . Parameter θ_3 then quantifies the causal effect of M on F [53].

Appendix B.2. Double/Debiased Machine Learning

DML removes the impact of regularization bias and overfitting on estimation of the parameter of interest θ_3 by using Neyman-orthogonal moments and cross-fitting [40]. One application of DML is in the context of a partial linear regression model,

$$F = M\theta_3 + k_0(Z) + N_f, \quad \mathbb{E}(N_f \mid M, Z) = 0, \tag{A5}$$

$$M = m_0(Z) + V, \quad \mathbb{E}(V \mid Z) = 0, \tag{A6}$$

with fatigue F , metacognition of allostatic control M , a VAS $Z = (A, G)$ consisting of confounding covariates and stochastic error terms N_f and V . The confounding covariates Z affect M and F via the functions m_0 and k_0 , respectively. DML can be used to estimate θ_3 ,

i.e., the main regression coefficient that we would like to infer, which can be interpreted as the average causal effect from M to F [55].

Appendix C. Results from Estimating the Average Causal Effect from F^*S to D

Table A1. Average causal effect of the interaction term F^*S to D using $\mathbf{Z} = (A, G)$. Displayed are estimates of the average causal effect of the interaction term F^*S to D $\hat{\theta}_{10}$ across three different methods to adjust for the covariates $\mathbf{Z} = (A, G)$. We report a point estimate of $\hat{\theta}_{10}$, the lower and upper bounds of a 95% confidence interval for $\hat{\theta}_{10}$, the value of the t -statistic, as well as the p -value for the one-sided t -test.

Estimation Method	$\hat{\theta}_{10}$	Confidence Interval		t Value	p -Value
linear regression	0.0281	−0.191	0.247	0.257	0.6010
propensity score	0.0142	−0.151	0.180	0.172	0.5680
DML	−0.2051	−0.476	0.066	−1.482	0.0691

Table A2. Average causal effect of the interaction term F^*S to D using $\mathbf{Z} = (A, G, M)$. Displayed are estimates of the average causal effect of the interaction term F^*S to D $\hat{\theta}_{10}$ across three different methods to adjust for the covariates $\mathbf{Z} = (A, G, M)$. We report a point estimate of $\hat{\theta}_{10}$, the lower and upper bounds of a 95% confidence interval for $\hat{\theta}_{10}$, the value of the t -statistic, as well as the p -value for the one-sided t -test.

Estimation Method	$\hat{\theta}_{10}$	Confidence Interval		t Value	p -Value
linear regression	0.0671	−0.122	0.257	0.711	0.7599
propensity score	0.0337	−0.159	0.227	0.350	0.6362
DML	0.0153	−0.283	0.314	0.100	0.5399

References

- Chaudhuri, A.; Behan, P.O. Fatigue in neurological disorders. *Lancet* **2004**, *363*, 978–988. [[CrossRef](#)] [[PubMed](#)]
- Wessely, S. Chronic Fatigue: Symptom and Syndrome. *Ann. Intern. Med.* **2001**, *134*, 838–843. [[CrossRef](#)] [[PubMed](#)]
- Fisk, J.D.; Ritvo, P.G.; Ross, L.; Haase, D.A.; Marrie, T.J.; Schlech, W.F. Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. *Clin. Infect. Dis.* **1994**, *18*, S79–S83. [[CrossRef](#)]
- World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*; World Health Organization: Geneva, Switzerland, 2004.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Washington, DC, USA, 2013.
- Manjaly, Z.M.; Harrison, N.A.; Critchley, H.D.; Do, C.T.; Stefanics, G.; Wenderoth, N.; Lutterotti, A.; Müller, A.; Stephan, K.E. Pathophysiological and cognitive mechanisms of fatigue in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **2019**, *90*, 642–651. [[CrossRef](#)] [[PubMed](#)]
- Stephan, K.E.; Manjaly, Z.M.; Mathys, C.D.; Weber, L.A.; Paliwal, S.; Gard, T.; Tittgemeyer, M.; Fleming, S.M.; Haker, H.; Seth, A.K.; et al. Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Front. Hum. Neurosci.* **2016**, *10*, 550. [[CrossRef](#)] [[PubMed](#)]
- Petzschner, F.H.; Weber, L.A.; Gard, T.; Stephan, K.E. Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. *Biol. Psychiatry* **2017**, *82*, 421–430. [[CrossRef](#)]
- Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. Biol. Sci.* **2005**, *360*, 815–836. [[CrossRef](#)]
- Friston, K.J.; Daunizeau, J.; Kilner, J.; Kiebel, S.J. Action and behavior: A free-energy formulation. *Biol. Cybern.* **2010**, *102*, 227–260. [[CrossRef](#)] [[PubMed](#)]
- Nord, C.L.; Garfinkel, S.N. Interoceptive pathways to understand and treat mental health conditions. *Trends Cogn. Sci.* **2022**, *26*, 499–513. [[CrossRef](#)] [[PubMed](#)]
- Khalsa, S.S.; Adolphs, R.; Cameron, O.G.; Critchley, H.D.; Davenport, P.W.; Feinstein, J.S.; Feusner, J.D.; Garfinkel, S.N.; Lane, R.D.; Mehling, W.E.; et al. Interoception and Mental Health: A Roadmap. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2018**, *3*, 501–513. [[CrossRef](#)] [[PubMed](#)]

13. Seth, A.K.; Suzuki, K.; Critchley, H.D. An interoceptive predictive coding model of conscious presence. *Front. Psychol.* **2012**, *3*, 395. [[CrossRef](#)] [[PubMed](#)]
14. Seth, A.K. Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **2013**, *17*, 565–573. [[CrossRef](#)]
15. Pezzulo, G.; Rigoli, F.; Friston, K. Active Inference, homeostatic regulation and adaptive behavioural control. *Prog. Neurobiol.* **2015**, *134*, 17–35. [[CrossRef](#)]
16. Ainley, V.; Apps, M.A.J.; Fotopoulou, A.; Tsakiris, M. ‘Bodily precision’: A predictive coding account of individual differences in interoceptive accuracy. *Philos. Trans. R. Soc. Biol. Sci.* **2016**, *371*, 20160003. [[CrossRef](#)] [[PubMed](#)]
17. Toussaint, B.; Heinzle, J.; Stephan, K.E. A computationally informed distinction of interoception and exteroception. *Neurosci. Biobehav. Rev.* **2024**, *159*, 105608. [[CrossRef](#)]
18. Fleming, S.M.; Dolan, R.J. The neural basis of metacognitive ability. *Philos. Trans. R. Soc. Biol. Sci.* **2012**, *367*, 1338–1349. [[CrossRef](#)]
19. Rouault, M.; Pereira, I.; Galioulline, H.; Fleming, S.M.; Stephan, K.E.; Manjaly, Z.M. Interoceptive and metacognitive facets of fatigue in multiple sclerosis. *Eur. J. Neurosci.* **2023**, *58*, 2603–2622. [[CrossRef](#)]
20. Dantzer, R.; O’Connor, J.C.; Freund, G.G.; Johnson, R.W.; Kelley, K.W. From inflammation to sickness and depression: When the immune system subjugates the brain. *Nat. Rev. Neurosci.* **2008**, *9*, 46–56. [[CrossRef](#)] [[PubMed](#)]
21. Dantzer, R.; Heijnen, C.J.; Kavelaars, A.; Laye, S.; Capuron, L. The neuroimmune basis of fatigue. *Trends Neurosci.* **2014**, *37*, 39–46. [[CrossRef](#)] [[PubMed](#)]
22. Heitmann, H.; Andlauer, T.F.M.; Korn, T.; Mühlau, M.; Henningsen, P.; Hemmer, B.; Ploner, M. Fatigue, depression, and pain in multiple sclerosis: How neuroinflammation translates into dysfunctional reward processing and anhedonic symptoms. *Mult. Scler. J.* **2022**, *28*, 1020–1027. [[CrossRef](#)] [[PubMed](#)]
23. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: Oxford, UK, 1989; 514p. [[CrossRef](#)]
24. Pearl, J. Causal Diagrams for Empirical Research. *Biometrika* **1995**, *82*, 669–688. [[CrossRef](#)]
25. Pearl, J. *Causality*; Cambridge University Press: Cambridge, UK, 2009.
26. Weichwald, S.; Peters, J. Causality in Cognitive Neuroscience: Concepts, Challenges, and Distributional Robustness. *J. Cogn. Neurosci.* **2021**, *33*, 226–247. [[CrossRef](#)]
27. Harrison, O.K.; Köchli, L.; Marino, S.; Luechinger, R.; Hennel, F.; Brand, K.; Hess, A.J.; Frässle, S.; Iglesias, S.; Vinckier, F.; et al. Interoception of breathing and its relationship with anxiety. *Neuron* **2021**, *109*, 4080–4093. [[CrossRef](#)] [[PubMed](#)]
28. Peters, J.; Janzing, D.; Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms*; The MIT Press: Cambridge, MA, USA, 2017.
29. Bongers, S.; Forré, P.; Peters, J.; Mooij, J.M. Foundations of structural causal models with cycles and latent variables. *Ann. Stat.* **2021**, *49*, 2885–2915. [[CrossRef](#)]
30. Lauritzen, S.L.; Lauritzen, S.L. *Graphical Models*; Oxford Statistical Science Series; Oxford University Press: Oxford, UK, 1996.
31. Reichenbach, H. *The Direction of Time*; Dover Publications: Mineola, NY, USA, 1956.
32. Lauritzen, S.L.; Dawid, A.P.; Larsen, B.N.; Leimer, H.G. Independence properties of directed markov fields. *Networks* **1990**, *20*, 491–505. [[CrossRef](#)]
33. Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *J. Stat. Softw.* **2010**, *35*, 1–22. [[CrossRef](#)]
34. Shah, R.D.; Peters, J. The hardness of conditional independence testing and the generalised covariance measure. *Ann. Stat.* **2020**, *48*, 1514–1538. [[CrossRef](#)]
35. Zhang, K.; Peters, J.; Janzing, D.; Schoelkopf, B. Kernel-based Conditional Independence Test and Application in Causal Discovery. *arXiv* **2012**, arXiv:1202.3775. [[CrossRef](#)]
36. Pearl, J. Belief networks revisited. *Artif. Intell.* **1993**, *59*, 49–56. [[CrossRef](#)]
37. Robins, J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **1986**, *7*, 1393–1512. [[CrossRef](#)]
38. Spirtes, P.; Glymour, C.N.; Scheines, R. *Causation, Prediction, and Search*; MIT Press: Cambridge, MA, USA, 2000.
39. Rosenbaum, P.R.; Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **1983**, *70*, 41–55. [[CrossRef](#)]
40. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **2018**, *21*, C1–C68. [[CrossRef](#)]
41. Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* **1977**, *84*, 191–215. [[CrossRef](#)] [[PubMed](#)]
42. Clark, I.; Dumas, G. The regulation of task performance: A trans-disciplinary review. *Front. Psychol.* **2016**, *6*, 1862. [[CrossRef](#)]
43. Hesp, C.; Tschantz, A.; Millidge, B.; Ramstead, M.; Friston, K.; Smith, R. Sophisticated Affective Inference: Simulating Anticipatory Affective Dynamics of Imagining Future Events. In *Active Inference: First International Workshop, IWAI 2020, Co-Located with ECML/PKDD 2020, Ghent, Belgium, September 14, 2020, Proceedings 1*; Verbelen, T., Lanillos, P., Buckley, C.L., De Boom, C., Eds.; Springer: Cham, Switzerland, 2020; pp. 179–186. [[CrossRef](#)]
44. Murray, E.J.; Kunicki, Z. As the Wheel Turns: Causal Inference for Feedback Loops and Bidirectional Effects. 2022. Available online: <https://osf.io/preprints/osf/9em5q> (accessed on 12 December 2024).
45. Daunizeau, J.; David, O.; Stephan, K.E. Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage* **2011**, *58*, 312–322. [[CrossRef](#)] [[PubMed](#)]
46. Faes, L.; Nollo, G.; Stramaglia, S.; Marinazzo, D. Multiscale Granger causality. *Phys. Rev. E* **2017**, *96*, 042150. [[CrossRef](#)] [[PubMed](#)]

47. Sugihara, G.; May, R.; Ye, H.; Hsieh, C.h.; Deyle, E.; Fogarty, M.; Munch, S. Detecting Causality in Complex Ecosystems. *Science* **2012**, *338*, 496–500. [[CrossRef](#)]
48. Friston, K.; Parr, T.; Zeidman, P.; Razi, A.; Flandin, G.; Daunizeau, J.; Hulme, O.; Billig, A.; Litvak, V.; Moran, R.; et al. Dynamic causal modelling of COVID-19 [version 2; peer review: 2 approved]. *Wellcome Open Res.* **2020**, *5*, 89. [[CrossRef](#)]
49. Nosek, B.A.; Ebersole, C.R.; DeHaven, A.C.; Mellor, D.T. The preregistration revolution. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2600–2606. [[CrossRef](#)]
50. Nociti, V.; Losavio, F.A.; Gnoni, V.; Losurdo, A.; Testani, E.; Vollono, C.; Frisullo, G.; Brunetti, V.; Mirabella, M.; Della Marca, G. Sleep and fatigue in multiple sclerosis: A questionnaire-based, cross-sectional, cohort study. *J. Neurol. Sci.* **2017**, *372*, 387–392. [[CrossRef](#)]
51. Runge, J. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, PMLR, Lanzarote, Canary Islands, Spain, 9–11 April 2018; pp. 938–947. ISSN: 2640-3498.
52. Robins, J.M.; Scheines, R.; Spirtes, P.; Wasserman, L. Uniform consistency in causal inference. *Biometrika* **2003**, *90*, 491–515. [[CrossRef](#)]
53. Wal, W.M.v.d.; Geskus, R.B. ipw: An R Package for Inverse Probability Weighting. *J. Stat. Softw.* **2011**, *43*, 1–23. [[CrossRef](#)]
54. Robins, J.M.; Hernan, M.A.; Brumback, B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* **2000**, *11*, 550. [[CrossRef](#)]
55. Bach, P.; Chernozhukov, V.; Kurz, M.S.; Spindler, M. DoubleML—An Object-Oriented Implementation of Double Machine Learning in R. *arXiv* **2023**, arXiv:2103.09603 [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.